On Causal Mediation Analysis with a Survival Outcome

by

Eric J. Tchetgen Tchetgen*

Departments of Epidemiology and Biostatistics, Harvard University

Abstract

Suppose that, having established a marginal total effect of a point exposure on a time-to-event outcome, an investigator wishes to decompose this effect into its direct and indirect pathways, also know as natural direct and indirect effects, mediated by a variable known to occur after the exposure and prior to the outcome. This paper proposes a theory of estimation of natural direct and indirect effects in two important semiparametric models for a failure time outcome. The underlying survival model for the marginal total effect and thus for the direct and indirect effects, can either

Key Words and Phrases: Cox Proportional Hazards Model, Additive Hazards Model, Natural direct effects, Natural indirect effects; multiple robust, mediation analysis

AMS 1991 Subject Classifications. Primary: 62G05 .

*Supported by NIH grant #R21ES019712

be a marginal structural Cox proportional hazards model, or a marginal structural additive hazards model. The proposed theory delivers new estimators for mediation analysis in each of these models, with appealing robustness properties. Specifically, in order to guarantee ignorability with respect to the exposure and mediator variables, the approach, which is multiply robust, allows the investigator to use several flexible working models to adjust for confounding by a large number of pre-exposure variables. Multiple robustness is appealing because it only requires a subset of working models to be correct for consistency; furthermore, the analyst need not know which subset of working models is in fact correct to report valid inferences. Finally, a novel semiparametric sensitivity analysis technique is developed for each of these models, to assess the impact on inference, of a violation of the assumption of ignorability of the mediator.

1 Introduction

Suppose that, upon establishing a marginal total effect of a point exposure on an outcome of interest, an investigator wishes to decompose this effect into its direct and indirect pathways, also know as natural or pure direct and indirect effects, mediated by a variable known to occur after the exposure and prior to the outcome (Robins and Greenland, 1992, Pearl, 2001). The literature on statistical methods for causal mediation analysis has blossomed in recent years with new results on idendification of direct and indirect effects, and a number of novel techniques for obtaining statistical inferences about these effects (van der Laan and Petersen, 2005, VandeWeele, 2009, Imai 2010a,b, Lange and Hansen, 2011, VanderWeele, 2011, Tchetgen Tchetgen and Shpitser, 2011a,b). With the exception of Tein and Mackinnon (2003), and the recent paper by Lange and Hansen (2011) and the accompanying commentary by VanderWeele (2011), who consider a survival context, the existing literature on causal mediation analysis has largely focused on structural models for a mean effect. The current paper aims to further develop methodology for mediation analysis for survival data. In fact, we propose a general theory of estimation of natural direct and indirect effects for two important semiparametric models of a failure time outcome. We assume that the underlying survival model for the marginal total effect and thus for the direct and indirect effects, can either be a marginal structural Cox proportional hazards model as in Robins (1998), or a marginal structural additive hazards model. Lange and Hansen (2011) were the first to consider the use of the additive hazards model for causal mediation analysis in a survival context; whereas Tein and Mackinnon (2003) and VanderWeele (2011) also consider the use of a Cox proportional hazards model for mediation analysis.

The current paper aims to extend these existing results in several important ways. Thus, we develop some new semiparametric estimators of direct and indirect effects for each of these models, with appealing robustness properties. Specifically, the proposed approach which is so-called multiply robust, allows the investigator to use several flexible working models in order to adjust for a possibly large number of pre-exposure confounders for both exposure and mediating variables. Multiple robustness is appealing because it only requires a subset of these working models to be correct for consistency; furthermore, the analyst need not know which subset of working models is in fact correct to report valid inferences. Finally, in this paper, a novel semiparametric sensitivity analysis technique is also developed for each model, to assess the impact on mediation inferences, of a violation of the assumption of ignorability of the mediating variable. This is an important contribution in its own right, particularly because no methodology currently exist for performing a sensitivity analysis in the current survival context.

The theory developed in this paper parallels similar theory recently proposed by Tchetgen Tchetgen and Shpitser (2011a,b) for making inferences about natural direct and indirect effects of the exposure on the mean of the outcome. In section 2, we adapt these previous results to obtain multiply robust inferences about natural direct and indirect effects of a binary exposure on the marginal survival curve in the presence of confounding and right censoring. Because the previous theory does not directly apply to semiparametric regression models for survival data, new methodology is developed in Section 3 for obtaining multiply robust inferences about natural direct and indirect effects under a Cox model and an additive hazards model. Then, we develop similar multiply robust estimators of natural indirect effects for each model. Finally, Section 4 gives new results on semiparametric sensitivity analysis in a survival context.

First we introduce some notation. Throughout, we suppose independent and identically distributed data on a vector (E, M, X, T^*, Δ) is collected for n subjects. Here, E is the binary exposure variable, M is a mediator variable with support $\mathcal{S}$, known to occur subsequently to E and prior to T^*, and X is a vector of pre-exposure variables with support $\mathcal{X}$ that confound the association between (E, M) and the underlying failure time of interest T. Because of censoring, we observe $\Delta = I(T \leq C)$ and $T^* = \min(T, C)$ where C denotes an individual's right censoring time. Throughout, we assume that conditional on E, censoring is independent of (M, X, T), although in principle, this latter assumption may be relaxed as in Robins and Rotnitzky (1992) and van der Laan and Robins (2003). To limit the amount of unmeasured confounding, we suppose that X contains several variables, and thus is likely of moderate to high dimension. We assume that for each level $\{E = e, M = m\}$, there exist a counterfactual variable $T_{e,m}$ corresponding to the outcome had possibly contrary to fact the exposure and mediator variables taken the value (e, m) and for $\{E = e\}$, there exist a counterfactual variable M_e corresponding to the mediator variable had possibly contrary to fact the exposure variable taken the value e.

Although the paper focuses on a binary exposure, we note that the extension to a polytomous exposure is trivially deduced from the exposition.

2 Mediation analysis for a marginal survival probability

Let $D^*(t)$ denote $I(T^* \geq t)$, $D(t)$ denote $I(T \geq t)$ and define the corresponding counterfactual at risk process $D_{em}(t) = I(T_{em} \geq t)$. Also, let $S_{em}(t) = \mathbb{E}\{D_{em}(t)\} = \mathbb{E}\{I(T_{em} \geq t)\}$ denote the survival probability at time t had possibly contrary to fact the exposure and mediator variables taken the value (e, m); and let $S_{T|E,M,X}(t|E, M, X)$ denote the conditional survival probability of T at t. Consider the following decomposition of the total effect of E on the survival probability at time t:

$$\overbrace{S_{1M_1}(t) - S_{0M_0}(t)}^{\text{total effect}} = \overbrace{S_{1M_1}(t) - S_{1M_0}(t)}^{\text{natural indirect effect}} + \underbrace{S_{1M_0}(t) - S_{0M_0}(t)}_{\text{natural direct effect}} \tag{1}$$

As shown in the display above, the natural direct effect captures the effect of the exposure when one intervenes to set the mediator to the (random) level it would have been in the absence of exposure (Robins and Greenland, 1992, Pearl 2001). Such an effect generally differs from the controlled direct effect which refers to the exposure effect that arises upon intervening to set the mediator to a fixed level that may differ from its actual observed value (Robins and Greenland, 1992, Pearl, 2001, Robins, 2003). As noted by Pearl (2001), controlled direct and indirect effects are particularly relevant for policy making whereas natural direct and indirect effects are more useful for understanding the underlying mechanism by which the exposure operates.

Identification of natural direct and indirect effects requires additional assumptions.

To proceed, we make the consistency assumption:

$$\text{if } E = e\,, \text{then } M_e = M \text{ w.p.1}$$

$$\text{and if } E = e \quad \text{and } M = m \quad \text{then } T_{e,m} = T \text{ w.p.1}$$

In addition, we adopt the sequential ignorability assumption of Imai et al (2010) which states that for $e, e' \in \{0, 1\}$:

$$\{T_{e',m}, M_e\} \amalg E|X \tag{2}$$

$$T_{e'm} \amalg M|E = e, X \tag{3}$$

paired with a standard positivity assumption:

$$f_{M|E,X}\,(m|E, X) > 0 \text{ w.p.1 for each } m \in \mathcal{S}$$

$$\text{and } f_{E|X}\,(e|X) > 0 \text{ w.p.1 for each } e \in \{0, 1\}$$

where $f_{M|E,X}$ is the density of $[M|E, X]$ and $f_{E|X}$ is the density of $[E|X]$. Then, under the consistency assumption, the first part of the sequential ignorability assumption (2) and the positivity assumption, one can show that $S_{eM_e}(t)$ is identified by the g-formula of Robins (1997); under the additional assumption given by the second part of the sequential ignorability assumption (3), one can further show as in Imai et al (2010a), that:

$$S_{1M_0}(t) = \theta_t \tag{4}$$
$$= \iint\limits_{\mathcal{S}\times\mathcal{X}} S_{T|E,M,X}(t|E=1, M=m, X=x) f_{M|E,X}\left(m|E=0, X\right) f_X(x) d\mu(m,x)$$

where $f_{M|E,X}$ and f_X are respectively the conditional density of the mediator M given (E, X) and the density of X, and μ is a dominating measure for the distribution of $[M, X]$. Thus $S_{1M_0}(t)$ is identified from the observed data (See Pearl, 2011 and van der Laan and Petersen (2005) for related identification results). We note that the second part of the sequential ignorability assumption (3) is particularly strong and must be made with care. This is partly because, it is always possible that there might be unobserved variables that confound the relationship between the outcome and the mediator variables even upon conditioning on the observed exposure and covariates. Furthermore, the confounders X must all be pre-exposure variables, i.e. they must precede E. In fact, Avin et al (2005) proved that without additional assumptions, one cannot identify natural direct and indirect effects if there are confounding variables that are affected by the exposure even if such variables are observed by the investigator. This implies that similar to the ignorability of the exposure in observational studies, ignorability of the mediator cannot be established with certainty even after collecting as many pre-exposure confounders as possible. Furthermore, as Robins and Richardson (2010) point out, whereas the first part of the sequential ignorability assumption (2) could in principle be enforced in a randomized study, by randomizing

E within levels of X; the second part of the sequential ignorability assumption (3) cannot similarly be enforced experimentally, even by randomization. And thus for this latter assumption to hold, one must entirely rely on expert knowledge about the mechanism under study. For this reason, it will be crucial in practice to supplement mediation analyses with a sensitivity analysis that accurately quantifies the degree to which results are robust to a potential violation of the sequential ignorability assumption. For this purpose, later in the paper, we adapt and extend the sensitivity analysis technique of Tchetgen Tchetgen and Shpitser (2011a,b) to a survival analysis setting.

Theorem 1 of Tchetgen Tchetgen and Shpitser (2011a) implies that in order to obtain a consistent and asymptotically normal (CAN) estimator of the functional displayed in equation (4) and thus a CAN estimator of $S_{1M_0}(t)$ under the three assumptions given above, one must consistently estimate a subset of the following quantities $\{S_{T|E,M,X}, f_{M|E,X}, f_{E|X}\}$. Thus, let $\{\widehat{S}_{T|E,M,X}, \widehat{f}_{M|E,X}, \widehat{f}_{E|X}\}$ denote estimates of these required quantities, based on standard parametric or semiparametric working models for regression and density estimation. Because of the curse of dimensionality due to a high dimensional X, nonparametric methods for estimating these quantities are likely impractical for the sample sizes encountered in practice, and thus parametric/semiparametric models must be used. We emphasize that these three models are not of primary scientific interest but as later demonstrated, are needed for making inferences about mediation.

In principle, one could simply evaluate the functional under the estimated model to obtain the maximum likelihood estimator (MLE):

$$\widehat{\theta}_t^{tm} = \mathbb{P}_n \int_{\mathcal{S}} \widehat{S}_{T|E,M,X}(t|E=1, M=m, X)\widehat{f}_{M|E,X}\left(m|E=0, X\right) d\mu(m)$$

where $\mathbb{P}_n\left[\cdot\right] \equiv n^{-1}\sum_i \left[\cdot\right]_i$. However, one should then be concerned that model mis-specification of either $\widehat{S}_{T|E,M,X}$ or $\widehat{f}_{M|E,X}$ will likely lead to biased estimates of direct and indirect effects. Note that the MLE does not rely on a model for $f_{E|X}$ and thus is completely robust to a mis-specified estimate $\widehat{f}_{E|X}$. Two alternative estimators similar to those proposed by Tchetgen Tchetgen and Shpitser (2011a) for mean effects, can be obtained here, that respectively use $\{\widehat{S}_{T|E,M,X}, \widehat{f}_{E|X}\}$ only and $\{\widehat{f}_{M|E,X}, \widehat{f}_{E|X}\}$ only, and thus are respectively robust to mis-specification of $\widehat{f}_{M|E,X}$ and $\widehat{S}_{T|E,M,X}$. Indeed, in the first case, one could use:

$$\widehat{\theta}_t^{te} = \mathbb{P}_n \left\{ \frac{I(E=0)}{\widehat{f}_{E|X}(0|X)} \widehat{S}_{T|E,M,X}(t|E=1, M=m, X) \right\}$$

and in the second case one could use :

$$\widehat{\theta}_t^{em} = \mathbb{P}_n \left\{ \frac{\Delta}{\widehat{S}_{C|E}\left(T^{*-}|E=1\right)} \frac{I(E=1)I\left(T^* \geq t\right)}{\widehat{f}_{E|X}(E|X)} \frac{\widehat{f}_{M|E,X}\left(M|E=0, X\right)}{\widehat{f}_{M|E,X}\left(M|E, X\right)} \right\}$$

where $\widehat{S}_{C|E}\left(T^{*-}|E=e\right)$ denotes the exposure arm specific Kaplan-Meier estimator of the survival curve of censoring. $\widehat{S}_{C|E}^{-1}\left(T^{*-}|E=1\right)$ weights are needed here to correctly account for censoring (Robins and Rotnitzky, 1992, Satten and Datta, 2001) under the current assumption that censoring is ignorable conditional on E, and an additional standard positivity assumption (Robins and Rotnitzky, 1992). Unfortunately, as was the case for the MLE, these alternative estimators are likely severely biased if either of the working models they require is incorrect.

