Estimating restricted mean treatment effects with additive-multiplicative hazards models

Jinhong Li[1], Jicai Liu[2], Yanbo Pei[3], Riquan Zhang[1*]

1. Department of Statistics, East China Normal University, Shanghai, China

2. School of Statistics and Mathematics

Shanghai Lixin University of Accounting and Finance, Shanghai, China

3. School of Statistics, Capital University of Economics and Business, Beijing, China

July 28, 2022

Abstract: The difference in restricted mean survival times between two groups is often of inherent interest in epidemiologic and medical studies. In this paper, we propose a general additive-multiplicative hazards (AMH) regression model to estimate the restricted mean treatment effects, where the survival time is subject to both dependent and independent censoring. The AMH specifies an additive and multiplicative form on the hazard functions for the survival and censored times associated with covariates, which contains the proportional hazards model and the additive hazards model. By an inverse probability censoring weights scheme, we obtain the estimators of the regression parameters and the restricted mean treatment effects. We establish the large sample properties of the proposed estimators. Monte Carlo simulation studies are conducted to examine the finite sample performance of the proposed procedures and the primary biliary cirrhosis patients data is analyzed for illustration.

Key words: Additive-multiplicative hazards model; dependent censoring; inverse probability of censoring weighting; restricted mean survival time; survival analysis.

1 Introduction

In epidemiologic and medical studies, it is of scientific interest to compare effects of different treatments. For example, in the primary biliary cirrhosis (PBC) study (Fleming and Harrington, 2011), researchers are interested in estimating and assessing the treatment effect of D-penicillamine. Estimating treatment effect has received considerable attention

*Corresponding author. E-mail address: zhangriquan@163.com.

in the statistics literature. See, Rubin (1974); Tsiatis (2006); van der Laan and Robins (2003), among others. In the absence of censoring, the treatment effect can be assessed by comparing some specific summary measure, such as mean and quantile. However, when the data are censored, the mean or quantile measure may be infeasible. For example, the mean of the survival time cannot be accurately estimated due to censoring.

Different from the above classical measure, such as mean or quantile, this paper focuses on the effect measure based on the restricted mean survival times (Karrison, 1987; Zucker, 1998). Specifically, the restricted mean of the survival time T is defined by $E\{\min(T, L)\} = \int_0^L P(T > t)dt$, for any positive L. By choosing different L, $E\{\min(T, L)\}$ often provides a clinically meaningful interpretation and is of more interest than $E\{T\}$ in some applications. Estimating the restricted mean treatment effects has been explored in recent literature. For example, Zhang and Schaubel (2011, 2012) estimated the restricted mean treatment effects by fitting treatment-specific Cox hazards models. However, it is well known that the proportionality assumption for Cox models may be violated in some situations. Thus, Zhang and Schaubel (2011, 2012)'s approaches may lead to biased and inefficient estimates of the restricted mean treatment effects in the case. To overcome the limitations, Yang (2013) replaced the Cox model by a short-term and long-term hazards model, which includes the proportional hazards model and the proportional odds model as special cases; Wey et al. (2016) used a stacked survival model, which can effectively estimate the conditional survival function in a wide range of scenarios. Recently, there have also been some works showing that additive models may be more applicable to causal inference than Cox model in some applications, see, for example, Li et al. (2015), Jiang et al. (2018) and Ying et al. (2019).

As a useful alternative to Cox or additive model, the AMH model (Lin and Ying, 1995) is commonly used in the literature of survival analysis. The AMH model enjoys many nice properties. For example, it has a good interpretability, which assumes that some covariates have multiplicative effects and others have additive effects or even allows certain covariates to have both additive and multiplicative effects; it is very flexible, which includes Cox model and additive hazards model, it can be easily estimated by several simple equations, and so on. These nice properties motivate us to estimate the restricted mean treatment effects under the AMH model, which can be viewed as a non-trivial extension to the approaches proposed by Zhang and Schaubel (2011, 2012).

In this paper, we use the AMH model to estimate treatment differences in restricted mean lifetime. Moreover, we assume that the survival time is subject to both dependent and independent censoring, which is commonly encountered in many applications. To handle the dependent censoring, we adopt the classical inverse probability of censoring weighting (IPCW) technique, see Robins and Rotnitzky (1992); Robins (1993); Robins and Finkelstein (2000). The proposed method is very useful in that it enjoys the flexility of the AMH model and moreover allows for group-specific baseline hazards and regression coefficients.

The rest of the paper is organized as follows. In Section 2, we introduce some notations and backgrounds. We further develop the estimate of the treatment differences in restricted mean lifetime. In Section 3, we establish the asymptotic properties of the estimator. The results from numerical studies are reported in Sections 4 and 5. We provide a brief discussion in Section 6. All the technical proofs are deferred to the Appendix.

2 Modeling and Estimation

2.1 Restricted Mean Treatment Effects

Let A be a binary treatment variable with $A = 1$ for treatment and $A = 0$ for control, and denote the survival time by T. In many applications, T may be censored due to various reasons. Here, we consider the following two types of censoring. Specifically, C_1 is a potential censoring time and independent of the survival time T conditional on baseline covariates $\mathbf{H} = (\mathbf{Z}^T, \mathbf{G}^T)^T_{p\times 1}$ and A. In practice, the type of censoring may occur at the end of the study period or when a subject fails to return for a study visit. Let C_2 be another potential censoring time, which may not be independent of T given $\mathbf{H}$ and A.

