Harvard University
Harvard University Biostatistics Working Paper Series

Year 2012 *Paper* 141

On Parametrization, Robustness and Sensitivity Analysis in a Marginal Structural Cox Proportional Hazards Model for Point Exposure

Eric J. Tchetgen Tchetgen* James M. Robins†

*Harvard University, etchetge@hsph.harvard.edu
†Harvard School of Public Health, robins@hsph.harvard.edu

http://biostats.bepress.com/harvardbiostat/paper141

On parametrization, robustness and sensitivity analysis in a marginal structural Cox proportional hazards model for point exposure

Eric J. Tchetgen Tchetgen and James Robins
Departments of Epidemiology and Biostatistics, Harvard University

February 11, 2012

Abstract

In this paper, some new statistical methods are proposed, for making inferences about the parameter indexing a Cox proportional hazards marginal structural model for point exposure. Under the key assumption that unmeasured confounding is absent, we propose a new class of closed-form estimators that are doubly robust in the sense that they remain consistent and asymptotically normal for the effect of treatment provided the marginal structural model is correctly specified and, at least one of the following holds: (i) a model for the treatment assignment mechanism is correctly specified or, (ii) a model for part of the observed data likelihood not involving the treatment assignment mechanism is correctly specified. In order to ensure that condition (ii) provides a genuine opportunity for valid inference, we propose a new parametrization of the observed data law, that is congenial with the marginal proportional hazards assumption. In addition, because the assumption of no unmeasured confounding can seldom be established with certainty with observational data, a second contribution of the current paper is to propose a general framework for estimation without the assumption of no unmeasured confounding. For this purpose, a sensitivity analysis technique is developed, that allows an investigator to assess, under model (i), the extent to which unmeasured confounding may alter inferences about causal effects.

The current article concerns the development of improved statistical methods for making inferences about the parameter indexing a Cox proportional hazards marginal structural model for point exposure. Under the key assumption that unmeasured confounding is absent, inverse-probability-of treatment-weighted estimation and augmented inverse-probability-of treatment-weighted estimation have previously been described, to obtain consistent and asymptotically normal estimators for this model. Unfortunately, estimators obtained using these previous methods rely on the crucial assumption that the treatment mechanism is consistently estimated. Furthermore, the assumption that there is no unmeasured confounding may be inappropriate, if the investigator fails to collect at least one key risk factor which predicts treatment assignment. In observational studies, investigators tend to collect and adjust for a large number of confounders, precisely to minimize the presence of unmeasured confounding. As a result, whether succesful in their effort to reduce unmeasured confounding or not, the curse of dimensionality implies that for good finite sample performance, parametric or semi-parametric models must be used to estimate the treatment mechanism. In the event that this latter model is incorrect, the corresponding inferences are likely to be severely biased even when all confounders are observed. As a remedy, here we develop a new class of estimators that are doubly robust when confounding is absent. In a marginal structural model, an estimator is doubly robust if it remains consistent and asymptotically normal for the effect of treatment provided that there is no unmeasured confounding, the marginal structural model is correctly specified and, at least one of the following holds;

i) the model for the treatment assignment mechanism $\mathcal{M}_1$ is correctly specified or,

ii) the model for part of the observed data likelihood not involving the treatment assignment mechanism $\mathcal{M}_2$ is correctly specified.

Doubly robust estimation extends inverse-probability-of treatment weighted estimation of marginal structural models and offers at least two major advantages over the latter which we emphasize. Firstly, as stated above, doubly robust estimation is more robust to model misspecification in the estimated weights used in inverse-probability weighting, and thus, an estimator that is doubly robust, is consistent and asymptotically normal under many more laws than one that is not. Secondly, doubly robust estimation can lead to more efficient estimation than inverse-probability-of-treatment-weighting. Existing literature on the theory of double robustness is too rich to summarize here; but see Sharfstein, Rotnitzky and Robins (1999), Robins (2000), Robins and Rotnizky (2001), van der Laan and Robins (2003) and Tsiatis (2006). van der Laan and Robins (2003). Yu and van der Laan (2003), and Tchetgen Tchetgen (2006) previously considered doubly robust methods for the Cox proportional hazards model in a more general setting which allowed for time-varying exposure and time-varying confounding. The current setting differs in two crucial ways. First, because studies with a point exposure play a pivotal role in several fields, including epidemiology, economics, biostatistics, political science and other social sciences, here we focus on the setting of a point exposure. The time-varying setting will be addressed elsewhere. Second, whereas all previous methods used a pooled logistic regression approximation to the Cox regression model, no such approximation is needed here, and inferences are obtained for a structural regression model in which the proportional hazards assumption is exact. Lastly, our proposed estimators are closed form, which is generally not the case in the time-varying setting, even under a pooled logistic regression approach. We emphasize that the proposed approach is new and does not immediately follow from available results on doubly robust estimation of pooled logistic regression.

Whilst the assumption of no unmeasured confounding may be enforced in an experimental context, mainly by randomizing treatment; there is seldom a guaranty that this assumption holds in an observational study. In addition, because this latter assumption is empirically untestable from nonexperimental data, a second contribution of the current paper is to propose a general framework for estimation without the assumption of no unmeasured confounding. For this purpose, a new sensitivity analysis technique is developed, that allows an investigator to assess, under model $\mathcal{M}_1$, and thus under the assumption that the treatment process is consistently estimated given the observed data, the extent to which unmeasured confounding may alter inferences about causal effects. We emphasize that to the best of our knowledge, there currently exist no sensitivity analysis methodology for unmeasured counfounding under a marginal Cox regression model, therefore the current paper aims to directly address this important gap in the causal inference literature.

1 Notation

We use capital letters to represent random variables and lower-case letters to represent possible realizations (values) of random variables. A_i is the dichotomous treatment variable, $i = 1, \ldots, n$. L_i is a vector of relevant prognostic factors for survival measured prior to treatment. We shall suppress the i subscript denoting individual, because we assume that the random vector for each subject is drawn independently from a distribution common to all subjects. Let T denote the underlying failure time of interest. 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, to focus our exposition, we assume that censoring is independent of (L, A, T), that is

$$C \amalg (L, A, T)$$

although in the appendix, this latter assumption is relaxed, and our methods are extended to produce inferences under the weaker assumption:

$$C \amalg T|A, L$$

The symbol $\amalg$ is used to indicate statistical independence. We assume that there exist counterfactual variables (T_0, T_1) corresponding to the outcome had possibly contrary to fact the treatment taken the value $a = 0, 1$. Finally, we define $\mathbb{P}_n(\cdot) = \sum_{i=1}^{n} (\cdot)_i / n$.