Theorem 2 of Tchetgen Tchetgen and Shpitser (2011a) partially resolves this potential difficulty, by providing a roadmap to construct an estimator $\widehat{\theta}_t \equiv \widehat{\theta}_t \left(\widehat{S}_{T|E,M,X}, \widehat{f}_{M|E,X}, \widehat{f}_{E|X} \right)$ that is partially robust to such model mis-specification, and remains CAN in the union model $\mathcal{M}_{union}$ that assumes at least one but not necessarily all of the following hold:

(a) the estimates of the conditional survival probability $\widehat{S}_{T|E,M,X}$ and of the conditional density of the mediator $\widehat{f}_{M|E,X}$ are consistent;

(b) the estimates of the conditional survival probability $\widehat{S}_{T|E,M,X}$ and of the conditional density of the exposure $\widehat{f}_{E|X}$ are consistent

(c) the estimates of the conditional densities of the exposure and mediator variables are consistent.

Clearly, such an estimator $\widehat{\theta}_t$ should generally be preferred to $\widehat{\theta}_t^{tm}$, $\widehat{\theta}_t^{te}$ and $\widehat{\theta}_t^{em}$ because an inference using $\widehat{\theta}_t$ is guaranteed to remain valid under many more data generating laws than an inference based on each of the other three estimators. $\widehat{\theta}_t$ is in fact so-called triply robust, as it delivers the correct inferences under the union of the three submodels (a), (b) and (c). By Theorem 2 of Tchetgen Tchetgen and

Shpitser (2011a), the following estimator is in fact triply robust:

$$\widehat{\theta}_t = \mathbb{P}_n \left[\begin{array}{c} \frac{\Delta}{\widehat{S}_{C,1}(T^{*-})} \frac{I(E=1)}{\widehat{f}_{E|X}(E|X)} \frac{\widehat{f}_{M|E,X}(M|E=0,X)}{\widehat{f}_{M|E,X}(M|E,X)} \left\{ I\left(T^* \geq t\right) - \widehat{S}_{T|E,M,X}(t|E=1,M,X) \right\} \\ + \frac{I(E=0)}{\widehat{f}_{E|X}(0|X)} \left\{ \widehat{S}_{T|E,M,X}(t|E=1,M,X) \ - \widehat{\eta}_t\left(1,0,X\right) \right\} \\ + \widehat{\eta}_t\left(1,0,X\right) \end{array} \right]$$

where

$$\widehat{\eta}_t\left(1,0,X\right) = \int_{\mathcal{S}} \widehat{S}_{T|E,M,X}(t|E=1,M=m,X) \widehat{f}_{M|E,X}\left(m|E=0,X\right) d\mu(m)$$

$\widehat{\theta}_t$ may in turn be combined as in Tchetgen Tchetgen and Shpitser (2011a) with an existing doubly robust estimator of the g-formula for $S_{eM_e}(t)$ (van der Laan and Robins, 2003, Bang and Robins, 2005), to obtain a triply robust estimator of the natural direct and indirect effects given in equation (1). To report confidence intervals, the nonparametric bootstrap could be used although an analytic expression and a corresponding estimator for the asymptotic variance of $\widehat{\theta}_t$ is easily derived from a standard Taylor series argument (see for example Tchetgen Tchetgen and Shpitser, 2011a).

3 Mediation analysis for two survival models

In this section, we consider the estimation of natural direct effects under two alternative structural models for the total effect of exposure: a Cox proportional hazards model (Cox PH) and an additive hazards model.

3.1 Proportional hazards model

The first model posits a Cox PH regression for the average total effect of the exposure, that is

$$\lambda_{T_e}(t) = \lambda_{T_0}(t) \exp(\beta_c e)$$

where $\lambda_{T_e}(t)$ denotes an individual's average hazard of experiencing an event at time t, had possibly contrary to fact, the person been exposed to $E = e$, and β_c encodes on the log-hazards scale, the total causal effect of exposure. As in VanderWeele (2011), one can decompose $\exp(\beta_c) = \lambda_{T_1}(t) / \lambda_{T_0}(t)$ into natural direct and indirect components:

$$\frac{\lambda_{T_1}(t)}{\lambda_{T_0}(t)} = \overbrace{\frac{\lambda_{T_{1M_1}}(t)}{\lambda_{T_{0M_0}}(t)}}^{\text{total effect}} = \frac{\overbrace{\lambda_{T_{1M_1}}(t)}^{\text{natural indirect effect}}}{\lambda_{T_{1M_0}}(t)} \times \underbrace{\frac{\lambda_{T_{1M_0}}(t)}{\lambda_{T_{0M_0}}(t)}}_{\text{natural direct effect}} \tag{5}$$

As we show next, unlike VanderWeele (2011) no rare outcome assumption is necessary for inference, but we further assume that the natural direct hazards ratio, and thus the indirect hazards ratio, agrees with the proportional hazards assumption of the total effect, and thus

$$\lambda_{T_{eM0}}(t) = \lambda_{T_{0M_0}}(t) \exp\left(\beta_c^{dir} e\right)$$

follows a Cox PH model where β_c^{dir} represents the direct effect of exposure, and similarly

$$\lambda_{T_{1M_e}}(t) = \lambda_{T_{1M_0}}(t) \exp\left(\beta_c^{ind} e\right)$$

where β_c^{ind} represents the indirect effect of exposure. This is an additional assumption since although unlikely in practice, in principle both direct and indirect effect could be functions of time in such a way that they combine to produce a time-constant total effect on the hazards ratio scale. Next, we describe some procedures for estimating the direct effect parameter β_c^{dir}.

Our first result generalizes the weighted strategy that previously gave $\widehat{\theta}_t^{em}$, and relies on the assumption that $\{\widehat{f}_{M|E,X}, \widehat{f}_{E|X}\}$ is consistent, however it does not use $\widehat{S}_{T|E,M,X}$.

Theorem 1: Under the consistency, sequential ignorability and positivity assumptions, $U^w\left(\beta_c^{dir}\right)$ is an unbiased estimating function for β_c^{dir}, where

$$U^w\left(\beta_c^{dir}\right) = U^w\left(\beta_c^{dir}; f_{M|E,X}, f_{E|X}\right) = \int dN^*(t) W \left[E - \frac{\xi_1\left(t; \beta_c^{dir}\right)}{\xi_2\left(t; \beta_c^{dir}\right)}\right], \qquad (6)$$

with

$$\xi_1\left(t; \beta_c^{dir}\right) = \mathbb{E}\left\{D^*\left(t\right) W E \exp\left(\beta_c^{dir} E\right)\right\},$$

$$\xi_2\left(t; \beta_c^{dir}\right) = \mathbb{E}\left\{D^*\left(t\right) W \exp\left(\beta_c^{dir} E\right)\right\},$$

$$W = \frac{f_{M|E,X}\left(M|E=0, X\right)}{f_{E|X}(E|X) f_{M|E,X}\left(M|E, X\right)}$$

and $N^*(t) = I(T^* \leq t, \Delta = 1)$ is the counting process of an observed failure time. Thus, β_c^{dir} is the solution of the equation:

$$\mathbb{E}\left\{U^w\left(\beta_c^{dir}\right)\right\} = 0$$

The proof of Theorem 1 is provided in the appendix; the result motivates the estimator $\widetilde{\beta}_c^{dir}$ that solves:

$$\mathbb{P}_n\left\{\widehat{U}^w\left(\widetilde{\beta}_c^{dir}\right)\right\}=0$$

where $\widehat{U}^w(\beta)=\widehat{U}^w\left(\beta;\widehat{f}_{M|E,X},\widehat{f}_{E|X}\right)$ is an empirical version of $U^w(\beta\ )$ defined as:

$$\int dN^*(t)\widehat{W}\left[E-\frac{\mathbb{P}_n\left\{D^*(t)\,\widehat{W}E\exp\left(\beta_c^{dir}E\right)\right\}}{\mathbb{P}_n\left\{D^*(t)\,\widehat{W}\exp\left(\beta_c^{dir}E\right)\right\}}\right] \tag{7}$$

with

$$\widehat{\xi}_1(t)=\mathbb{P}_n\left\{D^*(t)\,\widehat{W}E\exp\left(\beta_c^{dir}E\right)\right\},$$

$$\widehat{\xi}_2(t)=\mathbb{P}_n\left\{D^*(t)\,\widehat{W}\exp\left(\beta_c^{dir}E\right)\right\},$$

and $\widehat{W}$ defined as W under $\{\widehat{f}_{M|E,X},\widehat{f}_{E|X}\}$. Thus, under the key assumption that $\{\widehat{f}_{M|E,X},\widehat{f}_{E|X}\}$ is consistent (and converges in probability at rates faster than $n^{-1/4}$, see Newey (1994)), and under further standard regularity conditions $\widetilde{\beta}_c^{dir}$ is CAN with asymptotic variance that can be obtained by a standard Taylor expansion, or more conveniently by the nonparametric bootstrap. In the event that either $\widehat{f}_{M|E,X}$ or $\widehat{f}_{E|X}$ is not consistent, $\widetilde{\beta}_c^{dir}$ will generally be inconsistent. Thus we propose an alternative approach to estimate β_c^{dir}.

First, we note that because both $\xi_1(t)$ and $\xi_2(t)$ in equation (6) involve W, estimation of these functions of t requires correct models for $\{f_{M|E,X},f_{E|X}\}$. Thus, a key step in developing a multiply robust estimator of β_c^{dir} involves finding an alternative representation for these two functionals with better robustness properties. In this vein, for a given function $H = h(E)$ of E, let

$$
\begin{aligned}
&R(t,H;\beta_c^{dir}) = R(t,H;\beta_c^{dir}, S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}) \\
&= \left\{D^*(t) - S_{C|E}(t|E)\, S_{T|E,M,X}(t|E,M,X)\right\} W h(E) \exp\left(\beta_c^{dir} E\right) \\
&+ \left\{ \begin{array}{c} \sum_e \int S_{C|E}(t|E=e)\, S_{T|E,M,X}(t|E=e,M=m,X) \\ f_{M|E,X}(m|E=0,X)\, h(e) \exp\left(\beta_c^{dir} e\right) d\mu(m) \end{array} \right\} \\
&+ \tfrac{I(E=0)}{f(E|X)} \sum_e S_{C|E}(t|E=e)\, S_{T|E,M,X}(t|E=e,M\ ,X) h(e) \exp\left(\beta_c^{dir} e\right) \\
&- \tfrac{I(E=0)}{f(E|X)} \left[\begin{array}{c} \sum_e \int S_{C|E}(t|E=e)\, S_{T|E,M,X}(t|E=e,M=m,X) \\ f_{M|E,X}(m|E=0,X)\, h(e) \exp\left(\beta_c^{dir} e\right) d\mu(m) \end{array} \right]
\end{aligned}
$$

and for $H_1 = E$ and $H_2 = 1$, define

$\xi_j^{mr}\left(t;\beta_c^{dir}\right) = \xi_j^{mr}\left(t;\beta_c^{dir}, S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right) = \mathbb{E}\left\{R(t,H_j;\beta_c^{dir})\right\},\ j = 1,2.$

Next, define $R^{\ddagger}(t,H_j;\beta_c^{dir})$ as $R(t,H_j;\beta_c^{dir})$ under the law $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\}$. In the appendix, we establish that $\xi_j^{mr}\left(t;\beta_c^{dir}\right) = \xi_j\left(t;\beta_c^{dir}\right)$, $j = 1,2$, and in fact, we prove that this alternative representation is multiply robust, in the sense that $\xi_j^{mr,\ddagger}\left(t;\beta_c^{dir}\right) = \mathbb{E}\left\{R^{\ddagger}(t,H_j;\beta_c^{dir})\right\} = \xi_j\left(t;\beta_c^{dir}\right)$ provided that at least one of the following three conditions hold: either $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}\} = \{S_{T|E,M,X}, f_{M|E,X}\}$ or $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{E|X}\} = \{S_{T|E,M,X}, f_{E|X}\}$, or $\{f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\} = \{f_{M|E,X}, f_{E|X}\}$. In the appendix, we use this result to establish the following theorem:

Theorem 2: Under the consistency, sequential ignorability and positivity assumptions, $U^{mr}\left(\beta_c^{dir}\right) = U^{mr}\left(\beta_c^{dir}; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right)$ *is an unbiased estimat-*

ing function for β_c^{dir}, *where*

$$U^{mr}\left(\beta_c^{dir}\right) = \int \left\{dN^{*}(t) - S_{C|E}\left(t|E\right) f_{T|E,M,X}(t|E, M\ , X)dt\right\} W \left\{E - \frac{\xi_1^{mr}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr}(t)}\right\}$$

$$+\int\int\sum_{e\in\{0,1\}}\left[\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,m,X) f_{M|E,X}\left(m|E=0,X\right) \\ \times\left\{e - \frac{\xi_1^{mr}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr}(t)}\right\}\end{array}\right] d\left(\mu\left(m\right),t\right)$$

$$+\frac{I(E=0)}{f_{E|X}(E|X)}\int\sum_{e\in\{0,1\}}\left[\begin{array}{c} \left\{S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M\ ,X)\right\} \\ \times\left\{e - \frac{\xi_1^{mr}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr}\left(t;\beta_c^{dir}\right)}\right\}\end{array}\right] dt$$

$$-\frac{I(E=0)}{f_{E|X}(E|X)}\int\int\sum_{e\in\{0,1\}}\left[\begin{array}{c} \left\{\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M=m\ ,X) \\ \times f_{M|E,X}\left(m|E=0,X\right)\end{array}\right\} \\ \times\left\{e - \frac{\xi_1^{mr}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr}\left(t;\beta_c^{dir}\right)}\right\}\end{array}\right] d\left(\mu\left(m\right),t\right)$$

Furthermore,

$$\mathbb{E}\left\{U^{mr}\left(\beta_c^{dir}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}\right)\right\} = 0 \qquad (8)$$

if one but not necessarily all three of the following conditions holds: either $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}\} = \{S_{T|E,M,X}, f_{M|E,X}\}$ *or* $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{E|X}\} = \{S_{T|E,M,X}, f_{E|X}\}$, *or* $\{\ f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\} = \{\ f_{M|E,X}, f_{E|X}\}$; *with* $f_{T|E,M,X}(t|E=e, M=m\ , X) = -\partial S_{T|E,M,X}\left(t|E,M,X\right)/\partial t$, *the density of* $\left[T|E,M,X\right]$.