Denote the observed survival time by $U = T \wedge C_1 \wedge C_2$, and the censoring indicators by $\Delta_1 = I(T \leq C_1 \wedge C_2)$ and $\Delta_2 = I(C_2 \leq T \wedge C_1)$, where $a \wedge b = \min\{a, b\}$ and $I(\cdot)$ is the indicator function. Let $\widetilde{\mathbf{H}}(t) = \{(\mathbf{X}^T(u), \mathbf{W}^T(u))^T_{q\times 1}; u \in [0, t)\}$ be the history of time-dependent covariates up to just before time t. The observed data consist of $O_i = \{A_i, U_i, \Delta_{1i}, \Delta_{2i}, \mathbf{H}_i, \widetilde{\mathbf{H}}_i(U_i)\}$, $i = 1, 2, \ldots, n$, which are independent copies of $\{A, U, \Delta_1, \Delta_2, \mathbf{H}, \widetilde{\mathbf{H}}(U)\}$. In particular, the baseline covariates $\mathbf{H}$ are included in $\widetilde{\mathbf{H}}(U)$. However, for convenience, we still take this presentation.

Let $T^{(0)}$ and $T^{(1)}$ be the counterfactual survival times, which would be observed under control and treatment groups, respectively. In practice, each individual can only receive one treatment strategy. Thus, the observed survival time of subject is $T = T^{(0)}I(A = 0) + T^{(1)}I(A = 1)$. We can define the group-specific difference in restricted mean lifetime as

$$\delta = E\{\min(T^{(1)}, L)\} - E\{\min(T^{(0)}, L)\} = \int_0^L \{P(T^{(1)} > t) - P(T^{(0)} > t)\}dt.$$

By the conventional "no unmeasured confounders" assumption (Zhang and Schaubel, 2011, 2012), $(T^{(0)}, T^{(1)}) \perp\!\!\!\perp A|\mathbf{H}$, where "$\perp\!\!\!\perp$" indicates independence, we have that $P(T^{(0)} > t|A = 0, \mathbf{H}) = P(T > t|A = 0, \mathbf{H}) \triangleq S_0(t|\mathbf{H})$ and $P(T^{(1)} > t|A = 1, \mathbf{H}) = P(T > t|A = 1, \mathbf{H}) \triangleq S_1(t|\mathbf{H})$. These yield that

$$\begin{aligned} \delta &= \int_0^L E_{\mathbf{H}}\Big\{P(T^{(1)} > t|A = 1, \mathbf{H}) - P(T^{(0)} > t|A = 0, \mathbf{H})\Big\}dt \\ &= \int_0^L E_{\mathbf{H}}\{S_1(t|\mathbf{H}) - S_0(t|\mathbf{H})\}dt, \end{aligned} \tag{1}$$

where $E_{\mathbf{H}}\{\cdot\}$ is taken in respect to the marginal distribution of $\mathbf{H}$.

2.2 Estimation

By plugging the empirical estimates of $S_0(t|\mathbf{H})$ and $S_1(t|\mathbf{H})$ into (1), we can obtain an estimate of δ. However, a fully nonparametric estimation of $S_0(t|\mathbf{H})$ or $S_1(t|\mathbf{H})$ is often infeasible due to the curse of dimensionality when p is moderate or high. A commonly used strategy is to model the conditional survival function of T on A and $\mathbf{H}$.

In the paper, we focus on the following an AMH model for the survival time T, given by

$$\lambda_{ij}(t) = \lambda(t|\mathbf{H}_i, A_i = j) = \lambda_{0j}(t)h(\eta_j^T\mathbf{Z}_i) + g(\alpha_j^T\mathbf{G}_i), \quad j = 0, 1, \tag{2}$$

where $\lambda_{0j}(t)$ is an unknown baseline hazard function for $A = j$, $g(\cdot)$ and $h(\cdot)$ are two specified link functions, $\theta_j = (\eta_j^T, \alpha_j^T)^T$ is a p-vector of unknown regression parameters. In general, we take the link functions $h(x) = \exp(x)$ and $g(x) = x$.

Because there exists the dependent censoring time C_2, the classical estimating method proposed by Lin and Ying (1995) may obtain inconsistent estimators of $\lambda_{0j}(t)$ and θ_j, which has been shown in Zhang and Schaubel (2011) for the Cox model. To overcome the issues, according to Zhang and Schaubel (2011), we establish a hazard function for dependent censoring C_2 to estimate the dependent censoring probability. Assume that the dependence of C_{2i} and T_i occurs through the observed time-dependent process, $\widetilde{\mathbf{H}}_i(U_i)$. That is, we assume that C_{2i} is conditionally independent of T_i given $\{\mathbf{H}_i, A_i, \widetilde{\mathbf{H}}_i(U_i)\}$, which implies that

$$\begin{aligned} &\lim_{h\to 0} h^{-1}P\{t \le U_i \le t + h, \Delta_{2i} = 1 \mid U_i \ge t, A_i, \widetilde{\mathbf{H}}_i(t), T_i\} \\ =&\lim_{h\to 0} h^{-1}P\{t \le U_i \le t + h, \Delta_{2i} = 1 \mid U_i \ge t, A_i, \widetilde{\mathbf{H}}_i(t)\}. \end{aligned} \tag{3}$$

Under the assumption, we further establish an AMH model for the dependent censoring time C_2 as follows

$$\lambda_{ij}^C(t) = \lambda^C(t|\mathbf{H}_i(t), A_i = j) = \lambda_{0j}^C(t)h(\gamma_j^T\mathbf{X}_i(t)) + g(\beta_j^T\mathbf{W}_i(t)), \quad j = 0, 1, \tag{4}$$

where λ_{0j}^C is the unspecified baseline hazard function of C_2 for $A = j$, and $\theta_j^C = (\gamma_j^T, \beta_j^T)^T$ is a q-vector of unknown regression coefficients.