2 The Cox proportional hazards Marginal Structural Model

The Cox structural model of interest is a model for $\{T_a : a \in \{0, 1\}\}$ that assumes a proportional hazards model:

$$\lambda_{T_{\overline{a}}}(t) = \lambda_0(t) \exp(\beta_0 a) \tag{1}$$

and therefore, the parameter $\exp(\beta_0)$ can be interpreted as the causal hazards ratio for the total effect of A, so that $\beta_0 = 0$ encodes the null hypothesis of no treatment effect. Identification of total causal effects requires additional assumptions. To proceed, we make the consistency assumption:

$$\text{if } A = a\,, \text{then } T_a = T \text{ almost surely} \tag{2}$$

In addition, we assume no unmeasured confounding :

$$T_a \amalg A|L,\ a = 0, 1. \tag{3}$$

paired with a standard positivity assumption:

$$\text{if } f_L(L) > 0 \text{ then } f_{A|L}(a|L) > 0,\ \ a = 0, 1. \tag{4}$$

where $f_{A|L}$ is the density of $[A|L]$ and f_L is the density of L. Then, under assumptions (1)-(4), the survival curve of T_a , S_{T_a} is identified by the g-formula of Robins (1998):

$$S_{T_a}(t) = \int S_{T|A,L}(t|A = a, L = l) dF_L(l) = \exp\left\{- \exp(\beta a) \int_0^t \lambda_0(u)\, du\right\} \tag{5}$$

For estimation, Robins (1998) proposed using inverse-probability-of-treatment weighting, which entails finding an estimator $\widehat{\beta}^*_{ipw} = \widehat{\beta}_{ipw}(\widehat{g}^*)$ of β_0 by solving the weighted estimating equation:

$$\begin{aligned} 0 &= \mathbb{P}_n\left\{\widehat{U}^{ph}_{iptw}(\beta; \widehat{g}^*)\right\} = \mathbb{P}_n\left\{\widehat{U}^{ph}_{iptw}\left(\beta; \widehat{g}^*, \widehat{f}_{A|L}\right)\right\} \\ &= \mathbb{P}_n\left(\int \frac{dN^*(t)}{\widehat{f}_{A|L}(A|L)}\left[\widehat{g}^*(A,t) - \frac{\mathbb{P}_n\left\{\frac{\widehat{g}^*(A,t)\exp(\beta A)}{\widehat{f}_{A|L}(A|L)} Y(t)\right\}}{\mathbb{P}_n\left\{\frac{\widehat{g}^*(A,t)\exp(\beta A)}{\widehat{f}_{A|L}(A|L)} Y(t)\right\}}\right]\right) \end{aligned}$$

where $N^*(t) = I(T^* \leq t, \Delta = 1)$ is the counting process of an individual's failure time, $Y(t)$ is a 0-1 predictable process indicating, by the value 1, whether the subject is at risk at time t; $\widehat{f}_{A|L}$ is an estimator of $f_{A|L}$ and $\widehat{g}^*(A, t)$ is a user-specified function of A and t, and may be data dependent. In the following, it is convenient to choose

$$\widehat{g}^*(A, t) = \exp(-\beta A) \left[\widehat{g}(A, t) - \frac{\mathbb{P}_n \left\{ \widehat{g}(A, t) Y(t) \widehat{f}_{A|L}(A|L)^{-1} \right\}}{\mathbb{P}_n \left\{ Y(t) \widehat{f}_{A|L}(A|L)^{-1} \right\}} \right]$$

for a user-specified function $\widehat{g}(A, t)$ of A and t,that may be data dependent; which yields the simple closed-form estimator

$$\widehat{\beta}_{ipw} = \widehat{\beta}_{ipw}(\widehat{g}) = \log \frac{-\mathbb{P}_n \left(\int \frac{I(A=1) dN^*(t)}{\widehat{f}_{A|L}(A|L)} \left[\widehat{g}(A, t) - \frac{\widehat{\xi}_1^{ph}(t; \widehat{g})}{\widehat{\xi}_0^{ph}(t; \widehat{g})} \right] \right)}{\mathbb{P}_n \left(\frac{I(A=0) dN^*(t)}{\widehat{f}_{A|L}(A|L)} \left[\widehat{g}(A, t) - \frac{\widehat{\xi}_1^{ph}(t; \widehat{g})}{\widehat{\xi}_0^{ph}(t; \widehat{g})} \right] \right)}$$

where $\widehat{\xi}_j^{ph}(t; \widehat{g})$ is an estimator of $\xi_j^{ph}(t; g)$, with

$$\xi_j^{ph}(t, g) = \mathbb{E}\left\{ g(A, t)^j f_{A|L}(A|L)^{-1} Y(t) \right\}, \ j = 0, 1$$

and thus

$$\widehat{\xi}_j^{ph}(t, \widehat{g}) = \widehat{\xi}_j^{ph}\left(t, \widehat{g}, \widehat{f}_{A|L}\right) = \mathbb{P}_n \left\{ \widehat{g}(A, t)^j \widehat{f}_{A|L}(A|L)^{-1} Y(t) \right\}, \ j = 0, 1.$$

As previously mentioned, because of the curse of dimensionality due to a high dimensional L, nonparametric estimation will likely be impractical for estimating $f_{A|L}$ at sample sizes encountered in practice, and thus a parametric/semiparametric model $\mathcal{M}_1$ is typically used. We briefly note that the two sets of estimators $\{\widehat{\beta}_{ipw}(g^*) : g^* \text{ unrestricted}\}$ and $\{\widehat{\beta}_{ipw}(g) : g \text{ unrestricted}\}$ are equivalent. Hence, our decision to work with the latter representation is merely for the convenience of having a closed-form estimator. Nonetheless, under either representation, the estimator $\widehat{\beta}_{ipw}$ (as well as $\widehat{\beta}_{ipw}^*$) has two important potential limitations. Firstly, one should be concerned that model mis-specification of $\widehat{f}_{A|L}$ could potentially result in a biased estimate of β_0. Secondly, in the event that $\widehat{f}_{A|L}$ is consistent, $\widehat{\beta}_{ipw}$ is well known to be inefficient under the semiparametric model $\mathcal{M}_1$. The first difficulty is addressed below. To resolve the second difficulty, Robins (1998) proposed to improve efficiency for estimating β_0 under $\mathcal{M}_1$ by substracting from $\widehat{U}_{iptw}^{ph}(\beta; \widehat{g}^*)$, an estimate of its orthogonal projection onto the tangent space for the treatment process in a model for the observed data in which, except for the no unmeasured confounding and positivity assumptions, the latter is nonparametric. The tangent space for the treatment process is given by the closed linear span of scores for the treatment mechanism in the nonparametric model. Let $\mathcal{M}_2$ denote a working model for the conditional survival curve $S_{T|A,L}$, and let $\widehat{S}_{T|A,L}$ be an estimator under this working model. The result of Robins (1998) produces the following estimator:

$$\begin{aligned} \widehat{\beta}_{aug} &= \widehat{\beta}_{aug}(\widehat{g}) \\ &= \log \frac{-\mathbb{P}_n \int \left\{ \frac{I(A=1) dN^*(t)}{\widehat{f}_{A|L}(A|L)} + d\widehat{S}_{T|A,L}(t|1, L) \left\{ \frac{I(A=1)}{\widehat{f}_{A|L}(1|L)} - 1 \right\} \widehat{S}_C(t) \left\{ \widehat{g}(1, t) - \frac{\widehat{\xi}_1^{ph}(t; \widehat{g})}{\widehat{\xi}_0^{ph}(t; \widehat{g})} \right\} \right\}}{\mathbb{P}_n \int \left\{ \left(\frac{I(A=0) dN^*(t)}{\widehat{f}_{A|L}(A|L)} + d\widehat{S}_{T|A,L}(t|0, L) \left\{ \frac{I(A=0)}{\widehat{f}_{A|L}(0|L)} - 1 \right\} \widehat{S}_C(t) \right) \left\{ \widehat{g}(0, t) - \frac{\widehat{\xi}_1^{ph}(t; \widehat{g})}{\widehat{\xi}_0^{ph}(t; \widehat{g})} \right\} \right\}} \end{aligned}$$

with $\widehat{S}_C$ the Kaplan-Meier estimator of the survival function of censoring. A result due to Robins (1998) implies that in the absence of model mis-specification, $\widehat{\beta}_{aug}(\widehat{g})$ is consistent, with large sample variance guaranteed to be no larger than the large sample variance of $\widehat{\beta}_{ipw}(\widehat{g})$. In the event that $\mathcal{M}_1$ is incorrect, $\widehat{f}_{A|L}$ will fail to converge to $f_{A|L}$ due to model mis-specification and $\widehat{\beta}_{aug}(\widehat{g})$ will generally not be consistent, even if $\widehat{S}_{T|A,L}$ is consistent. It is straightforward to see that this is mainly because of model mis-specification $\widehat{\xi}_j^{ph}\left(t;\widehat{\beta}_{aug}\right)$ will also be inconsistent. As a remedy we propose an alternative estimator that is doubly robust; more precisely, we develop an estimator of β_0 that remains consistent in the union model $\mathcal{M}_1 \cup \mathcal{M}_2$. To do so, we proceed by first obtaining an estimator of $\xi_j^{ph}(t;\beta)$ that remains consistent in $\mathcal{M}_1 \cup \mathcal{M}_2$, and subsequently substitute this estimator for $\widehat{\xi}_j^{ph}(t;\cdot)$ in evaluating the augmented estimator given in the display above. To formally state the result, Let

$$
\begin{aligned}
\widehat{\xi}_{j,dr}^{ph}(t;\widehat{g}) &= \mathbb{P}_n\left\{\widehat{g}(A,t)^j \widehat{f}_{A|L}(A|L)^{-1}\left\{Y(t) - \widehat{S}_{T|A,L}(t|A,L)\,\widehat{S}_C(t)\right\}\right\} \\
&\quad + \mathbb{P}_n\left\{\widehat{S}_C(t)\sum_{a=0}^{1}\widehat{S}_{T|A,L}(t|A=a,L)\,\widehat{g}(a,t)^j\right\}
\end{aligned}
$$

and define $\widehat{\beta}_{dr}(\widehat{g})$ as $\widehat{\beta}_{aug}(\widehat{g})$ but with $\widehat{\xi}_{j,dr}^{ph}(t;\widehat{g})$ replacing $\widehat{\xi}_j^{ph}(t;\widehat{g})$.

Theorem 1: Under the consistency, no unmeasured confounding and positivity assumptions, the estimator $\widehat{\beta}_{dr}$ *is consistent in model* $\mathcal{M}_1 \cup \mathcal{M}_2$.

We briefly note that $\widehat{\beta}_{aug}(\widehat{g})$ is invariant to the choice of $\widehat{g}(0,t)$ and thus it is convenient in practice, to set $\widehat{g}(0,t) = 0$. Because $\widehat{\beta}_{dr}$ is regular and asymptotically normal under standard regularity conditions, inferences in the union model $\mathcal{M}_1 \cup \mathcal{M}_2$ are conveniently obtained via a version of the bootstrap.

2.1 A congenial parametrization of $S_{T|A,L}$

In order to hold true, Theorem 1 of the previous section implicitely assumes congeniality between the working model $\mathcal{M}_2$ of $S_{T|A,L}$ with the underlying marginal structural Cox proportional hazards model. This is because, if $\mathcal{M}_1$ is incorrect, then a model $\mathcal{M}_2$ cannot be used to obtain valid inferences about β_0 unless the model agrees with the underlying structural assumption of proportional marginal hazards. As equation (5) reveals, $S_{T|A,L}$ and the marginal Cox model are intimately related, since a model of the former must marginalize to a survival curve that satisfies the proportional hazards restriction. To ensure this property is always satisfied, we first observe that the survival curves $S_{T|A,L}$ and $S_{T_a}(t)$ are related by:

$$
S_{T|A,L}(t|A=a,L=l) = r(t,a,l) \times S_{T_a}(t) \tag{6}
$$

where

$$
\begin{aligned}
r(t,a,l) &= \exp\left\{m(t,a,l) - \overline{m}(t,a)\right\} \\
m(t,a,l) &= \log\frac{S_{T|A,L}(t|A=a,L=l)}{S_{T|A,L}(t|A=a,L=l_0)} \\
\overline{m}(t,a) &= \log\left[\int \exp\left\{m(a,l)\right\} dF_L(l)\right]
\end{aligned}
$$