According to Theorem 2, a multiply robust estimator $\widehat{\beta}_c^{dir}$ is obtained by solving the equation:

$$\mathbb{P}_n\left\{\widehat{U}^{mr}\left(\widehat{\beta}_c^{dir}; \widehat{S}_{T|E,M,X}, \widehat{f}_{M|E,X}, \widehat{f}_{E|X}, \widehat{S}_{C|E}\ \right)\right\} = 0$$

where $\widehat{U}^{mr}\left(\cdot;\cdot,\cdot,\cdot\right)$ is obtained by substituting $\mathbb{P}_n\left[\cdot\right]$ for all marginal expectations. so that under standard regularity conditions, $\widehat{\beta}_c^{dir}$ is CAN in model $\mathcal{M}_{union}$. An analytical expression for the asymptotic variance of $\widehat{\beta}_c^{dir}$ under $\mathcal{M}_{union}$ can be obtained by a standard Taylor series expansion and a standard calculus for martingale integrals which is not pursued here. Alternatively, one could also use the nonparametric bootstrap for inference which is more convenient.

To estimate the indirect log hazards ratio β_c^{ind}, we observe that by the decomposition given in equation (5), $\beta_c^{ind} = \beta_c - \beta_c^{dir}$ where β_c is the total log hazards ratio, i.e. $\lambda_{T_1}(t)/\lambda_{T_0}(t) = \exp(\beta_c)$. This immediately gives a simple approach for obtaining an estimator of the indirect effect. The approach entails first estimating β_c by using standard inverse-probability-of-treatment weighting for total effects. Following Robins (1998), $\widetilde{\beta}_c$ is obtained by solving equation (7) upon substituting $\widehat{W}$ with $\widehat{f}_{E|X}^{-1}(E|X)$. Then, we can define an estimator of $\widetilde{\beta}_c^{ind}$ by $\widetilde{\beta}_c - \widehat{\beta}_c^{dir}$ or alternatively by $\widetilde{\beta}_c - \widetilde{\beta}_c^{dir}$. Unfortunately, both of these estimators are likely biased if $\widehat{f}_{E|X}^{-1}(E|X)$ is not consistent. As a remedy, the next theorem gives a multiply robust estimating function of β_c^{ind} .

Theorem 3: Suppose that β_c^{dir} *is known, then under the consistency, sequential ignorability and positivity assumptions,* $V^{mr}(\beta_c^{dir}, \beta_c^{ind}) = V^{mr}(\beta_c^{dir}, \beta_c^{ind}; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E})$ *is an unbiased estimating function for* β_c^{ind}, *where*

$V^{mr}(\beta_c^{dir}, \beta_c^{ind})$

$$= \int \left[\begin{array}{c} \left\{ \begin{array}{c} dN^*(t) \\ -\int \left\{ \begin{array}{c} S_{C|E}\left(t|E\right) f_{T|E,M,X}(t|E, M = m\ , X) \\ \times f_{M|E,X}\left(m|E, X\right) \end{array} \right\} d\left(\mu\left(m\right), t\right) \end{array} \right\} f^{-1}_{E|X}(E|X) \\ \times \left\{ E - \frac{\vartheta_1^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}\right)}{\vartheta_2^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}\right)} \right\} \end{array} \right]$$

$$+ \int\int \sum\nolimits_{e\in\{0,1\}} \left[\begin{array}{c} S_{C|E}\left(t|E = e\right) f_{T|E,M,X}(t|E = e, M\ , X) f_{M|E,X}\left(m|E = e, X\right) \\ \times \left\{ e - \frac{\vartheta_1^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}\right)}{\vartheta_2^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}\right)} \right\} \end{array} \right] d\left(\mu\left(m\right), t\right),$$

with

$$\vartheta_j^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}\right) = \vartheta_j^{mr}\left(t;\beta_c^{dir},\beta_c^{ind}, S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right) = \mathbb{E}\left\{G(t, H_j;\beta_c^{dir},\beta_c^{ind})\right\},$$

$j = 1, 2$

$$G(t, H;\beta_c^{dir},\beta_c^{ind}) =$$

$$\left\{ \begin{array}{c} D^*\left(t\right) \\ -\int \left\{ \begin{array}{c} S_{C|E}\left(t|E\right) S_{T|E,M,X}(t|E, M = m\ , X) \\ f_{M|E,X}\left(m|E, X\right) \end{array} \right\} d\mu\left(m\right) \end{array} \right\}$$

$$\times f^{-1}_{E|X}(E|X) h(E) \exp\left\{ \left(\beta_c^{dir} + \beta_c^{ind}\right) E \right\}$$

$$+ \sum\nolimits_e \int \left\{ \begin{array}{c} S_{C|E}\left(t|E = e\right) S_{T|E,M,X}(t|E = e, M = m, X) \\ f_{M|E,X}\left(m|E = e, X\right) h(e) \exp\left\{ \left(\beta_c^{dir} + \beta_c^{ind}\right) e \right\} d\mu\left(m\right) \end{array} \right\}$$

Furthermore, $\left(\beta_c^{dir},\beta_c^{ind}\right)$ *solves*

$$\begin{aligned} \mathbb{E}\left\{V^{mr}\left(\beta_c^{ind},\beta_c^{dir}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}\right)\right\} &= 0 \\ \mathbb{E}\left\{U^{mr}\left(\beta_c^{dir}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}\right)\right\} &= 0 \end{aligned}$$

if one but not necessarily all three of the following conditions holds: either $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}\} =$

$\{S_{T|E,M,X}, f_{M|E,X}\}$ *or* $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{E|X}\} = \{S_{T|E,M,X}, f_{E|X}\}$, *or* $\{f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\} = \{f_{M|E,X}, f_{E|X}\}$; *with* $f_{T|E,M,X}(t|E = e, M = m\ , X) = -\partial S_{T|E,M,X}\left(t|E, M, X\right)/\partial t$, *the density of* $\left[T|E, M, X\right]$.

According to theorem 3, a multiply robust estimator $\widehat{\beta}_c^{ind}$ is obtained by solving the equation:

$$\mathbb{P}_n\left\{\widehat{V}^{mr}\left(\widehat{\beta}_c^{ind}, \widehat{\beta}_c^{dir}; \widehat{S}_{T|E,M,X}, \widehat{f}_{M|E,X}, \widehat{f}_{E|X}, \widehat{S}_{C|E}\right)\right\} = 0$$

where $\widehat{V}^{mr}\left(\cdot, \cdot; \cdot, \cdot, \cdot\right)$ is obtained by substituting $\mathbb{P}_n\left[\cdot\right]$ for all marginal expectations, so that under standard regularity conditions, $\widehat{\beta}_c^{ind}$ is CAN in model $\mathcal{M}_{union}$. we recommend the use of the nonparametric bootstrap for inference.

3.2 Additive hazards model

In some situations, assuming proportional hazards may not fit the data well, in which case, an additive hazards model will often fit the data better (Lin and Ying, 2004) . This alternative model assumes the average total effect of the exposure is additive on the hazards scale :

$$\lambda_{T_e}\left(t\right) = \lambda_{T_0}\left(t\right) + \beta_a e$$

where β_a encodes the total causal effect of exposure. As in Lange and Hansen (2011), one can decompose $\beta_a = \lambda_{T_1}(t) - \lambda_{T_0}(t)$ into natural direct and indirect components:

$$\begin{aligned} & \lambda_{T_1}(t) - \lambda_{T_0}(t) \\ = \; & \overbrace{\lambda_{T_{1M_1}}(t) - \lambda_{T_{0M_0}}(t)}^{\text{total effect}} = \overbrace{\lambda_{T_{1M_1}}(t) - \lambda_{T_{1M_0}}(t)}^{\text{natural indirect effect}} + \underbrace{\lambda_{T_{1M_0}}(t) - \lambda_{T_{0M_0}}(t)}_{\text{natural direct effect}} \end{aligned} \tag{9}$$

We further assume that the natural direct effect, and thus the indirect effect, agrees with the assumption of additive hazards, and thus

$$\lambda_{T_{eM0}}(t) = \lambda_{T_{0M_0}}(t) + \beta_a^{dir} e$$

where β_a^{dir} represents the direct effect of the exposure, and similarly

$$\lambda_{T_{1M_e}}(t) = \lambda_{T_{1M_0}}(t) + \beta_a^{ind} e$$

where β_a^{ind} represents the indirect effect of the exposure. As in the case of the Cox PH model, this is an assumption since although unlikely in practice, in principle both direct and indirect effects could be functions of time in such a way that they combine to produce an additive total effect. We describe some procedures for estimating the direct effect parameter β_a^{dir}.

The next result gives a weighted approach analoguous to that proposed for the Cox PH model

Theorem 4: Under the consistency, sequential ignorability and positivity assump-

tions, $Z^w\left(\beta_a^{dir}\right)$ *is an unbiased estimating function for* β_a^{dir}*, where*

$$Z^w\left(\beta_a^{dir}\right) = \int \left\{dN^*(t) - E\beta_a^{dir} D^*(t)\, dt\right\} W\left[E - \frac{\varpi_1(t)}{\varpi_2(t)}\right], \tag{10}$$

with

$$\begin{aligned} \varpi_1(t) &= \mathbb{E}\left\{D^*(t)\, WE\right\}, \\ \varpi_2(t) &= \mathbb{E}\left\{D^*(t)\, W\right\} \end{aligned}$$

Thus, β_c^{dir} *is the solution of the equation:*

$$\mathbb{E}\left\{Z^w\left(\beta_a^{dir}\right)\right\} = 0$$

The theorem implies that $\widetilde{\beta}_a^{dir}$ is CAN provided $\{\widehat{f}_{M|E,X}, \widehat{f}_{E|X}\}$ is consistent, where $\widetilde{\beta}_a^{dir}$ solves

$$\mathbb{P}_n\left\{\widehat{Z}^w\left(\widetilde{\beta}_a^{dir}\right)\right\} = 0$$

with

$$\widehat{Z}^w(\beta) = \int \{dN^*(t) - E\beta D^*(t)\, dt\}\, \widehat{W}\left[E - \frac{\widehat{\varpi}_1(t)}{\widehat{\varpi}_2(t)}\right] \tag{11}$$

an empirical version of Z^w. Thus, $\widetilde{\beta}_a^{dir}$ is not multiply robust. The next theorem provides a multiply robust estimating function of β_a^{dir}. First, we introduce some additional notation and let

$$\begin{aligned} &\varpi_j^{mr}(t) = \varpi_j^{mr}\left(t; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right) \\ &= \mathbb{E}\ \left[\left\{D^*(t) - S_{C|E}(t|E)\, S_{T|E,M,X}(t|E,M,X)\right\} W h_j(E)\right. \\ &+ \int \textstyle\sum_{e\in\{0,1\}} S_{C|E}(t|E=e)\, S_{T|E,M,X}(t|E=e, M=m, X) f_{M|E,X}(m|E=0,X)\, h_j(e)\, d\mu(m) \end{aligned}$$

$$+\frac{I(E=0)}{f_{E|X}(E|X)}\sum_{e\in\{0,1\}} S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M\ \ ,X)\ h_j\left(e\right)$$

$$\left.-\frac{I(E=0)}{f_{E|X}(E|X)}\int \sum_{e\in\{0,1\}}\left\{\begin{array}{c} S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M=m,X) \\ f_{M|E,X}\left(m|E=0,X\right) h_j\left(e\right)\end{array}\right\} d\mu\left(m\right)\right]$$

Theorem 5: Under the consistency, sequential ignorability and positivity assumptions, $Z^{mr}\left(\beta_a^{dir}\right) = Z^{mr}\left(\beta_a^{dir}; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right)$ *is an unbiased estimating function for* β_a^{dir}*, where*

$$Z^{mr}\left(\beta_a^{dir}\right) = \int \left\{\begin{array}{c} dN^{*}(t) - E\beta_a^{dir} D^{*}\left(t\right) dt \\ -S_{C|E}\left(t|E\right) f_{T|E,M,X}(t|E,M\ \ ,X)dt \\ +E\beta_a^{dir} S_{C|E}\left(t|E\right) S_{T|E,M,X}(t|E,M,X)dt \end{array}\right\} W\left\{E-\frac{\varpi_1^{mr}(t)}{\varpi_2^{mr}(t)}\right\}$$

$$+\int\int \sum_{e\in\{0,1\}}\left[\begin{array}{c} \left\{\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M=m,X) \\ -e\beta_a^{dir}\ S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M=m,X)\end{array}\right\} \\ \times\left\{e-\frac{\xi_1^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr,\ddagger}(t)}\right\} f_{M|E,X}\left(m|E=0,X\right) d\left(\mu\left(m\right),t\right)\end{array}\right]$$

$$+\frac{I(E=0)}{f_{E|X}(E|X)}\int \left\{\begin{array}{c} \sum_{e\in\{0,1\}}\left\{\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M,X) \\ -e\beta_a^{dir} S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M,X)\end{array}\right\} \\ \times\left\{e-\frac{\xi_1^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}\right\} d\left(t\right)\end{array}\right\}$$

$$-\frac{I(E=0)}{f_{E|X}(E|X)}\int\int \sum_{e\in\{0,1\}}\left[\begin{array}{c} \left\{\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M=m,X) \\ -e\beta_a^{dir} S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M=m,X)\end{array}\right\} \\ \times\left\{e-\frac{\xi_1^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}\right\} f_{M|E,X}\left(m|E=0,X\right) d\left(\mu\left(m\right),t\right)\end{array}\right]$$

Furthermore,

$$\mathbb{E}\left\{Z^{mr}\left(\beta_a^{dir}; S_{T|E,M,X}^{\ddagger}, f_{M|E,X}^{\ddagger}, f_{E|X}^{\ddagger}, S_{C|E}\right)\ \right\} = 0 \tag{12}$$

if one but not necessarily all three of the following conditions holds: either $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}\} = \{S_{T|E,M,X}, f_{M|E,X}\}$ *or* $\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{E|X}\} = \{S_{T|E,M,X}, f_{E|X}\}$, *or* $\{\ f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\} = \{\ f_{M|E,X}, f_{E|X}\}$

By theorem 5, a multiply robust estimator $\widehat{\beta}_a^{dir}$ is obtained by solving the equation:

$$\mathbb{P}_n\left\{\widehat{Z}^{mr}\left(\widehat{\beta}_a^{dir}; \widehat{S}_{T|E,M,X}, \widehat{f}_{M|E,X}, \widehat{f}_{E|X}, \widehat{S}_{C|E}\right)\right\} = 0$$

so that under standard regularity conditions, $\widehat{\beta}_a^{dir}$ is CAN in model $\mathcal{M}_{union}$. We recommend the nonparametric bootstrap for inference.