We next consider the estimation of the unknown quantities. By treating C_2 as a survival time and $T \wedge C_1$ as an independent censoring time, similar to Lin and Ying (1995), we define the observed-failure counting $N_{ij}^C(t) = I(U_i \le t, \Delta_{2i} = 1, A = j)$, the at-risk indicator $Y_{ij} = I(U_i \ge t, A = j)$, and the counting process martingale

$$dM_{ij}^C(t) = dN_{ij}^C(t) - Y_{ij}(t)[g(\beta_j^T\mathbf{W}_i(t))dt + h(\gamma_j^T\mathbf{X}_i(t))d\Lambda_{0j}^C(t)],$$

where $\Lambda_{0j}^C(t) = \int_0^t \lambda_{0j}^C(u)du$. Then, we obtain a Breslow-type estimator of $\Lambda_{0j}^C(t)$, given by

$$\hat{\Lambda}_{0j}^C(t) = \int_0^t \frac{\sum_{i=1}^n [dN_{ij}^C(s) - Y_{ij}(s)g(\beta_j^T\mathbf{W}_i(s))ds]}{\sum_{i=1}^n Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))}, \tag{5}$$

and an estimating equation of $\theta_j^C = (\gamma_j^T, \beta_j^T)^T$, given by

$$U(\tau; \theta_j^C) = \sum_{i=1}^{n} \int_0^\tau \{D_i(t; \theta_j^C) - \bar{D}_j(t; \theta_j^C)\}\{dN_{ij}^C(t) - Y_{ij}(t) g(\beta_j^T \mathbf{W}_i(t)) dt\}, \tag{6}$$

where τ is the maximum follow-up time and

$$\begin{aligned} D_i(t; \theta_j^C) &= \begin{bmatrix} \widetilde{\mathbf{X}}_i(t; \theta_j^C) \\ \widetilde{\mathbf{W}}_i(t; \theta_j^C) \end{bmatrix} = \begin{bmatrix} h'(\gamma_j^T \mathbf{X}_i(t)) \mathbf{X}_i(t) / h(\gamma_j^T \mathbf{X}_i(t)) \\ g'(\beta_j^T \mathbf{W}_i(t)) \mathbf{W}_i(t) / h(\gamma_j^T \mathbf{X}_i(t)) \end{bmatrix}, \\ \bar{D}_j(t; \theta_j^C) &= \frac{\sum_{i=1}^n Y_{ij}(t) h(\gamma_j^T \mathbf{X}_i(t)) D_i(t; \theta_j^C)}{\sum_{i=1}^n Y_{ij}(t) h(\gamma_j^T \mathbf{X}_i(t))}, \end{aligned}$$

where $g'(\cdot)$ and $h'(\cdot)$ are the first derivatives of $g(\cdot)$ and $h(\cdot)$, respectively.

To eliminate the influence of dependent censoring C_2 on the estimation of model (2), we can use the inverse probability censoring weights (IPCW) trick as that in Zhang and Schaubel (2011). Specifically, let $N_{ij}(t) = I(U_i \leq t, \Delta_{1i} = 1, A = j)$ and $dM_{ij}(t) = dN_{ij}(t) - Y_{ij}(t)[g(\alpha_j^T \mathbf{G}_i)dt + h(\eta_j^T \mathbf{Z}_i) d\Lambda_{0j}(t)]$. Similar to (5) and (6), we obtain the IPCW versions of a Breslow-type estimator of $\Lambda_{0j}(t)$ and an estimating equation of θ_j, given by

$$\hat{\Lambda}_{0j}(t) = \int_0^t \frac{\sum_{i=1}^n \rho_{ij}(s)[dN_{ij}(s) - Y_{ij}(s) g(\alpha_j^T \mathbf{G}_i) ds]}{\sum_{i=1}^n \rho_{ij}(s) Y_{ij}(s) h(\eta_j^T \mathbf{Z}_i)}, \tag{7}$$

$$U(\tau; \theta_j) = \sum_{i=1}^{n} \int_0^\tau \rho_{ij}(t)\{D_i(\theta_j) - \bar{D}_j(t; \theta_j)\}\{dN_{ij}(t) - Y_{ij}(t) g(\alpha_j^T \mathbf{G}_i) dt\}, \tag{8}$$

where

$$D_i(\theta_j) = \begin{bmatrix} h'(\eta_j^T \mathbf{Z}_i) \mathbf{Z}_i / h(\eta_j^T \mathbf{Z}_i) \\ g'(\alpha_j^T \mathbf{G}_i) \mathbf{G}_i / h(\eta_j^T \mathbf{Z}_i) \end{bmatrix}, \quad \bar{D}_j(t; \theta_j) = \frac{\sum_{i=1}^n \rho_{ij}(t) Y_{ij}(t) h(\eta_j^T \mathbf{Z}_i) D_i(\theta_j)}{\sum_{i=1}^n \rho_{ij}(t) Y_{ij}(t) h(\eta_j^T \mathbf{Z}_i)},$$

and $\rho_{ij}(t)$ is a weighted function, given by $\rho_{ij}(t) = \exp(\Lambda_{ij}^C(t)) \kappa(t; \mathbf{H}_i, A_i)$, where $\Lambda_{ij}^C(t) = \int_0^t \lambda_{ij}^C(u) du$ and $\kappa(t; \mathbf{H}_i, A_i)$ is a specified function, called stabilization factor (Zhang and Schaubel, 2011).

It is important to choose a suitable $\kappa(t; \mathbf{H}_i, A_i)$ to obtain stabilized estimators in (7) and (8). See, Robins and Finkelstein (2000); Hernán et al. (2001); Zhang and Schaubel (2011). As suggested by Robins and Finkelstein (2000); Hernán et al. (2001), we can choose $\kappa(t; \mathbf{H}_i, A_i)$ as $\exp(-\Lambda_{ij}^C(t|\mathbf{H}_i, A_i))$. Additionally, if the censoring is light or moderate, we can use $\kappa(t; \mathbf{H}_i, A_i) = 1$. In our numerical studies, we call the estimator based on $\kappa(t; H_i, A_i) = 1$ the "unstabilized" estimator.