with l_0 a reference value of L. Thus, a variation independent parametrization is obtained by directly modeling the quantities $\{m(t,a,l), F_L(l), S_{T_a}(\cdot)\}$ under the marginal Cox model. Let $\{m(t,a,l;\eta), F_L(l;\gamma), S_{T_a}(\cdot;\beta,\lambda_0)\}$ denote such a model indexed by variation independent parameters $(\eta,\gamma,\beta,\Lambda_0) \in \Phi \times \Gamma \times \mathbb{R} \times \Omega$, where recall under the marginal structural Cox proportional hazards model, $S_{T_a}(t;\beta,\Lambda_0) = \exp\{-\exp(\beta a)\Lambda_0(t)\}$ with $\Lambda_0 = \Lambda_0(\cdot) = \int_0^{\cdot} \lambda_0(u)\,du$, the unrestricted baseline cumulative hazard function; and $m(t,a,l;\eta)$ is a smooth parametric model with respect to η, that satisfies $m(t,a,l_0;\eta) = m(t,a,l;0) = 0$. For estimation, we note that equation (6) together with our choice of parametrization, implies the following model for the conditional hazard function $\lambda_{T|A,L}$:

$$\lambda_{T|A,L}(t|a,l;\eta,\gamma,\beta,\lambda_0) = \lambda_{T_a}(t;\beta,\lambda_0) + j(t,a,l;\eta,\gamma) = \lambda_0(t)\exp(\beta a) + j(t,a,l;\eta,\gamma) \qquad (7)$$

where $j(t,a,l;\eta,\gamma) = -\frac{\partial \log r(t,a,l;\eta,\gamma)}{\partial t}$ and $\log r(t,a,l;\eta,\gamma) = m(t,a,l;\eta) - \overline{m}(t,a;\eta,\gamma)$. Suppose for the moment that γ is known, then under the Cox MSM, model (7) corresponds to an instance of the so-called additive-multiplicative hazards model of Lin and Ying (1995), who provide a general semiparametric methodology to estimate the unknown parameters (η,β,Λ_0), with Λ_0 modeled nonparametrically. Since γ_0 is unknown, let $\widehat{\nu} = \left(\widehat{\eta}, \widehat{\beta}, \widehat{\Lambda}_0\right)$ denote their estimator which is given in the appendix, and is obtained upon replacing γ_0 with an estimator $\widehat{\gamma}$ conveniently obtained by maximizing the partial log-likelihood $\mathbb{P}_n \log f_L(L;\gamma)$. Finally, we may define $\widehat{S}_{T|A,L}(t|a,l)$ as

$$\widehat{S}_{T|A,L}(t|a,l) = \widehat{r}(t,a,l) \times \widehat{S}_{T_a}(t) = r(t,a,l;\widehat{\eta},\widehat{\gamma}) \times S_{T_a}\left(t;\widehat{\beta},\widehat{\Lambda}_0\right)$$

Because, as we have assumed throughout, L is likely high-dimensional with both discrete and continuous components; it may be difficult, if not impossible, to correctly specify a correct working model $F_L(l;\gamma)$. Hence, to further be robust to model mis-specification, we propose to instead use the empirical estimator $\widetilde{F}_L(l) = \mathbb{P}_n I(L \leq l)$ so that for a fixed value of η, $\exp\{\overline{m}(t,a;\eta)\}$ is now estimated by

$$\int \exp\{m(a,l;\eta)\}\, d\widetilde{F}_L(l) = \mathbb{P}_n \exp\{m(a,L;\eta)\}$$

which is actually more convenient because it obviates the need for numerical integration with respect to a continuous L.

2.2 Sensitivity analysis for unmeasured confounding

In this section, we develop a semiparametric sensitivity analysis technique to assess the extent to which a violation of (3) might alter inferences about β_0. Let

$$\overline{\gamma}(t,a,l) = \lambda_{T_a|A,L}(t|A=a, L=l) - \lambda_{T_a|A,L}(t|A=1-a, L=l)$$

then

$$T_a \not\perp\!\!\!\perp A | L = l$$

i.e. a violation of the no unmeasured confounding assumption, generally implies that $\overline{\gamma}(t,a,l) \neq 0$ for some (t,a,l). Suppose that larger values of T are beneficial for health, then if $\overline{\gamma}(t,1,l) < 0$ but $\overline{\gamma}(t,0,l) > 0$ for all t, then on average, individuals with $\{A=0, L=l\}$ have a higher hazard function for each of the potential outcomes $\{T_1, T_0\}$ than individuals with $\{A=1, L=l\}$; i.e. healthier individuals are more likely to receive the treatment. On the other hand, if $\overline{\gamma}(t,0,l) < 0$

but $\overline{\gamma}(t,1,l) > 0$ for all t, suggests confounding by indication for the mediator variable; i.e. unhealthier individuals are more likely to receive the treatment.

We proceed as in Robins et al (1999) who proposed using a selection bias function for the purposes of conducting a sensitivity analysis for average total effects. Tchetgen Tchetgen and Shpitser (2011a,b) and Tchetgen Tchetgen (2011) extended the approach for assessing the impact of unmeasured confounding on the estimation of natural direct and indirect effects in a causal mediation context. Note that $\overline{\gamma}$ is generally not identified from the observed data, thus, to proceed we propose to recover inferences by assuming the selection bias function $\overline{\gamma}(t,a,l)$ is known, which encodes the magnitude and direction of unmeasured confounding for the mediator. To motivate the proposed approach, suppose for the moment that $f_{A|L}$ is known, we show in the appendix that the following lemma holds:

Lemma 1:Let

$$\begin{aligned}\delta(t,a,l) &= \delta\left(t,a,l,f_{A|L}\right)\\ &= f_{A|L}(A=a|L=l) + \left\{1 - f_{A|L}(A=a|L=l)\right\}\exp\left\{-\int_0^t \overline{\gamma}(u,a,l)\,du\right\}\end{aligned}$$

and

$$\dot{\delta}(t,a,l) = \frac{\partial \log \delta(u,a,l)}{\partial u}|_{u=t}$$

Under the consistency assumption

$$S_{T_a|L}(t|L) = S_{T|A=a,L}(t|A=a,L) \times \delta(t,a,l)$$

Furthermore,

$$\begin{aligned}&\lambda_{T_a|L}(t|L=l)\\ =&\lambda_{T|A,L}(t|A=a,L=l) - \dot{\delta}(t,a,l)\end{aligned}$$