Suppose now that one wishes to estimate the indirect hazards difference β_a^{ind}. By the decomposition given in equation (9), $\beta_a^{dir} = \beta_a - \beta_a^{dir}$ where β_a^{total} is the total hazards difference, i.e. $\lambda_{T_1}(t) - \lambda_{T_0}(t) = \beta_a$. This decomposition immediately gives a simple estimator of the indirect effect based on a weighting scheme. The approach entails first estimating β_a by using inverse-probability-of-treatment weighting. Following Robins (1998), $\widetilde{\beta}_a$ is obtained by solving equation (11) upon replacing $\widehat{W}$ by $\widehat{f}^{-1}_{E|X}(E|X)$. Then, we can define an estimator of β_a^{ind} by $\widetilde{\beta}_a - \widehat{\beta}_a^{dir}$ or alternatively by $\widetilde{\beta}_a - \widetilde{\beta}_a^{dir}$. Unfortunately, just as in the Cox model, both of these estimators are likely biased if $\widehat{f}^{-1}_{E|X}(E|X)$ is not consistent. As a remedy, the next theorem gives a multiply robust estimating function of β_a^{ind}.

Theorem 6: Suppose β_a^{dir} *is known, then under the consistency, sequential ignorability and positivity assumptions,* $P^{mr}\left(\beta_a^{dir}, \beta_a^{ind}\right) = P^{mr}(\beta_a^{dir}, \beta_a^{ind}; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E})$ *is an unbiased estimating function for* β_a^{ind}, *where*

$$P^{mr}\left(\beta_a^{dir},\beta_a^{ind}\right)$$
$$=\int\left\{\begin{array}{c} dN^{*}(t)-E\left(\beta_a^{dir}+\beta_a^{ind}\right)D^{*}(t)\,dt \\ -\int S_{C|E}(t|E)\,f_{T|E,M,X}(t|E,m\ ,X)f_{M|E,X}(m|E,X)\,d(\mu(m),t) \\ +\int E\left(\beta_a^{dir}+\beta_a^{ind}\right)S_{C|E}(t|E)\,S_{T|E,M,X}(t|E,m,X)f_{M|E,X}(m|E,X)\,d(\mu(m),t) \end{array}\right\}f_{E|X}^{-1}(E|X$$
$$+\int\int\sum\nolimits_{e\in\{0,1\}}\left[\begin{array}{c}\left\{\begin{array}{c} S_{C|E}(t|E=e)\,f_{T|E,M,X}(t|E=e,M=m,X) \\ -e\left(\beta_a^{dir}+\beta_a^{ind}\right)\ S_{C|E}(t|E=e)\,S_{T|E,M,X}(t|E=e,M=m,X)\end{array}\right\} \\ \times\left\{e-\frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}f_{M|E,X}(m|E=e,X)\,d(\mu(m),t)\end{array}\right]$$

with $\phi_j^{mr}(t)=\phi_j^{mr}\left(t;S_{T|E,M,X},f_{M|E,X},f_{E|X},S_{C|E}\right)$

$$=\mathbb{E}\left[\left\{D^{*}(t)-\int S_{C|E}(t|E)\,S_{T|E,M,X}(t|E,m,X)f_{M|E,X}(m|E,X)\,d\mu(m)\right\}f_{E|X}^{-1}(E|X)h_j(E)\right.$$
$$\left.+\int\sum\nolimits_{e\in\{0,1\}}S_{C|E}(t|E=e)\,S_{T|E,M,X}(t|E=e,M=m,X)f_{M|E,X}(m|E=e,X)\,h_j(e)\,d\mu(m)\right]$$

Furthermore,

$$\begin{aligned}\mathbb{E}\left\{P^{mr}\left(\beta_a^{dir},\beta_a^{ind};S_{T|E,M,X}^{\ddagger},f_{M|E,X}^{\ddagger},f_{E|X}^{\ddagger},S_{C|E}\right)\right\} &= 0 \\ \mathbb{E}\left\{Z^{mr}\left(\beta_a^{dir};S_{T|E,M,X}^{\ddagger},f_{M|E,X}^{\ddagger},f_{E|X}^{\ddagger},S_{C|E}\right)\right\} &= 0\end{aligned}\qquad(13)$$

if one but not necessarily all three of the following conditions holds: either $\{S_{T|E,M,X}^{\ddagger},f_{M|E,X}^{\ddagger}\}=\{S_{T|E,M,X},f_{M|E,X}\}$ *or* $\{S_{T|E,M,X}^{\ddagger},f_{E|X}^{\ddagger}\}=\{S_{T|E,M,X},f_{E|X}\}$, *or* $\{f_{M|E,X}^{\ddagger},f_{E|X}^{\ddagger}\}=\{f_{M|E,X},f_{E|X}\}$.

According to theorem 6, a multiply robust estimator $\widehat{\beta}_a^{ind}$ is obtained by solving the equation:

$$\mathbb{P}_n\left\{\widehat{P}^{mr}\left(\widehat{\beta}_a^{ind},\widehat{\beta}_a^{dir};\widehat{S}_{T|E,M,X},\widehat{f}_{M|E,X},\widehat{f}_{E|X},\widehat{S}_{C|E}\right)\right\}=0$$

where $\widehat{P}^{mr}(\cdot,\cdot;\cdot,\cdot,\cdot,\cdot)$ is obtained by substituting $\mathbb{P}_n[\cdot]$ for all marginal expectations, then under standard regularity conditions, $\widehat{\beta}_a^{ind}$ is CAN in model $\mathcal{M}_{union}$; thus one can use the nonparametric bootstrap for inference.

4 A semiparametric sensitivity analysis

In this section, we extend the semiparametric sensitivity analysis technique proposed by Tchetgen Tchetgen and Shpitser (2011a,b), to assess the extent to which a violation of the ignorability assumption for the mediator might alter inferences about natural direct or indirect effects in the survival context. Let

$$\gamma(t,e,m,x) = \lambda_{T_{1,m}|E,M,X}(t|E=e,M=m,X=x) - \lambda_{T_{1,m}|E,M,X}(t|E=e,M\neq m,X=x)$$

then

$$T_{e',m} \not\perp\!\!\!\perp M | E=e, X$$

i.e. a violation of the ignorability assumption for the mediator variable, generally implies that $\gamma(t,e,m,x) \neq 0$ for some (t,e,m,x). Suppose M is binary and larger values of T are beneficial for health, then if $\gamma(t,e,1,x) < 0$ but $\gamma(t,e,0,x) > 0$ for all t, then on average, individuals with $\{E=e,X=x\}$ and mediator value $\{M=0\}$ have a higher hazard function for each of the potential outcomes $\{T_{11,}T_{10}\}$ than individuals with $\{E=e,X=x\}$ but $\{M=1\}$; i.e. healthier individuals are more likely to receive the mediator. On the other hand, if $\gamma(t,e,0,x) < 0$ but $\gamma(t,e,1,x) > 0$ for all t, suggests confounding by indication for the mediator variable; i.e. unhealthier individuals are more likely to receive the mediating factor.

We proceed as in Robins et al (1999) who proposed using a selection bias function for the purposes of conducting a sensitivity analysis for total effects, and Tchetgen Tchetgen and Shpitser (2011a,b) who adapted the approach for assessing the impact of unmeasured confounding on the estimation of average natural direct and indirect effects. Here we propose to recover inferences about natural direct effects on the hazard function, under either an additive or a proportional hazards model, by assuming the selection bias function $\gamma(t, e, m, x)$ is known, which encodes the magnitude and direction of the unmeasured confounding for the mediator. In the following, $\mathcal{S}$ is assumed to be finite. To motivate the proposed approach, suppose for the moment that $f_{M|E,X}$ is known, then under the assumption that the exposure is ignorable given X, we show in the appendix that the following lemma holds:

Lemma 1:Let

$$\begin{aligned}\delta(t, e, m, x) &= \delta\left(t, e, m, x; f_{M|E,X}\right) \\ &= \frac{f_{M|E,X}(m|E=e, X=x) + \left\{1 - f_{M|E,X}(m|E=e, X=x)\right\}\exp\left\{\int_0^t \gamma(u, e, m, x)\,du\right\}}{f_{M|E,X}(m|E=0, X=x) + \left\{1 - f_{M|E,X}(m|E=0, X=x)\right\}\exp\left\{\int_0^t \gamma(u, 0, m, x)\,du\right\}}\end{aligned}$$

and

$$\dot{\delta}(t, 1, m, x) = \frac{\partial \log \delta(u, 1, m, x)}{\partial u}|_{u=t}$$

Under the consistency assumption and the first part of the sequential ignorability *assumption (2)*

$$
\begin{aligned}
& S_{T_{1,M_0}|M_0,X}(t|M_0=m,X=x) \\
= \; & S_{T_{1,m}|E,M,X}(t|E=0,M=m,X=x) \\
= \; & S_{T|E,M,X}(t|E=1,M=m,X=x)\times\delta(t,1,m,x)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
& \lambda_{T_{1,M_0}|M_0,X}(t|M_0=m,X=x) \\
= \; & \lambda_{T_{1,m}|E,M,X}(t|E=0,M=m,X=x) \\
= \; & \lambda_{T|E,M,X}(t|E=1,M=m,X=x)-\dot{\delta}(t,1,m,x)
\end{aligned}
$$

Lemma 1 implies that $S_{T_{1,M_0}}(t)$ is identified by:

$$\mathbb{E}\left(\sum_{m\in\mathcal{S}} S_{T|E,M,X}(t|E=1,M=m,X=x)\,\delta(t,1,m,x)\,f_{M|E,X}(m|E=0,X)\right) \tag{14}$$

Below, we use this result to obtain consistent estimators of $\left\{\beta_j^{dir},\beta_j^{ind}: j=a,c\right\}$ assuming $\gamma(\cdot,\cdot,\cdot,\cdot)$ is known. A sensitivity analysis is then obtained as in Tchetgen Tchetgen and Shpitser (2011a,b) by repeating this process and by reporting inferences for each choice of $\gamma(\cdot,\cdot,\cdot,\cdot)$ in a finite set of user–specified functions $\Gamma=\{\gamma_\alpha(\cdot,\cdot,\cdot,\cdot):\alpha\}$ indexed by a finite dimensional parameter α with $\gamma_0(\cdot,\cdot,\cdot,\cdot)\in\Gamma$ corresponding to the ignorability assumption of M, i.e. $\gamma_0(\cdot,\cdot,\cdot,\cdot)\equiv 0$. Throughout, models for the probability mass functions of $[M|E,X]$ and $[E|X]$ are assumed to be correct. Thus, to implement the sensitivity analysis technique, we develop a semiparametric estimator of $\left\{\beta_j^{dir},\beta_j^{ind}:j=a,c\right\}$ in a model $\mathcal{M}_1$ that assumes the model for $[M,E|X]$ is known up to a set of finite dimensional parameters, and in which the selection bias function is known, $\gamma\left(\cdot,\cdot,\cdot,\cdot\right)=\gamma_{\alpha^*}\left(\cdot,\cdot,\cdot,\cdot\right)$ for α^* fixed .