Let $\hat{\theta}_j^C = (\hat{\gamma}_j^T, \hat{\beta}_j^T)^T$ be the root of $U(\tau; \theta_j^C) = 0$. By model (4), $\Lambda_{ij}^C(t)$ can be estimated by

$$\hat{\Lambda}_{ij}^C(t) = \int_0^t h(\hat{\gamma}_j^T \mathbf{X}_i(s)) d\hat{\Lambda}_{0j}^C(s) + \int_0^t g(\hat{\beta}_j^T \mathbf{W}_i(s)) ds. \tag{9}$$

Then, we can estimate the weight function $\rho_{ij}(t)$ by $\hat{\rho}_{ij}(t) = \exp(\hat{\Lambda}_{ij}^C(t))\kappa(t; \mathbf{H}_i, A_i)$. Replacing $\rho_{ij}(t)$ by $\hat{\rho}_{ij}(t)$ in (8), we can obtain the estimator for θ_j, denoted by $\hat{\theta}_j$. By $\hat{\theta}_j$ and $\hat{\Lambda}_{0j}(t)$, the estimator for $S_{ij}(t) \triangleq S_j(t|\mathbf{H}_i)$ can be given by

$$\hat{S}_{ij}(t) \triangleq \hat{S}_j(t|\mathbf{H}_i) = \exp\{-(\hat{\Lambda}_{0j}(t)h(\hat{\eta}_j^T\mathbf{Z}_i) + tg(\hat{\alpha}_j^T\mathbf{G}_i))\}, \quad j = 0, 1. \tag{10}$$

Let $\hat{S}_j(t) = n^{-1}\sum_{i=1}^n \hat{S}_{ij}(t)$ for $j = 0, 1$. By (1), we can estimate the group-specific difference δ by

$$\hat{\delta} = \int_0^L \{\hat{S}_1(t) - \hat{S}_0(t)\}dt. \tag{11}$$

3 Asymptotic properties

In this section, we establish the asymptotic properties of $\hat{\delta}$. To the end, we need to impose the following conditions.

(C1) $O_i = \{A_i, U_i, \Delta_{1i}, \Delta_{2i}, \mathbf{H}_i, \widetilde{\mathbf{H}}_i(U_i)\}$, $i = 1, 2, \ldots, n$, are independently and identically distributed.

(C2) $P(U \geq \tau) > 0$, $\Lambda_{0j}(\tau) < \infty$ and $\Lambda_{0j}^C(\tau) < \infty$, for $j = 0, 1$.

(C3) $|\mathbf{H}_{ik}| + \int_0^\tau d|\mathbf{H}_{ik}(t)| \leq M_H < \infty$, where $\mathbf{H}_{ik}$ and $\mathbf{H}_{ik}(t)$ are the kth elements of $\mathbf{H}_i$ and $\mathbf{H}_i(t)$.

(C4) $g(\cdot)$ and $h(\cdot)$ are continuously differentiable; $\partial D_i(t; \theta_j^C)/\partial\{\theta_j^C\}^T$ and $\partial D_i(\theta_j)/\partial\theta_j^T$ are equicontinuous in a neighborhood of θ_j^C and θ_j.

(C5) The matrices $A_j(\theta_j^C)$ and $A_j(\theta_j)$ are nonsingular, where

$$A_j^C(\theta_j^C) = E\left\{\int_0^\tau Y_{ij}(t)\{D_i(t;\theta_j^C) - \bar{d}_j(t;\theta_j^C)\}\begin{bmatrix} h'(\gamma_j^T\mathbf{X}_i(t))\mathbf{X}_i^T(t)d\Lambda_{0j}^C(t) \\ g'(\beta_j^T\mathbf{W}_i(t))\mathbf{W}_i^T(t)dt \end{bmatrix}^T\right\},$$

$$A_j(\theta_j) = E\left\{\int_0^\tau \rho_{ij}(t)Y_{ij}(t)\{D_i(\theta_j) - \bar{d}_j(t;\theta_j)\}\begin{bmatrix} h'(\eta_j^T\mathbf{Z}_i)\mathbf{Z}_i^T d\Lambda_{0j}(t) \\ g'(\alpha_j^T\mathbf{G}_i)\mathbf{G}_i^T dt \end{bmatrix}^T\right\},$$

with

$$\bar{d}_j(t;\theta_j^C) = \frac{E\{Y_{ij}(t)h(\gamma_j^T\mathbf{X}_i(t))D_i(t;\theta_j^C)\}}{E\{Y_{ij}(t)h(\gamma_j^T\mathbf{X}_i(t))\}}, \quad \bar{d}_j(t;\theta_j) = \frac{E\{\rho_{ij}(t)Y_{ij}(t)h(\eta_j^T\mathbf{Z}_i)D_i(\theta_j)\}}{E\{\rho_{ij}(t)Y_{ij}(t)h(\eta_j^T\mathbf{Z}_i)\}}.$$

(C6) $P\{A_i = j \mid \mathbf{H}_i\} \in (0, 1)$.

These conditions above are common in the survival analysis and causal inference literature and are satisfied in many applications. Conditions (C1)-(C6) are similar adopted in Zhang and Schaubel (2011).