Lemma 1 implies that $S_{T_a}(t)$ equals:

$$\mathbb{E}\left(S_{T|A=a,L}(t|A=a,L) \times \delta(t,a,L)\right) \tag{8}$$

Below, we use this result to obtain a consistent estimator of β_0. A sensitivity analysis is then obtained by repeating this process and by reporting inferences for each choice of $\overline{\gamma}(\cdot,\cdot,\cdot)$ in a set of user–specified functions $\Gamma = \{\overline{\gamma}_\alpha(\cdot,\cdot,\cdot) : \alpha\}$ indexed by a finite dimensional parameter α with $\overline{\gamma}_0(\cdot,\cdot,\cdot) \in \Gamma$ corresponding to the no unmeasured confounding assumption, i.e. $\overline{\gamma}_0(\cdot,\cdot,\cdot) \equiv 0$. Throughout, the working model for $f_{A|L}$ is assumed to be correct. Thus, to implement the sensitivity analysis technique, we develop a semiparametric estimator of β_0 in the model $\mathcal{M}_{1,\alpha^*}$ that assumes the Cox marginal structural model holds, the model for $f_{A|L}$ is correctly specified, and the selection bias function is known, that is $\overline{\gamma}(\cdot,\cdot,\cdot) = \overline{\gamma}_{\alpha^*}(\cdot,\cdot,\cdot)$ for α^* fixed .

For inference, we propose the following modified estimator of β_0, which carefully incorporates the selection bias function:

$$\begin{aligned}&\widehat{\beta}_{aug}(\alpha^*) = \widehat{\beta}_{aug}(\widehat{g},\alpha^*)\\ &= \log \frac{-\mathbb{P}_n \int \left(\left[\frac{I(A=1)\left\{dN^*(t) - \dot{\delta}_{\alpha^*}(t,1,L)Y(t)dt\right\}}{\widehat{f}_{A|L}(A|L)} + d\widehat{S}_{T|A,L}(t|1,L)\left\{\frac{I(A=1)}{\widehat{f}_{A|L}(1|L)} - 1\right\}\widehat{S}_C(t)\right]\delta_{\alpha^*}(t,1,L)\left\{\widehat{g}(1,t) - \frac{\widehat{\xi}_1^{ph}(t;\widehat{g},\alpha^*)}{\widehat{\xi}_0^{ph}\left(\widehat{t;g,\alpha^*}\right)}\right\}\right)}{\mathbb{P}_n \int \left(\left[\frac{I(A=0)\left\{dN^*(t) - \dot{\delta}_{\alpha^*}(t,0,L)D^*(t)dt\right\}}{\widehat{f}_{A|L}(A|L)} + d\widehat{S}_{T|A,L}(t|0,L)\left\{\frac{I(A=0)}{\widehat{f}_{A|L}(0|L)} - 1\right\}\widehat{S}_C(t)\right]\delta_{\alpha^*}(t,0,L)\left\{\widehat{g}(0,t) - \frac{\widehat{\xi}_1^{ph}(t;\widehat{g},\alpha^*)}{\widehat{\xi}_0^{ph}(t;\widehat{g},\alpha^*)}\right\}\right)}\end{aligned}$$

and
$\widehat{\xi}^{ph}_{j,dr}(t;\widehat{g},\alpha^*) = \mathbb{P}_n\left(\widehat{g}(A,t)^j \widehat{f}_{A|L}(A|L)^{-1}\left[Y(t) - \widehat{S}_{T|A,L}(t|A,L)\,\widehat{S}_C(t)\right]\delta_{\alpha^*}(t,A,L)\right)$
$+\mathbb{P}_n\left\{\widehat{S}_C(t)\sum_{a=0}^{1}\widehat{S}_{T|A,L}(t|A=a,L)\,\delta_{\alpha^*}(t,a,L)\,\widehat{g}(a,t)^j\right\}$

Theorem 2:Suppose that model $\mathcal{M}_{1,\alpha^*}$ holds, then under the consistency and positivity assumptions, $\widehat{\beta}_{aug}(\alpha^*)$ converges to β_0.

Thus, under model $\mathcal{M}_{1,\alpha^*}$ a sensitivity analysis entails reporting the set$\left\{\widehat{\beta}(\alpha^*):\alpha\right\}$ (and associated confidence intervals) which summarizes how sensitive inferences are to a deviation from the no unmeasured confounding assumption $\alpha = 0$.

It is helpful for practice, to briefly describe possible functional forms for the selection bias function $\overline{\gamma}_\alpha(\cdot,\cdot,\cdot)$. In general, a single parameter model is attractive because it is most tractable, thus one may use one of the following:

$$\begin{array}{ll} \overline{\gamma}_{\alpha,1}(t,a,l) = \alpha(2a-1) & \overline{\gamma}_{\alpha,2}(t,a,l) = \alpha a \\ \overline{\gamma}_{\alpha,3}(t,a,l) = \alpha t(2a-1) & \overline{\gamma}_{\alpha,4}(t,a,l) = \alpha ta \\ \overline{\gamma}_{\alpha,5}(t,a,l) = \alpha(2a-1)tl_1 & \overline{\gamma}_{\alpha,6}(t,e,m,x) = \alpha atl_1 \end{array}$$

where L_1 is a component of L; and for each of the above functional forms, the scalar parameter α encodes the magnitude and direction of unmeasured confounding for the mediator.

The functions $\overline{\gamma}_{\alpha,1}$ and $\overline{\gamma}_{\alpha,2}$ assume the selection bias is time invariant, whereas $\overline{\gamma}_{\alpha,3}, \overline{\gamma}_{\alpha,4}, \overline{\gamma}_{\alpha,5}$ and $\overline{\gamma}_{\alpha,6}$ model interactions with time and possibly a covariate $L_1,$ thus allowing for heterogeneity in the selection bias function over time. Since the functional form of $\overline{\gamma}_\alpha$ is not identified from the observed data, we generally recommend reporting results for a variety of functional forms.

References

[1] Lin and Ying (1995). Semiparametric analysis of general additive-multiplicative intensity models for counting processes. Annals of Statistics, 23:1712-1734.

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

[3] 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

[4] Robins JM, Rotnitzky A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer" Statistica Sinica, 11(4):920-936.

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

[6] 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.

[7] 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.

[8] Robins JM. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science 1999, pp. 6-10.

[9] Satten G, Datta S, Robins J.(2001) Estimating the marginal survival function in the presence of time dependent covariates.Statistics & Probability Letters. Volume 54, Issue 4, Pages 397-403.