For the Cox PH model , we propose to use the following modified estimating function for estimating the direct effect under $\mathcal{M}_1$, which carefully incorporates the selection bias function:

$$\begin{aligned} U^w\left(\beta_c^{dir},\alpha^*\right) &= \int \delta_{\alpha^*}\left(t,E,M,X\right)\left\{dN^*(t)-\dot{\delta}_{\alpha^*}\left(t,E,M,X\right)D^*\left(t\right)dt\right\}W \\ &\quad \times\left\{E-\frac{\mathbb{E}\left\{D^*\left(t\right)WE\delta_{\alpha^*}\left(t,E,M,X\right)\exp\left(\beta_c^{dir}E\right)\right\}}{\mathbb{E}\left\{D^*\left(t\right)W\delta_{\alpha^*}\left(t,E,M,X\right)\exp\left(\beta_c^{dir}E\right)\right\}}\right\}\end{aligned}$$

where $\delta_{\alpha^*}\left(\cdot,\cdot,\cdot,\cdot\right)$ is defined as $\delta\left(\cdot,\cdot,\cdot,\cdot\right)$ under $\gamma_{\alpha^*}\left(\cdot,\cdot,\cdot,\cdot\right)$. For the additive model, one can use the following modified estimating function under $\mathcal{M}_1$:

$$\begin{aligned} Z^w\left(\beta_a^{dir},\alpha^*\right) &= \int\left\{\begin{array}{c} dN^*(t)-\dot{\delta}_{\alpha^*}\left(t,E,M,X\right)D^*\left(t\right)dt \\ -E\beta_a^{dir}D^*\left(t\right)\dot{\delta}_{\alpha^*}\left(t,E,M,X\right)dt\end{array}\right\}\delta_{\alpha^*}\left(t,E,M,X\right)W \\ &\quad \times\left\{E-\frac{\mathbb{E}\left\{D^*\left(t\right)WE\delta_{\alpha^*}\left(t,E,M,X\right)\right\}}{\mathbb{E}\left\{D^*\left(t\right)W\delta_{\alpha^*}\left(t,E,M,X\right)\right\}}\right\}\end{aligned}$$

In the appendix, we show the following result holds:

Theorem 7:Suppose $\gamma\left(\cdot,\cdot,\cdot,\cdot\right)=\gamma_{\alpha^*}\left(\cdot,\cdot,\cdot,\cdot\right)$, *then under the consistency and positivity assumptions, and the ignorability assumption for the exposure, and under the Cox PH model,* $\beta_c^{dir}=\beta_c^{dir}\left(\alpha^*\right)$ *solves the equation*

$$\mathbb{E}\left\{U^w\left(\beta_c^{dir},\alpha^*\right)\right\}=0 \tag{15}$$

Similarly, under the additive hazards model, $\beta_a^{dir} = \beta_a^{dir}(\alpha^*)$ *solves the equation*

$$\mathbb{E}\left\{Z^w\left(\beta_a^{dir}, \alpha^*\right)\right\} = 0$$

Thus, under model $\mathcal{M}_1$ and the Cox PH assumption, a sensitivity analysis then entails reporting the set $\left\{\widehat{\beta}_c^{dir}(\alpha) : \alpha\right\}$ (and the associated confidence intervals) which summarizes how sensitive inferences are to a deviation from the ignorability assumption $\alpha = 0$, where $\widetilde{\beta}_c^{dir}(\alpha)$ solves an empirical version of equation (15) with unknown quantities estimated under the model. A sensitivity analysis is similarly obtained for the additive hazards model, and inferences about indirect effects are obtained as in Section 3, upon substituting $\left\{\widehat{\beta}_c^{dir}(\alpha), \widehat{\beta}_a^{dir}(\alpha) : \alpha\right\}$ for $\left\{\widehat{\beta}_c^{dir}, \widehat{\beta}_a^{dir}\right\}$. In the appendix, we describe a doubly robust sensitivity analysis technique which further extends these results, by recovering correct sensitivity analyses under a union model in which, $\widehat{f}_{M|E,X}$ is assumed to be consistent, however, only one but not necessarily both $f_{T|M,E,X}$ and $f_{E|X}$ need to be consistently estimated.

It is helpful for practice, to briefly describe possible functional forms for the selection bias function $\gamma_\alpha(\cdot,\cdot,\cdot,\cdot)$. In the simple case where M is binary, it may be convenient to specify a single parameter model such as one of the following:

$$\begin{aligned}
\gamma_{\alpha,1}(t,e,m,x) &= \alpha t(2m-1) & \gamma_{\alpha,2}(t,e,m,x) &= \alpha tm \\
\gamma_{\alpha,3}(t,e,m,x) &= \alpha t(2m-1)e & \gamma_{\alpha,4}(t,e,m,x) &= \alpha tme \\
\gamma_{\alpha,5}(t,e,m,x) &= \alpha t(2m-1)ex_1 & \gamma_{\alpha,6}(t,e,m,x) &= \alpha tmex_1
\end{aligned}$$

where for each of the above functional forms, the scalar parameter α encodes the magnitude and direction of unmeasured confounding for the mediator.

The functions $\gamma_{\alpha,3}, \gamma_{\alpha,4}, \gamma_{\alpha,5}$ and $\gamma_{\alpha,6}$ model interactions with the exposure variable and a component X_1 of $X,$ thus allowing for heterogeneity in the selection bias function over time. Since the functional form of γ_α is not identified from the observed data, we generally recommend reporting results for a variety of functional forms.

It is important to note that the sensitivity analysis technique introduced above appears to be the first of its kind for survival data. While a variety of techniques have previously been proposed for conducting sensitivity analyses for unmeasured confounding in the context of mediation, for example, VanderWeele (2010), Imai et al (2010a), Tchetgen Tchetgen and Shpitser (2011a,b), none of the existing techniques apply to mediation in the survival context under either a Cox PH model or an additive hazards model. It is also important to note that by concisely encoding a possible violation of the ignorability assumption for the mediator through a selection bias function the proposed approach avoids having to spell out in detail, the possible nature of the unmeasured confounding; although in practice, as illustrated above, a parsimonuous model must be used for the selection bias function. A further appeal of the approach is that it may be used to perform a sensitivity analysis, in settings where the ignorability violation arises due to a confounder of the mediator-outcome relationship that is also an effect of the exposure variable; in which case, as observed in

Section 2, such a variable even when observed, cannot be used towards identification of natural direct and indirect effects without additional assumptions.

Finally, we note that while in this section, the support of M was finite, the proposed sensitivity analysis methodology can be extended to accomodate a continuous mediator by further adapting the approach of Robins et al (1999) to the present setting.

5 Discussion

The current paper makes a number of contributions to the study of statistical methods for causal mediation analysis. Focusing on survival data, we have proposed a number of new estimators of natural direct and indirect effects for the Cox PH and the additive hazards models. The weighted approach developed in section 3 is appealing for its simplicity and because it is easy to implement in existing software, provided individual-specific weights are accomodated. We should note that, whereas it is common practice when estimating total effects via inverse-probability-weights, to report conservative standard errors based on the sandwich variance formula, that ignores the first stage estimation of the treatment weights, results by Tchetgen Tchetgen and Shpitser (2011a) imply that such a practice gives the wrong answer for natural direct and indirect effects. For this reason, we recommend the bootstrap for inference. We also note that, in general, the more involved multiply robust approach of Section

3 should be preferred to the simpler weighted approach on theoretical grounds, because the former delivers valid inferences under weaker assumptions than the latter. However, implementing these improved methods for routine application presents a significant challenge that we plan to take on elsewhere.

References

Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of path-specific effects. In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pp. 357–363.

Bang H, Robins J. (2005). Doubly robust estimation in Missing data and causal inference models. Biometrics, 61:692-972.

van der Laan MJ, Robins JM. (2003). Unified Methods for Censored Longitudinal Data and Causality. Springer Verlag: New York.

van der Laan, M, Petersen,M.(2005) Direct Effect Models. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 187. http://www.bepress.com/ucbbiostat/paper187

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

Imai, K, Keele L and Tingley D. (2010b). “A General Approach to Causal Mediation Analysis.” Psychological Methods, Vol. 15, No. 4 (December), pp. 309-334.

Lin DY, Ying Z. 1994. Semiparametric analysis of the additive risk model. Biometrika, 81:61-71.

Newey W. 1994. Semiparametric efficiency bounds. Journal of Applied Econometric 5 (2), 99-135.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), San Francisco, CA, pp. 411–42. Morgan Kaufmann.

Pearl J (2011) The Mediation Formula: A guide to the assessment of causal pathways in nonlinear models.Technical report <http://ftp.cs.ucla.edu/pub/stat_ser/r379.pdf>

Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143–155.

Robins JM, Rotnitzky A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. AIDS Epidemiology - Methodological Issues. Eds: Jewell N., Dietz K., Farewell V. Boston, MA: Birkhäuser. pp. 297-331

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

Robins JM. (1998). Marginal structural models. 1997 Proceedings of the American Statistical Association. Section on Bayesian Statistical Science, pp. 1-10. Reproduced courtesy of the American Statistical Association.

Robins, J. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson (Eds.), Highly Structured Stochastic Systems, pp. 70–81. Oxford, UK: Oxford University Press.

Robins JM, Richardson, TS. (2010). Alternative graphical causal models and the identification of direct effects. To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. P. Shrout, Editor. Oxford University Press

Scharfstein DO, Rotnitzky A, Robins JM. (1999). Rejoinder to comments on

"Adjusting for non-ignorable drop-out using semiparametric non-response models". Journal of the American Statistical Association, 94:1096-1120. Journal of the American Statistical Association, 94:1121-1146.

Tchetgen Tchetgen EJ, Shpitser I. Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness and sensitivity analysis. (2011b) Submitted

Tchetgen Tchetgen EJ, Shpitser I. Semiparametric Estimation of Models for Natural Direct and Indirect Effects. (2011a) Submitted

Tsiatis AA (2006) Semiparametric Theory and Missing Data. Springer.Verlag: New York.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology, 20:18-26.

APPENDIX

PROOF OF THEOREM 1: Under the consistency, sequential ignorability and positivity assumptions,

$$\mathbb{E}\left\{D^*(t)\,\frac{f_{M|E,X}(M|E=0,X)}{f_{E|X}(E|X)f_{M|E,X}(M|E,X)}h(E)\exp\left(\beta_c^{dir}E\right)\right\}$$

$$=\mathbb{E}\left\{S_{T|E,M,X}(t|E,M,X)\,S_{C|E}(t|E)\,\frac{f_{M|E,X}(M|E=0,X)}{f_{E|X}(E|X)f_{M|E,X}(M|E,X)}h(E)\exp\left(\beta_c^{dir}E\right)\right\}$$

$$=\mathbb{E}\left\{\begin{array}{c}\sum_e\int S_{T|E,M,X}(t|E=e,M=m,X)\,S_{C|E}(t|E=e)\\ \times f_{M|E,X}(m|E=0,X)\,h(e)\exp\left(\beta_c^{dir}e\right)d\mu(m)\end{array}\right\}$$

$$=\left\{\begin{array}{c}\sum_e S_{C|E}(t|E=e)\,h(e)\exp\left(\beta_c^{dir}e\right)\\ \times\mathbb{E}\left[\int S_{T|E,M,X}(t|E=e,M=m,X)\,f_{M|E,X}(m|E=0,X)\,d\mu(m)\right]\end{array}\right\}$$

$$=\left\{\sum_e S_{C|E}(t|E=e)\,h(e)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}$$

$$=\left\{\sum_e S_{C|E}(t|E=e)\,h(e)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}$$

and $\mathbb{E}\left\{dN^*(t)Wh(E)\right\}$

$$=\mathbb{E}\left\{\lambda_{T|E,M,X}(t|E,M,X)\,S_{T|E,M,X}(t|e,M,X)\,S_{C|E}(t|E)\,Wh(E)\,dt\right\}$$

$$=\mathbb{E}\left\{f_{T|E,M,X}(t|E,M,X)\,S_{C|E}(t|E)\,Wh(E)\right\}dt$$

$$=\sum_e\left[\begin{array}{c}S_{C|E}(t|E=e)\,h(e)\,dt\\ \times\mathbb{E}\left\{\int f_{T|E,M,X}(t|E=e,M=m,X)\,f_{M|E,X}(m|E=0,X)\,d\mu(m)\right\}\end{array}\right]$$

$$=\sum_e f_{T_{e,M_0}}(t)\,S_{C|E}(t|E=e)\,h(e)\,dt$$

$$=\sum_e S_{C|E}(t|E=e)\,h(e)\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\,dt$$

Therefore $\mathbb{E}\left\{\int dN^*(t)W\left[E-\frac{\xi_1\left(t;\beta_c^{dir}\right)}{\xi_2\left(t;\beta_c^{dir}\right)}\right]\right\}$

$$=\mathbb{E}\left\{\int dN^*(t)W\left[E-\frac{\mathbb{E}\left\{D^*(t)\frac{f_{M|E,X}(M|E=0,X)}{f_{E|X}(E|X)f_{M|E,X}(M|E,X)}h_1(E)\exp\left(\beta_c^{dir}E\right)\right\}}{\mathbb{E}\left\{D^*(t)\frac{f_{M|E,X}(M|E=0,X)}{f_{E|X}(E|X)f_{M|E,X}(M|E,X)}h_2(E)\exp\left(\beta_c^{dir}E\right)\right\}}\right]\right\}$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\int dN^*(t)W\left[E - \frac{\left\{\sum_e S_{C|E}(t|E=e)e\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}{\left\{\sum_e S_{C|E}(t|E=e)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}\right]\right\} \\
&= \int\left[\begin{array}{c} \mathbb{E}\left\{dN^*(t)Wh_1(E)\right\} \\ -\frac{\mathbb{E}\{dN^*(t)Wh_2(E)\}\left\{\sum_e S_{C|E}(t|E=e)e\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}{\left\{\sum_e S_{C|E}(t|E=e)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}\end{array}\right] \\
&= \int dt\left[\begin{array}{c} \sum_e S_{C|E}(t|E=e)\,e\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t) \\ -\frac{\left\{\sum_e S_{C|E}(t|E=e)\;\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}\left\{\sum_e S_{C|E}(t|E=e)e\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}{\left\{\sum_e S_{C|E}(t|E=e)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}\end{array}\right] \\
&= \int dt\left[\begin{array}{c} \sum_e S_{C|E}(t|E=e)\,e\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t) \\ -\frac{\left\{\sum_e S_{C|E}(t|E=e)\;\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}\left\{\sum_e S_{C|E}(t|E=e)e\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}{\left\{\sum_e S_{C|E}(t|E=e)\lambda_{T_0}(t)\exp\left(\beta_c^{dir}e\right)S_{T_{e,M_0}}(t)\right\}}\end{array}\right] \\
&= 0
\end{aligned}
$$

The following lemma will be used repeatedly to establish multiple robustness of a given estimating function.