Theorem 1 *Suppose the conditions* $(C1) - (C6)$ *hold, then we have*

(i) $\hat{\delta}$ *converges in probability to* δ*;*

(ii) When $\kappa(t;\mathbf{H}_i,A_i)=1$, $n^{1/2}(\hat{\delta}-\delta)$ *is asymptotically normal with mean zero and variance* $E\{(\phi_{i1}-\phi_{i0})^2\}$*, where*

$$\phi_{ij}=-E\left\{\begin{bmatrix}\int_0^L S_{ij}(t)\Lambda_{0j}(t)h'(\eta_j^T\mathbf{Z}_i)\mathbf{Z}_i^T dt\\ \int_0^L S_{ij}(t)tg'(\alpha_j^T\mathbf{G}_i)\mathbf{G}_i^T dt\end{bmatrix}^T\right\}A_j^{-1}(\theta_j)U_{ij}(\theta_j)$$
$$-\int_0^L E\Big\{h(\eta_j^T\mathbf{Z}_i)[u_{ij}(L)-u_{ij}(t)]\Big\}d\Phi_{ij}(t)+(u_{ij}-u_j),$$

$\Phi_{ij}(t)$ *and* $U_{ij}(\theta_j)$ *are defined in the Appendix A, and* $u_{ij}=u_{ij}(L)$ *with* $u_{ij}(t)=\int_0^t S_{ij}(s)ds$*.*

It is easy to see that Theorem 1 is an extension to that for the proportional hazards model in Zhang and Schaubel (2011). In fact, when taking $g(x)=0$, Theorem 1 reduces to the results in Theorem 1 of Zhang and Schaubel (2011). The asymptotical variance $E\{(\phi_{i1}-\phi_{i0})^2\}$ in Theorem 1 (ii) can be consistently estimated by $n^{-1}\sum_{i=1}^n(\hat{\phi}_{i1}-\hat{\phi}_{i0})^2$, where $\hat{\phi}_{ij}$ is the empirical estimate of ϕ_{ij}. However, as shown in the Appendix A, $U_{ij}(\theta_j)$ and $\Phi_{ij}(t)$ have very complicated forms. Thus, $\hat{\phi}_{ij}$ and $n^{-1}\sum_{i=1}^n(\hat{\phi}_{i1}-\hat{\phi}_{i0})^2$ are not easily computed.

As suggested by Zhang and Schaubel (2011), we can estimate $U_{ij}(\theta_j)$ and $\Phi_{ij}(t)$ by $\hat{U}_{ij}^{\dagger}(\hat{\theta}_j)$ and $\hat{\Phi}_{ij}^{\dagger}(t)$, where

$$\hat{U}_{ij}^{\dagger}(\hat{\theta}_j)=\int_0^{\tau}\{D_i-\bar{D}_j(t;\hat{\theta}_j,\hat{\rho})\}\hat{\rho}_{ij}(t)\{dN_{ij}(t)-Y_{ij}(t)g(\hat{\alpha}_j^T\mathbf{G}_i)dt\},$$
$$\hat{\Phi}_{ij}^{\dagger}(t)=\begin{bmatrix}-\int_0^t a_{g'}(s)ds\\ -\int_0^t a_{h'}(s)d\hat{\Lambda}_{0j}(s)\end{bmatrix}^T\hat{A}(\hat{\theta}_j)^{-1}\hat{U}_{ij}^{\dagger}(\hat{\theta}_j)$$
$$+\int_0^t\hat{\rho}_{ij}(s)a_0^{-1}(s)\{dN_{ij}(s)-Y_{ij}(s)[g(\hat{\alpha}_j^T\mathbf{G}_i)ds+h(\hat{\eta}_j^T\mathbf{Z}_i)d\hat{\Lambda}_{0j}(s)]\},$$

where $a_0(s)=n^{-1}\sum_{i=1}^n\hat{\rho}_{ij}(s)Y_{ij}(s)h(\hat{\eta}_j^T\mathbf{Z}_i)$,

$$a_{g'}(s)=\frac{n^{-1}\sum_{i=1}^n\hat{\rho}_{ij}(s)Y_{ij}(s)g'(\hat{\alpha}_j^T\mathbf{G}_i)\mathbf{G}_i^T}{n^{-1}\sum_{i=1}^n\hat{\rho}_{ij}(s)Y_{ij}(s)h(\hat{\eta}_j^T\mathbf{Z}_i)},\quad a_{h'}(s)=\frac{n^{-1}\sum_{i=1}^n\hat{\rho}_{ij}(s)Y_{ij}(s)h'(\hat{\eta}_j^T\mathbf{Z}_i)\mathbf{Z}_i^T}{n^{-1}\sum_{i=1}^n\hat{\rho}_{ij}(s)Y_{ij}(s)h(\hat{\eta}_j^T\mathbf{Z}_i)}.$$

In summary, $E\{(\phi_{i1}-\phi_{i0})^2\}$ can be estimated by $n^{-1}\sum_{i=1}^n(\hat{\phi}_{i1}^{\dagger}-\hat{\phi}_{i0}^{\dagger})^2$, where

$$\hat{\phi}_{ij}^{\dagger}=-n^{-1}\sum_{l=1}^n\begin{bmatrix}\int_0^L\hat{S}_{lj}(t)\hat{\Lambda}_{0j}(t)h'(\hat{\eta}_j^T\mathbf{Z}_l)\mathbf{Z}_l^T dt\\ \int_0^L\hat{S}_{lj}(t)tg'(\hat{\alpha}_j^T\mathbf{G}_l)\mathbf{G}_l^T dt\end{bmatrix}^T\hat{A}_j^{-1}(\hat{\theta}_j)\hat{U}_{ij}^{\dagger}(\hat{\theta}_j)$$
$$-\int_0^L\Big[n^{-1}\sum_{l=1}^n h(\hat{\eta}_j^T\mathbf{Z}_l)\{\hat{u}_{lj}(L)-\hat{u}_{lj}(t)\}\Big]d\hat{\Phi}_{ij}^{\dagger}(t)+(\hat{u}_{ij}-\hat{u}_j),$$

where $\hat{S}_{lj}(t)$, $\hat{u}_{ij}$ and $\hat{u}_j$ are the empirical counterparts of $S_{lj}(t)$, u_{ij} and u_j.