[10] Tchetgen Tchetgen EJ (2006) Statistical methods for robust inference in causal and missing data models. PhD Thesis. Department of Biostatistics. Harvard University.

[11] Tchetgen Tchetgen EJ (2011) On causal mediation analysis with a survival outcome. International Journal of Biostatistics (in press)

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

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

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

[15] Yu and van der Laan (2006)" and change in reference list " Yu Z and van der Laan M (2006) Doubly robust estimation in longitudinal marginal structural models. Journal of Statistical Planning and Inference (2006) Volume: 136, Issue: 3, Pages: 1061-1089.

APPENDIX

PROOF OF THEOREM 1: Let $\left\{f^*_{A|L}, S^*_{T|A,L}\right\}$ denote the probability limit of $\left\{\widehat{f}_{A|L}, \widehat{S}_{T|A,L}\right\}$. First, suppose that $\widehat{f}_{A|L}$ is consistent, then $\widehat{\xi}^{ph}_{j,dr}\left(t;\widehat{g}\right)$ converges to

$$\begin{aligned}
&\mathbb{E}\left\{g(A,t)^j f_{A|L}\left(A|L\right)^{-1}\left\{Y(t) - S^*_{T|A,L}\left(t|A,L\right)S_C(t)\right\}\right\} \\
&+\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S^*_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\} \\
&=\mathbb{E}\left\{g(A,t)^j f_{A|L}\left(A|L\right)^{-1}Y(t)\right\} \\
&=\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\}
\end{aligned}$$

similarly,if $\widehat{S}_{A|L}$ is consistent, then $\widehat{\xi}^{ph}_{j,dr}\left(t;\widehat{g}\right)$ converges to

$$\begin{aligned}
&\mathbb{E}\left\{g(A,t)^j f^*_{A|L}\left(A|L\right)^{-1}\left\{Y(t) - S_{T|A,L}\left(t|A,L\right)S_C(t)\right\}\right\} \\
&+\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\} \\
&=\mathbb{E}\left\{g(A,t)^j f^*_{A|L}\left(A|L\right)^{-1}\left\{\mathbb{E}\left\{Y(t)|A,L\right\} - S_{T|A,L}\left(t|A,L\right)S_C(t)\right\}\right\} \\
&+\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\} \\
&=\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\} \\
&=\mathbb{E}\left\{S_C(t)\sum_{a=0}^{1} S_{T|A,L}\left(t|A=a,L\right)g(a,t)^j\right\}
\end{aligned}$$

therefore in the union model -exp$(\widehat{\beta}_{aug}) = -\exp\left\{\widehat{\beta}_{aug}\left(\widehat{g}\right)\right\}$ converges to

$$\begin{aligned}
&\frac{\mathbb{E}\int\left\{\frac{I(A=1)dN^*(t)}{f^*_{A|L}(A|L)}+dS^*_{T|A,L}(t|1,L)\left\{\frac{I(A=1)}{f^*_{A|L}(1|L)}-1\right\}S_C(t)\left\{g(1,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}\right\}}{\mathbb{E}\int\left\{\left(\frac{I(A=0)dN^*(t)}{f^*_{A|L}(A|L)}+dS^*_{T|A,L}(t|0,L)\left\{\frac{I(A=0)}{f^*_{A|L}(0|L)}-1\right\}S_C(t)\right)\left\{g(0,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}\right\}} \\
&=\frac{\mathbb{E}\int\left[\left\{\frac{f_{A|L}(1|L)}{f^*_{A|L}(1|L)}-1\right\}\left\{dS^*_{T|A,L}(t|1,L)-dS_{T|A,L}(t|1,L)\right\}+dS_{T|A,L}(t|1,L)S_C(t)\right]\left\{g(1,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}}{\mathbb{E}\int\left[\left\{\frac{f_{A|L}(0|L)}{f^*_{A|L}(0|L)}-1\right\}\left\{dS^*_{T|A,L}(t|0,L)-dS_{T|A,L}(t|0,L)\right\}+dS_{T|A,L}(t|0,L)S_C(t)\right]\left\{g(0,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}} \\
&=\frac{\int\left[d\mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}S_C(t)\right]\left\{g(1,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}}{\int\left[d\mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}S_C(t)\right]\left\{g(0,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}} \\
&=\frac{exp(\beta)\int\mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}S_C(t)\left\{g(1,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}dt}{\int\mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}S_C(t)\left\{g(0,t)-\frac{\xi^{ph}_1(t;g)}{\xi^{ph}_0(t;g)}\right\}dt}
\end{aligned}$$

which proves the result because

$$
\frac{\int \mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}S_C(t)\left\{g(1,t)-\frac{\xi_1^{ph}(t;g)}{\xi_0^{ph}(t;g)}\right\}dt}{\int \mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}S_C(t)\left\{g(0,t)-\frac{\xi_1^{ph}(t;g)}{\xi_0^{ph}(t;g)}\right\}dt}
$$

$$
= \frac{\int \mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}S_C(t)\left\{g(1,t)-\frac{\mathbb{E}\left\{S_C(t)\sum_{a=0}^1 S_{T|A,L}(t|A=a,L)g(a,t)\right\}}{\mathbb{E}\left\{S_C(t)\sum_{a=0}^1 S_{T|A,L}(t|A=a,L)\right\}}\right\}dt}{\int \mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}S_C(t)\left\{g(0,t)-\frac{\mathbb{E}\left\{S_C(t)\sum_{a=0}^1 S_{T|A,L}(t|A=a,L)g(a,t)\right\}}{\mathbb{E}\left\{S_C(t)\sum_{a=0}^1 S_{T|A,L}(t|A=a,L)\right\}}\right\}dt}
$$

$$
= \frac{\int\left\{\begin{array}{c} g(1,t)\mathbb{E}\left\{\sum_{a=0}^1 S_{T|A,L}\left(t|A=a,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\} \\ -\mathbb{E}\left\{\sum_{a=0}^1 S_{T|A,L}\left(t|A=a,L\right)g(a,t)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\}\end{array}\right\}dt}{\int\left\{\begin{array}{c} g(0,t)\mathbb{E}\left\{\sum_{a=0}^1 S_{T|A,L}\left(t|A=a,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\} \\ -\mathbb{E}\left\{\sum_{a=0}^1 S_{T|A,L}\left(t|A=a,L\right)g(a,t)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\}\end{array}\right\}dt}
$$