LEMMA A.1 Given i.i.d data (O, M, E, X), define the weighted functional $\beta(l)$ with weight $L = l(E)$ as:

$$
\kappa(l) = \sum_{e=0}^{1} L(e)\mathbb{E}\left\{\int \mathbb{E}(O|M=m, E=e, X)\, f_{M|E,X}(m|E=0, X)\, d\mu(m)\right\}
$$

Let $B(m, e, x) = \mathbb{E}(O|M=m, E=e, X=x)$. Then, the random variable $J = J(B, f_{M|E,X}, f_{E|X})$ satisfies the triply robust unbiasedness property

$\mathbb{E}\left\{J(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X})\right\} = \kappa(l)$

if at least one but not necessarily all of the following conditions hold: either $\{B^{\ddagger}, f^{\ddagger}_{M|E,X}\} = \{B, f_{M|E,X}\}$ or $\{B^{\ddagger}, f^{\ddagger}_{E|X}\} = \{B, f_{E|X}\}$, or $\{f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}\} = \{f_{M|E,X}, f_{E|X}\}$; where

$$J(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}) = \frac{f^{\ddagger}_{M|E,X}(M|E=0,X)}{f^{\ddagger}_{E|X}(E|X)f^{\ddagger}_{M|E,X}(M|E,X)} \left\{O - B^{\ddagger}(M,E,X)\right\} L(E)$$

$$+ \sum_{e=0}^{1} \int L(e) B^{\ddagger}(M,e,X) f^{\ddagger}_{M|E,X}(m|E=0,X)\, d\mu(m)$$

$$+ \frac{I(E=0)}{f^{\ddagger}_{E|X}(0|X)} \left\{ \sum_{e=0}^{1} L(e) \left[\begin{array}{c} B^{\ddagger}(M,e,X) \\ -\int B^{\ddagger}(M,e,X) f^{\ddagger}_{M|E,X}(m|E=0,X)\, d\mu(m) \end{array} \right] \right\}$$

PROOF OF LEMMA A.1: The bias of $J(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X})$ can be expressed

$$\mathbb{E}\left\{J(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X})\right\} - \kappa(l)$$

$$= \mathbb{E}\left[\sum_{e=0}^{1} \int \frac{f_{E|X}(e|X) f_{M|E,X}(m|E=e,X) f^{\ddagger}_{M|E,X}(m|E=0,X)}{f^{\ddagger}_{E|X}(e|X) f^{\ddagger}_{M|E,X}(m|E=e,X)} \left\{B(m,e,X) - B^{\ddagger}(m,e,X)\right\} L(e) d\mu(m)\right.$$

$$+ \sum_{e=0}^{1} \int L(e) B^{\ddagger}(M,e,X) f^{\ddagger}_{M|E,X}(m|E=0,X)\, d\mu(m)$$

$$- \sum_{e=0}^{1} \int L(e) B(M,e,X) f_{M|E,X}(m|E=0,X)\, d\mu(m)$$

$$\left. + \frac{f_{E|X}(0|X)}{f^{\ddagger}_{E|X}(0|X)} \left\{ \sum_{e=0}^{1} L(e) \left[\begin{array}{c} \int B^{\ddagger}(m,e,X) f_{M|E,X}(m|E=0,X)\, d\mu(m) \\ -\int B^{\ddagger}(m,e,X) f^{\ddagger}_{M|E,X}(m|E=0,X)\, d\mu(m) \end{array} \right] \right\} \right]$$

$$= \mathbb{E}\left[\sum_{e=0}^{1} \int \left\{ \begin{array}{c} \frac{f_{E|X}(e|X) f_{M|E,X}(m|E=e,X)}{f^{\ddagger}_{E|X}(e|X) f^{\ddagger}_{M|E,X}(m|E=e,X)} \\ -1 \end{array} \right\} \left\{ \begin{array}{c} B(m,e,X) \\ -B^{\ddagger}(m,e,X) \end{array} \right\} f^{\ddagger}_{M|E,X}(m|E=0,X) L(e)\right.$$

$$- \sum_{e=0}^{1} \int \left\{ \begin{array}{c} f_{M|E,X}(m|E=0,X) \\ -f^{\ddagger}_{M|E,X}(m|E=0,X) \end{array} \right\} \left\{ \begin{array}{c} B(m,e,X) \\ -B^{\ddagger}(m,e,X) \end{array} \right\} L(e) d\mu(m)$$

$$\left. + \sum_{e=0}^{1} \int \left\{ \begin{array}{c} f_{M|E,X}(m|E=0,X) \\ -f^{\ddagger}_{M|E,X}(m|E=0,X) \end{array} \right\} \left\{ \frac{f_{E|X}(e|X)}{f^{\ddagger}_{E|X}(e|X)} - 1 \right\} B(m,e,X) L(e) d\mu(m) \right]$$

$= 0$ if at least one of the three conditions of the Lemma holds, proving the result.

PROOF OF THEOREM 2: Under the consistency, sequential ignorability and positivity assumptions, in the proof of Theorem 1 we showed

$$\xi_1\left(t;\beta_c^{dir}\right)=\left\{\begin{array}{c}\sum_e h(e)\exp\left(\beta_c^{dir}e\right)\\ \times\mathbb{E}\left[\begin{array}{c}\int S_{C|E}\left(t|E=e\right)S_{T|E,M,X}\left(t|E=e,M=m,X\right)\\ f_{M|E,X}\left(m|E=0,X\right)d\mu\left(m\right)\end{array}\right]\end{array}\right\}$$

which is of the form $\kappa\left(l\right)$, with $L(e)=h(e)\exp\left(\beta_c^{dir}e\right)$ and $O=D^*\left(t\right)$. Therefore, by Lemma 1, $R^{\ddagger}(t,H;\beta_c^{dir})$ has the desired triply robust unbiasedness property, such that $L(e)=S_{C|E}\left(t|E=e\right)h(e)\exp\beta_c^{dir}e$, thus we have $\mathbb{E}\left\{R^{\ddagger}(t,H;\beta_c^{dir})\right\}=\xi_1\left(t;\beta_c^{dir}\right)$ under the conditions of the Theorem. Similarly, we have previously established in the proof of Theorem 1, that

$$\mathbb{E}\left\{dN^*(t)Wh\left(E\right)\right\}=\sum_e\left[\begin{array}{c}h\left(e\right)\\ \times\mathbb{E}\left\{\begin{array}{c}\int S_{C|E}\left(t|E=e\right)f_{T|E,M,X}\left(t|E=e,M=m,X\right)dt\\ f_{M|E,X}\left(m|E=0,X\right)d\mu\left(m\right)\end{array}\right\}\end{array}\right]$$

which is of the form $\kappa\left(l\right)$, with $L(e)=h(e)$ and $O=dN^*(t)$.Therefore, by Lemma 1, the theorem holds upon setting $h(E)=\left\{E-\frac{\xi_1^{mr,\ddagger}\left(t;\beta_c^{dir}\right)}{\xi_2^{mr,\ddagger}(t)}\right\}$.

The following Lemma is key to proving Theorem 1

LEMMA A.2 Define the weighted functional $\sigma\left(l\right)$ with weight $L=l(E)$ as:

$$\begin{aligned}\sigma\left(l\right) &= \sum_{e=0}^{1}L(e)\mathbb{E}\left\{\int\mathbb{E}\left(O|\ E=e,X\right)\ d\mu\left(m\right)\right\}\\ &= \sum_{e=0}^{1}L(e)\mathbb{E}\left\{\int\mathbb{E}\left(O|M=m,E=e,X\right)f_{M|E,X}\left(m|E=e,X\right)d\mu\left(m\right)\right\}\end{aligned}$$

The random variable $A=A(B,f_{M|E,X},f_{E|X})$ satisfies the double robust unbiasedness property which states that $\mathbb{E}\left\{A(B^{\ddagger},f_{M|E,X}^{\ddagger},f_{E|X}^{\ddagger})\right\}=\sigma\left(l\right)$ if at least one but not necessarily both of the following conditions hold: either $\{B^{\ddagger},f_{M|E,X}^{\ddagger}\}=\{B,f_{M|E,X}\}$

or $f^{\ddagger}_{E|X} = f_{E|X}$; where

$$A(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}) = \frac{1}{f^{\ddagger}_{E|X}(E|X)} \left\{ O - \int B^{\ddagger}(m,E,X) f^{\ddagger}_{M|E,X}\left(m|E,X\right) d\mu\left(m\right) \right\} L(E)$$
$$+ \sum_{e=0}^{1} \int L(e) B^{\ddagger}(M,e,X) f^{\ddagger}_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right).$$

PROOF OF LEMMA A.2: $\mathbb{E}\left\{ A(B^{\ddagger}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}) \right\} - \sigma\left(l\right) =$

$$\sum_{e=0}^{1} \mathbb{E} \left\{ \begin{array}{c} \frac{f_{E|X}(e|X)}{f^{\ddagger}_{E|X}(e|X)} \left\{ \begin{array}{c} \int B(m,e,X) f_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right) \\ - \int B^{\ddagger}(m,e,X) f^{\ddagger}_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right) \end{array} \right\} L(e) \\ + \sum_{e=0}^{1} \int L(e) B^{\ddagger}(M,e,X) f^{\ddagger}_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right) \\ - \int B(m,e,X) f_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right) \end{array} \right\}$$

$$= \sum_{e=0}^{1} \mathbb{E} \left\{ \left\{ \frac{f_{E|X}(e|X)}{f^{\ddagger}_{E|X}(e|X)} - 1 \right\} \left\{ \begin{array}{c} \int B(m,e,X) f_{M|E,X}\left(m|E=e,X\right) d\mu\left(m\right) \\ - \int B^{\ddagger}(m,e,X) f^{\ddagger}_{M|E,X}\left(m|e,X\right) d\mu\left(m\right) \end{array} \right\} L(e) \right\}$$

$= 0$ under the assumptions of the theorem.

PROOF OF THEOREM 3: We note that $\vartheta^{mr}_{j}\left(t; \beta^{dir}_{c}, \beta^{ind}_{c}\right)$

$$= \vartheta_{j}\left(t; \beta^{dir}_{c}, \beta^{ind}_{c}\right) = \vartheta_{j}\left(t; \beta^{dir}_{c}, \beta^{ind}_{c}, S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right)$$
$$= \mathbb{E}\left[\sum_{e} \int \left\{ \begin{array}{c} S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e, M=m, X) \\ f_{M|E,X}\left(m|E=e,X\right) h(e) \exp\left\{ \left(\beta^{dir}_{c} + \beta^{ind}_{c}\right) e \right\} d\mu\left(m\right) \end{array} \right\} \right]$$

is of the form of the weighted functional $\sigma\left(l\right)$ with $L(e) = h(e) \exp\left\{ \left(\beta^{dir}_{c} + \beta^{ind}_{c}\right) e \right\}$ and $O = D^{*}\left(t\right)$ thus by Lemma A.2,

$\mathbb{E}\left[G(t, H_{j}; \beta^{dir}_{c}, \beta^{ind}_{c}, S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}) \right] = \vartheta_{j}\left(t; \beta^{dir}_{c}, \beta^{ind}_{c}\right)$ if either $f^{\ddagger}_{E|X} = f_{E|X}$ or $\left\{ S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X} \right\} = \left\{ S_{T|E,M,X}, f_{M|E,X} \right\}$ but not necessarily both. Furthermore, note that

$$= \mathbb{E}\left\{ \int \sum_{e\in\{0,1\}} \left[\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M\ ,X) f_{M|E,X}\left(m|E=e,X\right) \\ \times \left\{ e - \frac{\vartheta_1^{mr}(t;\beta_1,\beta_2)}{\vartheta_2^{mr}(t;\beta_1,\beta_2)} \right\} dt \end{array} \right] d\mu\left(m\right) \right\}$$

is of the form of the weighted functional $\sigma\left(l\right)$ with $L(e) = \left\{ e - \frac{\vartheta_1^{mr}(t;\beta_1,\beta_2)}{\vartheta_2^{mr}(t;\beta_1,\beta_2)} \right\} dt$ and $O = dN^*\left(t\right)$, therefore, by Lemma A.2, if either $f_{E|X}^{\ddagger} = f_{E|X}$ or $\left\{ S_{T|E,M,X}^{\ddagger}, f_{M|E,X}^{\ddagger} \right\} = \left\{ S_{T|E,M,X}, f_{M|E,X} \right\}$

$$\mathbb{E}\left\{ V^{mr}(\beta_c^{dir}, \beta_c^{ind}; S_{T|E,M,X}^{\ddagger}, f_{M|E,X}^{\ddagger}, f_{E|X}^{\ddagger}, S_{C|E}) \right\}$$

$$= \mathbb{E}\left\{ \int\int \sum_{e\in\{0,1\}} \left[\begin{array}{c} S_{C|E}\left(t|E=e\right) f_{T|E,M,X}(t|E=e,M\ ,X) f_{M|E,X}\left(m|E=e,X\right) \\ \times \left\{ e - \frac{\vartheta_1\left(t;\beta_c^{dir},\beta_c^{ind}\right)}{\vartheta_2(t;\beta_1,\beta_2)} \right\} dt \end{array} \right] d\mu\left(m\right) \right\}$$

$$= \mathbb{E}\left\{ \int \sum_{e\in\{0,1\}} \left[\begin{array}{c} S_{C|E}\left(t|E=e\right) S_{T_e}\left(t\right) \lambda_{T_0}\left(t\right) \exp\left(\beta_c e\right) \\ \times \left\{ e - \frac{\sum_{e\in\{0,1\}} S_{C|E}(t|E=e) e S_{T_e}(t) \lambda_{T_0}(t) \exp(\beta_c e)}{\sum_{e\in\{0,1\}} S_{C|E}(t|E=e) S_{T_e}(t) \lambda_{T_0}(t) \exp(\beta_c e)} \right\} dt \end{array} \right] \right\} = 0$$

The result then follows by noting that β_c^{dir} solves $\mathbb{E}\left\{ U^{mr}\left(\beta_c^{dir}; S_{T|E,M,X}^{\ddagger}, f_{M|E,X}^{\ddagger}, f_{E|X}^{\ddagger}, S_{C|E} \right) \right\} = 0$ which is triply robust by Theorem 2.