In Theorem 1 (ii), we focus on $\kappa(t;\mathbf{H}_i;A_i)=1$. For general $\kappa(t;\mathbf{H}_i;A_i)$, the asymptotic properties for the resulting estimator can be obtained in a similar way. However, its asymptotic variance is more complicated. As discussed by Zhang and Schaubel (2011), we still compute the asymptotic variance of $\hat{\delta}$ by $n^{-1}\sum_{i=1}^n(\hat{\phi}_{i1}^{\dagger}-\hat{\phi}_{i0}^{\dagger})^2$ in this case.

4 Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to assess the finite sample performance of the proposed methods. For our methods, we consider the two stabilization factors: $\kappa(t;\mathbf{H}_i,A_i)=1$ (denoted by "Unstable") and $\kappa(t;\mathbf{H}_i,A_i)=\exp(-\hat{\Lambda}_{ij}^{C}(t|\mathbf{H}_i,A_i))$ (denoted by "Stable"), where $\hat{\Lambda}_{ij}^{C}(t|\mathbf{H}_i,A_i)$ is similarly obtained by (9). As a counterpart, we consider an unweighted method by using $\rho_{ij}(t)=1$ in (7) and (8) (denoted by "Unweight"). In fact, the unweighted method is similar to that proposed by Chen and Tsiatis (2001). We can easily see that, unweighted method is efficient when C_2 is independent censoring, whereas it may loss efficiency if C_2 is dependent censoring.

In this simulation studies, we consider the following process of generating data:

1): Baseline covariates $Z\in\mathbb{R}$ and $G\in\mathbb{R}$ are both generated from Uniform$(0,1)$.

2): Group indicator $A\sim$ Bernoulli$(1,p)$, where

$$p=\exp(-0.15Z-0.15G)/(1+\exp(-0.15Z-0.15G)).$$

3): Survival time T is generated by the following AMH model

$$T=\begin{cases}-\log(\epsilon_1)/(\exp(-1.6+1.3Z)+0.15G), & \text{if } A=0,\\ -\log(\epsilon_1)/(\exp(-2.6+1.1Z)+0.15G), & \text{if } A=1,\end{cases}$$

where $\epsilon_1\sim$ Uniform$(0,1)$.

4): Generate intermediate variable X by $X=-5\log(A\epsilon_1+(1-A)\epsilon_1)+\epsilon_2$, where $\epsilon_2\sim$ Uniform$(0,1)$.

5): Generate the time-dependent covariates by $X(t)=I(X\geq t)$ and $W(t)=I(W\geq t)$, where $W\sim$ Uniform$(0,1)$.

6): Generate the dependent censoring time C_2 based on the AMH model

$$\lambda^{C}(t|X(t),W(t),A)=\exp(\gamma_1+\gamma_2 Z+0.2A+\gamma_3 X(t))+\gamma_4 W(t).$$

7): Generate the independent censoring time C_1 by the exponential distribution with the parameter $0.1\gamma_5$.

Using the above process of generating data, we consider the following five scenarios, corresponding to different confounding mechanisms. In each scenario, we consider two different percentages of censoring: light censoring and heavy censoring. Specifically, for the light censoring, about 5% subjects are censored by C_1 and about 20% subjects are censored by C_2; for the heavy censoring, about 30% subjects are censored by C_2. Throughout our experiments, we choose $L=10$, and repeat each setting 1000 times with sample size $n=800$.

Table 1: Estimation of δ under Scenarios 1-4 for light censoring case

Method	True	MC Bias	MSE	MC SD	Ave.SE	CP
Light censoring, Scenario 1						
Stable	2.097	0.032	0.041	0.206	0.200	0.959
Unstable	2.097	0.031	0.041	0.201	0.200	0.954
Unweight	2.097	0.053	0.049	0.237	0.216	0.962
Light censoring, Scenario 2						
Stable	2.097	0.009	0.042	0.211	0.206	0.958
Unstable	2.097	0.009	0.042	0.201	0.206	0.947
Unweight	2.097	0.017	0.046	0.233	0.214	0.967
Light censoring, Scenario 3						
Stable	2.097	0.073	0.052	0.203	0.211	0.939
Unstable	2.097	0.074	0.053	0.199	0.214	0.943
Unweight	2.097	0.110	0.057	0.241	0.221	0.957
Light censoring, Scenario 4						
Stable	2.762	0.018	0.054	0.054	0.054	0.943
Unstable	2.762	0.019	0.055	0.051	0.054	0.932
Unweight	2.762	0.059	0.064	0.074	0.061	0.955

Scenario 1 *In this scenario, we consider the setting that there exist both the baseline and time-dependent confounders. We set* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-4.0, 0.5, 1.6, 0.2, 0.1)$ *for light censoring case, and* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-3.6, 0.5, 1.6, 0.2, 0.1)$ *for heavy censoring case.*

Scenario 2 *In the scenario, we consider the case that there only exists the baseline confounders. To the end, we choose* $\gamma_3 = 0$*, which implies that* $X(t)$ *correlates only with* T *but not with* C_2*. We set* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$ *to (-3.5, 1, 0, 0.2, 0.1) for light censoring case and to (-2.8, 1, 0, 0.2, 0.1) for heavy censoring case.*

Scenario 3 *In the scenario, we consider the case that* $X(t)$ *is conditionally independent of* T *but correlated with* C_2*, which implies that* $X(t)$ *is not a confounder either. Data are generated the same as in Scenario 1 except that* $X = -5\log(A\epsilon_3 + (1-A)\epsilon_3) + \epsilon_2$*, where* $\epsilon_3 \sim$ *Uniform*$(0, 1)$*.*