$$
= \frac{\int\left\{\begin{array}{c} \mathbb{E}\left\{S_{T|A,L}\left(t|A=1,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\}g(1,t) \\ +\mathbb{E}\left\{S_{T|A,L}\left(t|A=0,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\}g(1,t) \\ -\mathbb{E}\left\{S_{T|A,L}\left(t|A=1,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\}g(1,t) \\ -\mathbb{E}\left\{S_{T|A,L}\left(t|A=0,L\right)g(0,t)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|1,L\right)\right\}\end{array}\right\}dt}{\int\left\{\begin{array}{c} g(0,t)\mathbb{E}\left\{S_{T|A,L}\left(t|A=1,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\} \\ +g(0,t)\mathbb{E}\left\{S_{T|A,L}\left(t|A=0,L\right)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\} \\ -\mathbb{E}\left\{S_{T|A,L}\left(t|A=1,L\right)g(1,t)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\} \\ -\mathbb{E}\left\{S_{T|A,L}\left(t|A=0,L\right)g(0,t)\right\}\mathbb{E}\left\{S_{T|A,L}\left(t|0,L\right)\right\}\end{array}\right\}dt}
$$

$$
= \frac{\int\left\{\mathbb{E}\left\{S_{T|A,L}(t|A=0,L)\right\}\mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}g(1,t)-\mathbb{E}\left\{S_{T|A,L}(t|A=0,L)g(0,t)\right\}\mathbb{E}\left\{S_{T|A,L}(t|1,L)\right\}\right\}dt}{\int\left\{g(0,t)\mathbb{E}\left\{S_{T|A,L}(t|A=1,L)\right\}\mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}-\mathbb{E}\left\{S_{T|A,L}(t|A=1,L)g(1,t)\right\}\mathbb{E}\left\{S_{T|A,L}(t|0,L)\right\}\right\}dt}
$$

$$
= -1
$$

□

PROOF OF LEMMA 1: First note that

$$
S_{T_a|L}\left(t|L\right) = f_{A|L}\left(A=a|L=l\right)S_{T_a|A=a,L}\left(t|A=a,L\right)
$$

$$
+\left\{1-f_{A|L}\left(A=a|L=l\right)\right\}S_{T_a|A=a,L}\left(t|A=1-a,L\right)
$$

$$
=\left[\begin{array}{c} f_{A|L}\left(A=a|L=l\right) \\ +\left\{1-f_{A|L}\left(A=a|L=l\right)\right\}S_{T_a|A=a,L}\left(t|A=1-a,L\right)/S_{T_a|A=a,L}\left(t|A=a,L\right)\end{array}\right]S_{T_a|A=a,L}\left(t|A=a,L\right)
$$

$$
=S_{T_a|A=a,L}\left(t|A=a,L\right)\left[f_{A|L}\left(A=a|L=l\right)+\left\{1-f_{A|L}\left(A=a|L=l\right)\right\}\exp\left\{-\int_0^t\overline{\gamma}\left(u,a,l\right)du\right\}\right]
$$

Differentiating wrt t yields:

$$
-\lambda_{T_a|L}\left(t|L\right)S_{T_a|L}\left(t|L\right)
$$

$$
= -S_{T_a|A=a,L}\left(t|A=a,L\right)\lambda_{T_a|A=a,L}\left(t|A=a,L\right)\delta\left(t,a,l\right)
$$

$$
+S_{T_a|A=a,L}\left(t|A=a,L\right)\partial\delta\left(t,a,l\right)/\partial t
$$

which is equivalent to

$$
\lambda_{T_a|L}\left(t|L\right) = \frac{S_{T_a|A=a,L}(t|A=a,L)}{S_{T_a|L}(t|L)}\lambda_{T_a|A=a,L}\left(t|A=a,L\right)\delta\left(t,a,l\right)
$$

$$
-\frac{S_{T_a|A=a,L}(t|A=a,L)}{S_{T_a|L}(t|L)}\partial\delta\left(t,a,l\right)/\partial t
$$

$$
= \lambda_{T_a|A=a,L}\left(t|A=a,L\right) - \frac{\partial\delta(t,a,l)/\partial t}{\delta(t,a,l)}
$$

$$
=\lambda_{T|A=a,L}\left(t|A=a,L\right) - \frac{\partial\delta(t,a,l)/\partial t}{\delta(t,a,l)}
$$

□

PROOF OF THEOREM 2: We begin by observing that by Lemma 1

$$
\mathbb{E}\left\{dN^*(t)-\dot{\delta}_{\alpha^*}\left(t,a,L\right)Y\left(t\right)dt|A=a,L\right\}
$$

$$
= \left\{\lambda_{T_a|A=a,L}\left(t|A=a,L\right)-\dot{\delta}_{\alpha^*}\left(t,1,L\right)\right\}S_C(t)S_{T_a|A=a,L}\left(t|A=a,L\right)\times\delta\left(t,a,l\right)dt
$$

$= \lambda_{T_a|L}\left(t|L\right) S_{T_a|L}\left(t|L\right) S_C(t)dt$

and $\mathbb{E}\left[g(A,t)^j f_{A|L}\left(A|L\right)^{-1}\left\{Y(t) - S^*_{T|A,L}\left(t|A,L\right) S_C(t)\right\}\delta_{\alpha^*}\left(t,A,L\right)\right]$

$+\mathbb{E}\left[S_C(t)\sum_{a=0}^{1} S^*_{T|A,L}\left(t|A=a,L\right)\delta_{\alpha^*}\left(t,a,L\right) g(a,t)^j\right]$

$= \mathbb{E}\left[g(A,t)^j f_{A|L}\left(A|L\right)^{-1} S_{T|A,L}\left(t|A,L\right) S_C(t)\delta_{\alpha^*}\left(t,A,L\right)\right]$

$=\mathbb{E}\left[g(A,t)^j f_{A|L}\left(A|L\right)^{-1} S_{T_a|L}\left(t|L\right) S_C(t)\right]$

The result then holds by Theorem 1.