PROOF OF THEOREM 4: It is straightforward to verify that $\varpi_j\left(t\right)$ is of the form of $\kappa(l)$ with $L(e) = h_j(e)$ and $O = D^*\left(t\right)$. Thus,

$$\varpi_1\left(t\right) = \sum_{e=0}^{1} e\mathbb{E}\left\{ \int S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M\ ,X) f_{M|E,X}\left(m|E=0,X\right) d\mu\left(m\right) \right\}$$

$$\varpi_2\left(t\right) = \sum_{e=0}^{1} \mathbb{E}\left\{ \int S_{C|E}\left(t|E=e\right) S_{T|E,M,X}(t|E=e,M\ ,X) f_{M|E,X}\left(m|E=0,X\right) d\mu\left(m\right) \right\}$$

Under the assumed structural model, and the consistency, sequential ignorability and positivity assumptions,

$$\mathbb{E}\left[\left\{ dN^*(t) - E\beta_a^{dir} D^*\left(t\right) dt \right\} W h_j(E) \right]$$

$$= \sum_{e=0}^{1} h_j(e)\lambda_{T_0}(t)dt\mathbb{E}\left\{\begin{array}{c}\int S_{C|E}\left(t|E=e\right)S_{T|E,M,X}(t|E=e,M\ \ ,X)\\ \times f_{M|E,X}\left(m|E=0,X\right)d\mu\left(m\right)\end{array}\right\}$$

proving the result.

PROOF OF THEOREM 5: The proof is similar to that of Theorem 2, by applying Lemma A.1 to the three functionals $\varpi_1\left(t\right) = \varpi_1^{mr}, \varpi_2\left(t\right) = \varpi_2^{mr}\left(t\right)$ and $\mathbb{E}\left[\left\{dN^*(t) - E\beta_a^{dir}D^*\left(t\right)dt\right\}Wh_j(E)\right]$.

PROOF OF THEOREM 6: The proof is similar to that of Theorem 3, by applying Lemma A.2 to the four key functionals:

$$\phi_j^{mr}\left(t\right) = \sum_{e\in\{0,1\}} h_j\left(e\right)\mathbb{E}\left\{\begin{array}{c}\int S_{C|E}\left(t|E=e\right)S_{T|E,M,X}(t|E=e,M=m,X)\\ \times f_{M|E,X}\left(m|E=e,X\right)d\mu\left(m\right)\end{array}\right\}, j = 1,2$$

and thus $L(e) = h_j(e)$ respectively and $O = D^*\left(t\right)$;

furthermore, $\mathbb{E}\left[\left\{dN^*(t) - E\left(\beta_a^{dir} + \beta_a^{ind}\right)D^*\left(t\right)dt\right\}f_{E|X}^{-1}\left(E|X\right)\left\{E - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}\right]$

$$= \mathbb{E}\left[\left\{dN^*(t) - E\beta_a\ D^*\left(t\right)dt\right\}f_{E|X}^{-1}\left(E|X\right)\left\{E - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}\right]$$

$$= \int \sum_{e\in\{0,1\}}\left\{e - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}\left[\begin{array}{c}\left\{S_{C|E}\left(t|E=e\right)f_{T|E,M,X}(t|E=e,M=m,X)\right\}\\ \times f_{M|E,X}\left(m|E=e,X\right)d\mu\left(m\right)\end{array}\right]$$

$$-\int \sum_{e\in\{0,1\}} e\left(\beta_a^{dir} + \beta_a^{ind}\right)\ \left\{e - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}\left[\begin{array}{c}S_{C|E}\left(t|E=e\right)S_{T|E,M,X}(t|E=e,M=m,X)\\ \times f_{M|E,X}\left(m|E=e,X\right)d\mu\left(m\right)\end{array}\right]$$

is a difference of two $\sigma\left(l\right)-$functionals with respectively $L(e) = \left\{e - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}$ and $O = dN^*\left(t\right)$ for the first functional, and $L(e) = e\left(\beta_a^{dir} + \beta_a^{ind}\right)\ \left\{e - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\}$ and $O = D^*\left(t\right)$ for the second functional. This implies that if if either $f_{E|X}^{\ddagger} = f_{E|X}$ or

$\left\{S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}\right\} = \left\{S_{T|E,M,X}, f_{M|E,X}\right\}$

$\mathbb{E}\left\{P^{mr}(\beta_a^{dir}, \beta_a^{ind}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E})\right\}$

$= \int \lambda_{T_0}(t) \sum_{e=0}^{1} S_{C|E}(t|E=e) S_{T_e}(t) \left\{e - \frac{\phi_1^{mr}(t)}{\phi_2^{mr}(t)}\right\} dt$

$= \int \lambda_{T_0}(t) \sum_{e=0}^{1} S_{C|E}(t|E=e) S_{T_e}(t) \left\{e - \frac{\sum_{e=0}^{1} e S_{C|E}(t|E=e) S_{T_e}(t)}{\sum_{e=0}^{1} S_{C|E}(t|E=e) S_{T_e}(t)}\right\} dt$

$= 0$ and therefore $P^{mr}(\beta_a^{dir}, \beta_a^{ind}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E})$ is a doubly robust estimating function for $\beta_a = \beta_a^{dir} + \beta_a^{ind}$.

The result then follows by noting that $\beta_a^{ind} = \beta_a - \beta_a^{dir}$ and β_a^{dir} solves

$\mathbb{E}\left\{Z^{mr}\left(\beta_a^{dir}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}\right)\right\} = 0$

which is triply robust by Theorem 5, and thus, β_a^{ind} solves

$\mathbb{E}\left\{P^{mr}(\beta_a^{dir}, \beta_a^{ind}; S^{\ddagger}_{T|E,M,X}, f^{\ddagger}_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E})\right\} = 0$ provided one of the three conditions given in the theorem hold.

PROOF OF LEMMA 1: We observe that

$S_{T_{1,m}|E,M,X}(t|E=e, X=x) =$

$S_{T_{1,m}|E,M,X}(t|E=e, M=m, X=x) f_{M|E,X}(m|E=e, X=x)$

$+ S_{T_{1,m}|E,M,X}(t|E=e, M\neq m, X=x)\left\{1 - f_{M|E,X}(m|E=e, X=x)\right\}$

$= \exp\left\{-\int_0^t \left(\lambda_{T_{1,m}|E,M,X}(u|E=e, M=m, X=x)\right) du\right\} f_{M|E,X}(m|E=e, X=x)$

$+ \exp\left\{-\int_0^t \left(\lambda_{T_{1,m}|E,M,X}(u|E=e, M\neq m, X=x) du\right)\right\}\left\{1 - f_{M|E,X}(m|E=e, X=x)\right\}$

$= \exp\left\{-\int_0^t \left(\lambda_{T_{1,m}|E,M,X}(u|E=e, M=m, X=x) du\right)\right\}$

$$\times \left[\begin{array}{c} f_{M|E,X}(m|E=e, X=x) \\ + \exp\left\{\int_0^t \gamma(u,e,m,x) du\right\}\left\{1 - f_{M|E,X}(m|E=e, X=x)\right\} \end{array}\right]$$

Thus, by ignorability of E, we obtain

$$\exp\left\{-\int_0^t \left(\lambda_{T_{1,m|E,M,X}}\left(u|E=0,M=m,X=x\right)du\right)\right\}$$

$$=\exp\left\{-\int_0^t \left(\lambda_{T_{1,m|E,M,X}}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\times\frac{\left[f_{M|E,X}(m|E=1,X=x)+\exp\left\{\int_0^t \gamma(u,1,m,x)du\right\}\left\{1-f_{M|E,X}(m|E=1,X=x)\right\}\right]}{\left[f_{M|E,X}(m|E=0,X=x)+\exp\left\{\int_0^t \gamma(u,0,m,x)du\right\}\left\{1-f_{M|E,X}(m|E=0,X=x)\right\}\right]}$$

$$=\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\times\frac{\left[f_{M|E,X}(m|E=1,X=x)+\exp\left\{\int_0^t \gamma(u,1,m,x)du\right\}\left\{1-f_{M|E,X}(m|E=1,X=x)\right\}\right]}{\left[f_{M|E,X}(m|E=0,X=x)+\exp\left\{\int_0^t \gamma(u,0,m,x)du\right\}\left\{1-f_{M|E,X}(m|E=0,X=x)\right\}\right]}$$

proving the first result by consistency.

Furthermore, by differentiating with respect to t :

$$-\lambda_{T_{1,m|E,M,X}}\left(t|E=0,M=m,X=x\right)\exp\left\{-\int_0^t \left(\lambda_{T_{1,m|E,M,X}}\left(u|E=0,M=m,X=x\right)du\right)\right\}$$

$$=-\lambda_{T|E,M,X}\left(t|E=1,M=m,X=x\right)\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\times\delta\left(t,e,m,x\right)+\dot{\delta}\left(t,e,m,x\right)\times\delta\left(t,e,m,x\right)\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\Leftrightarrow -\lambda_{T_{1,m|E,M,X}}\left(t|E=0,M=m,X=x\right)\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\times\delta\left(t,e,m,x\right)=-\lambda_{T|E,M,X}\left(t|E=1,M=m,X=x\right)$$

$$\times\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\times\delta\left(t,e,m,x\right)+\dot{\delta}\left(t,e,m,x\right)\times\delta\left(t,e,m,x\right)\exp\left\{-\int_0^t \left(\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)du\right)\right\}$$

$$\Leftrightarrow \lambda_{T_{1,m|E,M,X}}\left(t|E=0,M=m,X=x\right)$$

$$=\lambda_{T|E,M,X}\left(u|E=1,M=m,X=x\right)-\dot{\delta}\left(t,e,m,x\right)$$

proving the second part of the Lemma.

PROOF OF THEOREM 7 By Lemma 1 and the assumptions of the theorem,

$$\mathbb{E}\left[\delta_{\alpha^*}(t,E,M,X)\left\{dN^*(t)-\dot{\delta}_{\alpha^*}(t,E,M,X)\,D^*(t)\,dt\right\}Wh_j(E)\right]$$

$$=\mathbb{E}\left[\left\{\lambda_{T|E,M,X}(t|E,M,X)-\dot{\delta}_{\alpha^*}(t,E,M,X)\right\}\delta_{\alpha^*}(t,E,M,X)\,D^*(t)\,Wh_j(E)dt\right]$$

$$=\mathbb{E}\left[\begin{array}{c}\left\{\lambda_{T|E,M,X}(t|E,M,X)-\dot{\delta}_{\alpha^*}(t,E,M,X)\right\}\\ \times S_{C|E}(t|E)\,S_{T|E,M,X}(t|E,M,X)\delta_{\alpha^*}(t,E,M,X)\,Wh_j(E)dt\end{array}\right]$$

$$=\sum_{e\in\{0,1\}}\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\left\{\begin{array}{c}\lambda_{T|E,M,X}(t|e,m,X)\\ -\dot{\delta}_{\alpha^*}(t,e,M,X)\end{array}\right\}\left\{\begin{array}{c}S_{T|E,M,X}(t|e,m,X)\\ \times\delta_{\alpha^*}(t,e,M,X)\end{array}\right\}\\ f_{M|E,X}(m|E=0,X)\ \ h_j(e)dt\end{array}\right]$$

$$=\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\left\{\begin{array}{c}\lambda_{T|E,M,X}(t|1,m,X)\\ -\dot{\delta}_{\alpha^*}(t,1,M,X)\end{array}\right\}\left\{\begin{array}{c}S_{T|E,M,X}(t|1,m,X)\\ \times\delta_{\alpha^*}(t,1,M,X)\end{array}\right\}\\ f_{M|E,X}(m|E=0,X)\,d\mu(m)\,h_j(1)dt\end{array}\right]+$$

$$\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\left\{\lambda_{T|E,M,X}(t|0,m,X)\right\}\left\{S_{T|E,M,X}(t|0,m,X)\right\}\\ f_{M|E,X}(m|E=0,X)\ \ h_j(0)dt\end{array}\right]$$

$$=\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\lambda_{T_{1,M_0}|M_0,X}(t|M_0=m,X)\left\{S_{T_{1,M_0}|M_0,X}(t|M_0=m,X)\right\}\\ f_{M|E,X}(m|E=0,X)\ \ h_j(1)dt\end{array}\right]+$$

$$\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\lambda_{T_{0,M_0}|M_0,X}(t|M_0=m,X)\left\{S_{T_{0,M_0}|M_0,X}(t|_{M_0=m,X})\right\}\\ f_{M|E,X}(m|E=0,X)\ \ h_j(0)dt\end{array}\right]$$

$$=\sum_{e\in\{0,1\}}S_{C|E}(t|e)\,\lambda_{T_{e,M_0}}(t)\,S_{T_{e,M_0}}(t)\,h_j(e)dt$$

$$=\sum_{e\in\{0,1\}}S_{C|E}(t|e)\,\lambda_{T_{0,M_0}}(t)\exp\left(\beta_c^{dir}\right)S_{T_{e,M_0}}(t)\,h_j(e)dt$$