Scenario 4 *In the scenario, we consider that some covariates have both additive and multiplicative effects on* T*. Other parameters are set as those in Scenario 1. Let the AMH functions of the group 0 and 1 be* $\exp(-1.6+1.3Z)+0.15Z$ *and* $\exp(-2.6+1.1Z)+0.15Z$*, respectively.*

Tables 1 and 2 summarize the simulation results for Scenarios 1-4. In the two tables, "True" denotes the true δ. In the Appendix B, we provide more details on calculating the

Table 2: Estimation of δ under Scenarios 1-4 for heavy censoring case

Method	True	MC Bias	MSE	MC SD	Ave.SE	CP
Heavy censoring, Scenario 1						
Stable	2.097	0.042	0.046	0.201	0.211	0.940
Unstable	2.097	0.042	0.046	0.194	0.211	0.925
Unweight	2.097	0.087	0.063	0.253	0.235	0.952
Heavy censoring, Scenario 2						
Stable	2.097	0.031	0.047	0.216	0.215	0.949
Unstable	2.097	0.030	0.047	0.195	0.216	0.917
Unweight	2.097	0.042	0.054	0.246	0.230	0.956
Heavy censoring, Scenario 3						
Stable	2.097	0.087	0.057	0.201	0.223	0.935
Unstable	2.097	0.086	0.058	0.197	0.217	0.938
Unweight	2.097	0.147	0.065	0.247	0.253	0.944
Heavy censoring, Scenario 4						
Stable	2.762	0.007	0.056	0.053	0.056	0.927
Unstable	2.762	0.007	0.055	0.048	0.055	0.913
Unweight	2.762	0.098	0.078	0.075	0.068	0.951

true $\delta = 2.907$ in Scenarios 1-3. By the same way, we can calculate $\delta = 2.762$ in Scenario 4. We also report the Monte Carlo absolute of bias (MC Bias for short), the mean squared error (MSE), the Monte Carlo standard deviation of estimates (MC SD), the Monte Carlo average of estimated standard errors (Ave.SE), and the coverage probability of nominal 95% Wald confidence interval (CP).

From Tables 1 and 2, we can see that our two proposed "Unstable" and "Stable" estimators perform better than the Unweighted method across all scenarios. Among the four scenarios, Scenario 2 performs best. This is because the time-dependent confounding factors do not influence T and C_2, which improves the estimation accuracy.

Theoretically, the Stable method ($\kappa(t; \mathbf{H}, A) \neq 1$) should be better than Unstable methods ($\kappa(t; \mathbf{H}, A) = 1$). However, from Tables 1 and 2, the performance difference between the two methods is not obvious. The main reason may be that, when the sample size is finite, the stabilization factor for the Stable method is very complicated, which may lead to loss of its estimation accuracy.

From Tables 1 and 2 for Scenario 4, we can also see that the estimation of δ has a reasonable accuracy in the AHM model, in which the common covariate Z has both additive and multiplicative effects on T. In practice, this modeling approach in Scenario 4 is very useful, because it does not need to identify which covariates have additive effects and which have multiplicative effects before modeling. See the following empirical analysis of the PBC data.

Different from the settings in Scenarios 1-4, the following Scenario 5 considers more general distributions of Z and G other than uniform distributions.

Scenario 5 *In the scenario, we consider that baseline covariates sample from other distribution. Assume that* $Z \sim N(0, 0.4)$ *and* $G \sim B(1, 0.5)$. *Data generating show as follows:*

S1 *This case is similar to Scenario 1. We set* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-3.5, 0.8, 1.6, 0.2, 0.1)$ *for light censoring case, and* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-4.5, 0.8, 1.1, 0.2, 0.1)$ *for heavy censoring case.*

S2 *This case is similar to Scenario 2. We set* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-3.5, 0.8, 0, 0.2, 0.1)$ *for light censoring case, and* $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (-4.5, 0.8, 0, 0.2, 0.1)$ *for heavy censoring case. The other settings are the same as that of* ***S1****.*

S3 *Data are generated as those in* ***S1****, except that* $X = -5\log(A\epsilon_3 + (1-A)\epsilon_3) + \epsilon_2$, *where* $\epsilon_3 \sim$ *Uniform*$(0, 1)$.

S4 *Other parameters are set as those in* ***S1****. Let the AMH function of group 0 be:* $\exp(-1.6 + 1.3Z) + 0.15Z$, *and that of group 1 be:* $\exp(-2.6 + 1.1Z) + 0.15Z$, *where* $Z \sim$ *Poisson*(0.5).

The simulation results for Scenario 5 are reported in Tables 3 and 4. From the results, we can see that the performances are similar to those in Scenarios 1-4. Thus, we conclude that our proposed methods work well in the settings of $Z \sim N(0, 0.4)$ and $G \sim$ *Bernoulli*$(1, 0.5)$.

5 Real data analysis

In this section, we illustrate the proposed method by empirical analysis of the PBC data, which is the follow-up record of patients with PBC in Mayo Clinic (Fleming and Harrington, 2011). The PBC is an immune-mediated chronic cholestatic liver disease with a slowly progressive course which can eventually lead to death. The data can be found from the JM package in R.

This dataset contains 312 patients, of which 154 patients are in the control group ($A = 0$) and 158 patients are in the drug D-penicillamine (D-penicil) treatment group ($A = 1$). Here, our goal is to estimate and assess the treatment effect of the D-penicil. For this data, dependent censoring occurred due to liver transplantation, and independent censoring was due to loss to follow-up and administrative censoring at the end of the study period.