ESTIMATION APPROACH OF LIN AND YING (2005) OF AN ADDITIVE-MULTIPLICATIVE HAZARDS MODEL

Given an estimator $\widehat{\overline{\gamma}}$ of $\overline{\gamma}_0$, we wish to estimate $\nu = (\eta_0, \beta_0, \Lambda_0)$ under the additive-multiplicative hazards model $\lambda_{T|A,L}\left(t|a,l;\eta,\overline{\gamma},\beta,\lambda_0\right)$ given by by equation (7). The estimator proposed by Lin and Ying (2005) is given by $\widehat{\nu} = \left(\widehat{\eta}, \widehat{\beta}, \widehat{\Lambda}_0\right)$ where $\left(\widehat{\eta}, \widehat{\beta}\right)$ solves the equation

$$\mathbb{P}_n\left[\int\left\{dN^*\left(t\right) - j\left(t,a,l;\widehat{\eta},\widehat{\overline{\gamma}}\right)\ Y(t)\right\}\left\{h\left(A,L,t\right) - \frac{\mathbb{P}_n\left\{h\left(A,L,t\right)Y(t)\exp\left(\beta A\right)\right\}}{\mathbb{P}_n\left\{\ Y(t)\exp\left(\beta A\right)\right\}}\right\}\right] = 0$$

and

$$\widehat{\Lambda}_0\left(t\right) = \frac{\int_0^t \mathbb{P}_n\left\{dN^*\left(u\right) - j\left(u,a,l;\widehat{\eta},\widehat{\overline{\gamma}}\right)\ Y(u)\right\}}{\int_0^t \mathbb{P}_n\left\{\exp\left(\widehat{\beta}A\right)\ Y(u)\right\}}$$

ESTIMATION UNDER THE ASSUMPTION $C \amalg T|A,L$

Under this assumption, let $\widehat{S}_{C|A,L}$ denote a estimate of $S_{C|A,L}$ the survival curve of censoring under a parametric or semiparametric model. To model $S_{C|A,L}$, one may proceed as in Robins and Rotnizky (1992) who use a Cox Proportional hazards model, or alternatively, one may adapt the additive hazards model of Satten et al (2001).Then, redefine

$$\widehat{\xi}^{ph}_{j,dr}\left(t;\widehat{g}\right) = \mathbb{P}_n\left\{\widehat{g}(A,t)^j\left\{\widehat{f}_{A|L}\left(A|L\right)\widehat{S}_{C|A,L}(t|A.L)\right\}^{-1}\left\{Y(t) - \widehat{S}_{T|A,L}\left(t|A,L\right)\right\}\right\}$$
$$+\mathbb{P}_n\left\{\sum_{a=0}^{1}\widehat{S}_{T|A,L}\left(t|A=a,L\right)\widehat{g}(a,t)^j\right\}$$

$$\widehat{\beta}_{aug} = \widehat{\beta}_{aug}\left(\widehat{g}\right)$$
$$= \log\frac{-\mathbb{P}_n\int\left\{\frac{I(A=1)dN^*(t)}{\widehat{f}_{A|L}(A|L)\widehat{S}_{C|A,L}(t|A.L)} + d\widehat{S}_{T|A,L}\left(t|1,L\right)\left\{\frac{I(A=1)}{\widehat{f}_{A|L}(1|L)} - 1\right\}\ \left\{\widehat{g}(1,t) - \frac{\widehat{\xi}^{ph}_1(t;\widehat{g})}{\widehat{\xi}^{ph}_0(t;\widehat{g})}\right\}\right\}}{\mathbb{P}_n\int\left\{\left(\frac{I(A=0)dN^*(t)}{\widehat{f}_{A|L}(A|L)\widehat{S}_{C|A,L}(t|A.L)} + d\widehat{S}_{T|A,L}\left(t|0,L\right)\left\{\frac{I(A=0)}{\widehat{f}_{A|L}(0|L)} - 1\right\}\ \right)\left\{\widehat{g}(0,t) - \frac{\widehat{\xi}^{ph}_1(t;\widehat{g})}{\widehat{\xi}^{ph}_0(t;\widehat{g})}\right\}\right\}}$$

Then, Theorem 1 continues to hold upon redefining $\mathcal{M}_1$ to be the submodel in which models for both $f_{A|L}$ and $S_{C|A,L}$ are correctly specified. Similarly, to Theorem 2 continues to hold upon redefining $\widehat{\beta}_{aug}\left(\alpha^*\right) = \widehat{\beta}_{aug}\left(\widehat{g},\alpha^*\right)$

$$= \log\frac{-\mathbb{P}_n\int\left(\left[\frac{I(A=1)\left\{dN^*(t) - \dot{\delta}_{\alpha^*}(t,1,L)Y(t)dt\right\}}{\widehat{f}_{A|L}(A|L)\widehat{S}_{C|A,L}(t|A.L)} + d\widehat{S}_{T|A,L}(t|1,L)\left\{\frac{I(A=1)}{\widehat{f}_{A|L}(1|L)} - 1\right\}\right]\delta_{\alpha^*}(t,1,L)\left\{\widehat{g}(1,t) - \frac{\widehat{\xi}^{ph}_1(t;\widehat{g},\alpha^*)}{\widehat{\xi}^{ph}_0\left(\widehat{t;g,\alpha^*}\right)}\right\}\right)}{\mathbb{P}_n\int\left(\left[\frac{I(A=0)\left\{dN^*(t) - \dot{\delta}_{\alpha^*}(t,0,L)D^*(t)dt\right\}}{\widehat{f}_{A|L}(A|L)\widehat{S}_{C|A,L}(t|A.L)} + d\widehat{S}_{T|A,L}(t|0,L)\left\{\frac{I(A=0)}{\widehat{f}_{A|L}(0|L)} - 1\right\}\right]\delta_{\alpha^*}(t,0,L)\left\{\widehat{g}(0,t) - \frac{\widehat{\xi}^{ph}_1(t;\widehat{g},\alpha^*)}{\widehat{\xi}^{ph}_0(t;\widehat{g},\alpha^*)}\right\}\right)}$$

and

$$\widehat{\xi}_{j,dr}^{ph}\left(t;\widehat{g},\alpha^{*}\right)=\mathbb{P}_{n}\left(\widehat{g}(A,t)^{j}\left\{\widehat{S}_{C|A,L}(t|A.L)\widehat{f}_{A|L}\left(A|L\right)\right\}^{-1}\left[Y(t)-\widehat{S}_{T|A,L}\left(t|A,L\right)\right]\delta_{\alpha^{*}}\left(t,A,L\right)\right)$$
$$+\mathbb{P}_{n}\left\{\sum_{a=0}^{1}\widehat{S}_{T|A,L}\left(t|A=a,L\right)\delta_{\alpha^{*}}\left(t,a,L\right)\widehat{g}(a,t)^{j}\right\}$$