One can similarly show that

$$\mathbb{E}\left\{D^*(t)\,Wh_j(E)\delta_{\alpha^*}(t,E,M,X)\exp\left(\beta_c^{dir}E\right)\right\}$$

$$= \sum_{e\in\{0,1\}} S_{C|E}(t|e)\exp\left(e\beta_c^{dir}\right)S_{T_{e,M_0}}(t)\,h_j(e)dt$$

which implies the result since

$$\mathbb{E}\left\{U^w\left(\beta_c^{dir},\alpha^*\right)\right\}$$

$$= \int \sum_{e\in\{0,1\}} S_{C|E}(t|e)\,\lambda_{T_{0,M_0}}(t)\exp\left(\beta_c^{dir}\right)S_{T_{e,M_0}}(t)$$

$$\times\left\{e - \frac{\sum_{e'\in\{0,1\}} S_{C|E}(t|e')\exp\left(e'\beta_c^{dir}\right)S_{T_{e',M_0}}(t)e'}{\sum_{e''\in\{0,1\}}\int S_{C|E}(t|e'')\exp\left(e''\beta_c^{dir}\right)S_{T_{e'',M_0}}(t)dt}\right\}dt$$

$$= 0$$

For the case of an additive structural model

$$\mathbb{E}\left[\int W\left\{dN^*(t) - \dot{\delta}_{\alpha^*}(t,E,M,X)\,D^*(t)\,dt - E\beta_a^{dir}D^*(t)\,dt\right\}\delta_{\alpha^*}(t,E,M,X)\,h_j(E)\right]$$

$$= \mathbb{E}\left[\left\{\lambda_{T|E,M,X}(t|E,M,X) - \dot{\delta}_{\alpha^*}(t,E,M,X) - E\beta_a^{dir}\right\}\delta_{\alpha^*}(t,E,M,X)\,D^*(t)\,Wh_j(E)dt\right]$$

$$= \mathbb{E}\left[\left\{\lambda_{T|E,M,X}(t|E,M,X) - \dot{\delta}_{\alpha^*}(t,E,M,X) - E\beta_a^{dir}\right\}\delta_{\alpha^*}(t,E,M,X)\,D^*(t)\,Wh_j(E)dt\right]$$

$$= \mathbb{E}\left[\begin{array}{c}\left\{\lambda_{T|E,M,X}(t|E,M,X) - \dot{\delta}_{\alpha^*}(t,E,M,X) - E\beta_a^{dir}\right\}\\ \times S_{C|E}(t|E)\,S_{T|E,M,X}(t|E,M,X)\delta_{\alpha^*}(t,E,M,X)\,Wh_j(E)dt\end{array}\right]$$

$$= \sum_{e\in\{0,1\}}\sum_{m\in\mathcal{S}}\int S_{C|E}(t|e)\,\mathbb{E}\left[\begin{array}{c}\left\{\begin{array}{c}\lambda_{T|E,M,X}(t|e,m,X)\\ -\dot{\delta}_{\alpha^*}(t,e,M,X) - e\beta_a^{dir}\end{array}\right\}\left\{\begin{array}{c}S_{T|E,M,X}(t|e,m,X)\\ \times\delta_{\alpha^*}(t,e,M,X)\end{array}\right\}\\ \times f_{M|E,X}(m|E=0,X)\,h_j(e)dt\end{array}\right]$$

$$= \sum_{m\in\mathcal{S}}\int S_{C|E}(t|0)\,\mathbb{E}\left[\begin{array}{c}\lambda_{T|E,M,X}(t|0,m,X)\left\{S_{T|E,M,X}(t|0,m,X)\right\}\\ \times f_{M|E,X}(m|E=0,X)\ \ h_j(0)dt\end{array}\right]$$

$$+\sum_{m\in\mathcal{S}}\int S_{C|E}(t|1)\,\mathbb{E}\left[\begin{array}{c}\left\{\begin{array}{c}\lambda_{T|E,M,X}(t|1,m,X)\\ -\dot{\delta}_{\alpha^*}(t,1,M,X)-\beta_a^{dir}\end{array}\right\}\left\{\begin{array}{c}S_{T|E,M,X}(t|1,m,X)\\ \times\delta_{\alpha^*}(t,1,M,X)\end{array}\right\}\\ \times f_{M|E,X}(m|E=0,X)\ \ h_j(1)dt\end{array}\right]$$

$$=\sum_{m\in\mathcal{S}}\int S_{C|E}(t|0)\,\mathbb{E}\left[\begin{array}{c}\lambda_{T_{0,M_0}|M_0,X}(t|M_0=m,X)\,S_{T_{0,M_0}|M_0,X}(t|M_0=m,X)\\ \times f_{M|E,X}(m|E=0,X)\,d\mu(m)\,h_j(0)dt\end{array}\right]$$

$$+\sum_{m\in\mathcal{S}}\int S_{C|E}(t|1)\,\mathbb{E}\left[\begin{array}{c}\left\{\lambda_{T_{1,M_0}|M_0,X}(t|M_0=m,X)-\beta_a^{dir}\right\}\left\{S_{T_{1,M_0}|M_0,X}(t|M_0=m,X)\right\}\\ \times f_{M|E,X}(m|E=0,X)\ \ h_j(1)dt\end{array}\right]$$

$$=\ S_{C|E}(t|0)\left[\lambda_{T_{0,M_0}}(t)\,S_{T_{0,M_0}}(t)\,h_j(0)dt\right]$$

$$+\int S_{C|E}(t|1)\left\{\lambda_{T_{1,M_0}}(t)-\beta_a^{dir}\right\}S_{T_{1,M_0}}(t)\,h_j(1)dt$$

$$=\lambda_{T_{0,M_0}}(t)\left\{S_{C|E}(t|0)\,S_{T_{0,M_0}}(t)\,h_j(0)dt\right.$$

$$\left.+\int S_{C|E}(t|1)\,S_{T_{1,M_0}}(t)\,h_j(1)dt\right\}$$

$$=\lambda_{T_{0,M_0}}(t)\sum_{e\in\{0,1\}}S_{C|E}(t|e)\,S_{T_{e,M_0}}(t)\,h_j(e)dt$$

One can similarly show that

$$\mathbb{E}\left\{D^*(t)\,Wh_j(E)\delta_{\alpha^*}(t,E,M,X)\right\}=\sum_{e\in\{0,1\}}S_{C|E}(t|e)\,S_{T_{e,M_0}}(t)\,h_j(e)dt$$

which gives the result.

DOUBLY ROBUST SENSITIVITY ANALYSIS FOR SURVIVAL SEMI-PARAMETRIC REGRESSION MODELS:

We propose a sensitivity analysis that is doubly robust under the Cox PH model. For each fixed $\alpha=\alpha^*$, consider the following modified estimating function for β_c^{dir} :

$$U^{w,dr}\left(\beta_c^{dir},\alpha^*\right)=U^{w,dr}\left(\beta_c^{dir},\alpha^*;S_{T|E,M,X},f_{M|E,X},f_{E|X},S_{C|E}\right)$$

$$= \int \left\{ \delta_{\alpha^*}(t,E,M,X) \left\{ \begin{array}{c} dN^*(t) \\ -\dot{\delta}_{\alpha^*}(t,E,M,X)\, D^*(t)\, dt \\ -E\beta_a^{dir} D^*(t)\, \dot{\delta}_{\alpha^*}(t,E,M,X)\, dt \end{array} \right\} W \times \left\{ E - \frac{\chi_1\left(t;\beta_c^{dir},\alpha^*\right)}{\chi_2\left(t;\beta_c^{dir},\alpha^*\right)} \right\} \right.$$

$$\left. -\delta_{\alpha^*}(t,E,M,X)\, S_{C|E}(t|E) \left\{ \begin{array}{c} f_{T|E,M,X}(t|E,M,X)dt \\ -\dot{\delta}_{\alpha^*}(t,E,M,X)\, S_{T|E,M,X}(t|E,M,X)dt \\ -E\beta_a^{dir} S_{T|E,M,X}(t|E,M,X)dt\dot{\delta}_{\alpha^*}(t,E,M,X)\, dt \end{array} \right\} W \right\} \times$$

$$\left\{ E - \frac{\chi_1\left(t;\beta_c^{dir},\alpha^*\right)}{\chi_2\left(t;\beta_c^{dir},\alpha^*\right)} \right\}$$

$$+ \sum_{m\in\mathcal{S}} \sum_{e\in\{0,1\}} \delta_{\alpha^*}(t,e,m,X)\, S_{C|E}(t|e) \left\{ \begin{array}{c} f_{T|E,M,X}(t|e,m,X)dt \\ -\dot{\delta}_{\alpha^*}(t,e,m,X)\, S_{T|E,M,X}(t|e,m\ ,X)dt \\ -e\beta_a^{dir}\dot{\delta}_{\alpha^*}(t,e,m,X)\, S_{T|E,M,X}(t|e,m\ ,X)dt \end{array} \right\}$$

$$\times \left\{ e - \frac{\chi_1\left(t;\beta_c^{dir},\alpha^*\right)}{\chi_2\left(t;\beta_c^{dir},\alpha^*\right)} \right\} f_{M|E,X}(m|E=0,X)$$

with

$\chi_j\left(t;\beta_c^{dir},\alpha^*\right) =$

$\mathbb{E}\left\{ \left(D^*(t) - S_{C|E}(t|E)\, S_{T|E,M,X}(t|E,M,X)dt\right) W h_j(E)\delta_{\alpha^*}(t,E,M,X)\exp\left(\beta_c^{dir}E\right)\right\}$

$+\sum_{m\in\mathcal{S}}\sum_{e\in\{0,1\}}\left(\ S_{C|E}(t|e)\, S_{T|E,M,X}(t|e,m,X)dt\right)$

$\times f_{M|E,X}(m|E=0,X)\ \ h_j(e)\delta_{\alpha^*}(t,e,m,X)\exp\left(\beta_c^{dir}e\right)$

One can then easily verify that $U^{w,dr}\left(\beta_c^{dir},\alpha^*;S_{T|E,M,X},f_{M|E,X},f_{E|X},S_{C|E}\right)$ is doubly robust in the sense that

$\mathbb{E}\left\{U^{w,dr}\left(\beta_c^{dir},\alpha^*;S^{\ddagger}_{T|E,M,X},f_{M|E,X},f^{\ddagger}_{E|X},S_{C|E}\right)\right\}=0$ if either $S^{\ddagger}_{T|E,M,X}=S_{T|E,M,X}$ or $f^{\ddagger}_{E|X}=f_{E|X}$.

For the additive hazards model, we propose to use the following modified estimating function of β_a^{dir}:

$$Z^{w,dr}\left(\beta_a^{dir},\alpha^*\right) = Z^{w,dr}\left(\beta_a^{dir},\alpha^*; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right)$$

$$= \int \left\{ \delta_{\alpha^*}\left(t,E,M,X\right) \left\{ \begin{array}{c} dN^*(t) \\ -\dot{\delta}_{\alpha^*}\left(t,E,M,X\right) D^*\left(t\right) dt \end{array} \right\} W \times \left\{E - \frac{\zeta_1(t,\alpha^*)}{\zeta_2(t;\alpha^*)}\right\} \right.$$

$$\left. -\delta_{\alpha^*}\left(t,E,M,X\right) S_{C|E}\left(t|E\right) \left\{ \begin{array}{c} f_{T|E,M,X}(t|E,M,X)dt \\ -\dot{\delta}_{\alpha^*}\left(t,E,M,X\right) S_{T|E,M,X}(t|E,M,X)dt \end{array} \right\} W \right\} \times$$

$$\left\{E - \frac{\zeta_1(t,\alpha^*)}{\zeta_2(t;\alpha^*)}\right\}$$

$$+\sum_{m\in\mathcal{S}}\sum_{e\in\{0,1\}} \delta_{\alpha^*}\left(t,e,m,X\right) S_{C|E}\left(t|e\right) \left\{ \begin{array}{c} f_{T|E,M,X}(t|e,m,X)dt \\ -\dot{\delta}_{\alpha^*}\left(t,e,m,X\right) S_{T|E,M,X}(t|e,m\ ,X)dt \end{array} \right\}$$

$$\times \left\{e - \frac{\zeta_1(t,\alpha^*)}{\zeta_2(t;\alpha^*)}\right\} f_{M|E,X}\left(m|E=0,X\right)$$

with

$$\zeta_j\left(t,\alpha^*\right) =$$

$$\mathbb{E}\left\{\left(D^*\left(t\right) - S_{C|E}\left(t|E\right) S_{T|E,M,X}(t|E,M,X)dt\right) W h_j(E)\delta_{\alpha^*}\left(t,E,M,X\right)\right\}$$

$$+\sum_{m\in\mathcal{S}}\sum_{e\in\{0,1\}}\left(\ S_{C|E}\left(t|e\right) S_{T|E,M,X}(t|e,m,X)dt\right) f_{M|E,X}\left(m|E=0,X\right)\ h_j(e)\delta_{\alpha^*}\left(t,e,m,X\right)$$

One can easily verify that $Z^{w,dr}\left(\beta_a^{dir},\alpha^*; S_{T|E,M,X}, f_{M|E,X}, f_{E|X}, S_{C|E}\right)$ is doubly robust in the sense that

$$\mathbb{E}\left\{Z^{w,dr}\left(\beta_a^{dir},\alpha^*; S^{\ddagger}_{T|E,M,X}, f_{M|E,X}, f^{\ddagger}_{E|X}, S_{C|E}\right)\right\} = 0 \text{ if either } S^{\ddagger}_{T|E,M,X} = S_{T|E,M,X}$$

or $f^{\ddagger}_{E|X} = f_{E|X}$.