The data have been studied by many authors to develop the parameter estimation or variable selection in various survival models, such as, Fleming and Harrington (2011); Shows et al. (2010); Tibshirani (1997); Zhang and Lu (2007). Different from the previous

Table 3: Estimation of δ under Scenarios 5 for light censoring case

Method	True	MC Bias	MSE	MC SD	Ave.SE	CP
Light censoring, **S1**						
Stable	1.854	0.053	0.057	0.254	0.233	0.947
Unstable	1.854	0.053	0.057	0.255	0.233	0.944
Unweight	1.854	0.114	0.058	0.262	0.240	0.937
Light censoring, **S2**						
Stable	1.854	0.072	0.062	0.259	0.238	0.955
Unstable	1.854	0.071	0.063	0.261	0.241	0.946
Unweight	1.854	0.094	0.063	0.263	0.246	0.967
Light censoring, **S3**						
Stable	1.854	0.008	0.057	0.260	0.238	0.952
Unstable	1.854	0.008	0.057	0.261	0.238	0.953
Unweight	1.854	0.110	0.069	0.264	0.240	0.927
Light censoring, **S4**						
Stable	2.185	0.018	0.054	0.054	0.054	0.943
Unstable	2.185	0.019	0.055	0.051	0.054	0.932
Unweight	2.185	0.059	0.064	0.074	0.061	0.955

Table 4: Estimation of δ under Scenarios 5 for heavy censoring case

Method	True	MC Bias	MSE	MC SD	Ave.SE	CP
Heavy censoring, **S1**						
Stable	1.854	0.058	0.073	0.270	0.264	0.947
Unstable	1.854	0.056	0.074	0.273	0.266	0.951
Unweight	1.854	0.072	0.093	0.311	0.296	0.954
Heavy censoring, **S2**						
Stable	1.854	0.073	0.070	0.267	0.254	0.956
Unstable	1.854	0.073	0.071	0.271	0.256	0.954
Unweight	1.854	0.035	0.072	0.275	0.266	0.959
Heavy censoring, **S3**						
Stable	1.854	0.012	0.065	0.265	0.254	0.953
Unstable	1.854	0.013	0.065	0.267	0.255	0.953
Unweight	1.854	0.133	0.083	0.271	0.256	0.937
Heavy censoring, **S4**						
Stable	2.185	0.018	0.054	0.054	0.054	0.943
Unstable	2.185	0.019	0.055	0.051	0.054	0.932
Unweight	2.185	0.059	0.064	0.074	0.061	0.955

studies, our interest is the survival time of patients in different treatment groups. Here we focus on the following baseline adjustment covariates $\mathbf{Z}$, including Age (in years), Sex (0 for female woman and 1 for male) and Histologic (four stages). In practice, it is hard to determine which covariates have additive effects and which have multiplicative effects. To overcome the issue, similar to Scenario 4 in the above section, we establish the following AMH model for the survival time T

$$\lambda(t|\mathbf{Z}_i, A_i = j) = \lambda_{0j}(t)\exp\{\eta_j^T\mathbf{Z}_i\} + \alpha_j^T\mathbf{Z}_i,\ \ j = 0, 1, \tag{12}$$

where $\eta_j = (\eta_{j,\text{Age}}, \eta_{j,\text{Sex}}, \eta_{j,\text{Histologic}})^T$ and $\alpha_j = (\alpha_{j,\text{Age}}, \alpha_{j,\text{Sex}}, \alpha_{j,\text{Histologic}})^T$. For the dependent censoring time C_2, we consider the following the AMH model

$$\lambda_{ij}^C(t|\mathbf{X}_i(t), A_i = j) = \lambda_{0j}^C 0(t)\exp\{\gamma_j^T\mathbf{X}_i(t)\} + \beta_j^T\mathbf{X}_i(t), \tag{13}$$

where $\mathbf{X}_i(t)$ is the time-dependent covariates, including SerBilir (serum bilirubin in mg/dl) and Albumin (albumin in gm/dl), $\gamma_j = (\gamma_{j,SerBilir}, \gamma_{j,Albumin})^T$ and $\beta_j = (\beta_{j,SerBilir}, \beta_{j,Albumin})^T$.

Table 5 summarizes the estimation of the coefficients in model (12) under different treatment groups. It is easy to see that Histologic has obvious effect both in the additive and multiplicative parts. In addition, Age also has additive and multiplicative effects in the control group. These results suggest that the explanation of this factor may not be enough if we only use a Cox model or additive hazards model.

Table 5: Parameters estimation of failure time T.

Treatment	Coefficient	Estimation	SE	p-value
Treatment group	η_{Age}	0.4420	0.3112	0.07
	η_{Sex}	0.2593	0.2476	0.14
	$\eta_{\text{Histologic}}$	0.2055	0.0593	<0.01
	α_{Age}	0.0187	0.6202	0.48
	α_{Sex}	0.0433	0.1221	0.36
	$\alpha_{\text{Histologic}}$	0.2885	0.1199	<0.01
Control group	η_{Age}	0.3016	0.1364	0.01
	η_{Sex}	0.0446	0.2816	0.43
	$\eta_{\text{Histologic}}$	0.4391	0.2732	0.05
	α_{Age}	0.0310	0.0154	0.02
	α_{Sex}	0.0447	0.9475	0.48
	$\alpha_{\text{Histologic}}$	0.2723	0.1772	0.06

From Table 5, it can be seen that the Histologic is significant in model (12). In the following contents, we are interested in the effects of Histologic on the restricted mean lifetime. That is, we again consider model (12) using only Histologic (excluding Sex and

Table 6: Analysis results for the PBC data for L = 1 to 5 years.

	Stable					Unstable				
L	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\delta}$	SE($\hat{\delta}$)	p-value	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\delta}$	SE($\hat{\delta}$)	p-value
1	0.954	0.972	0.018	0.017	0.145	0.954	0.972	0.018	0.017	0.139
2	1.838	1.901	0.063	0.049	0.100	1.838	1.901	0.063	0.047	0.090
3	2.644	2.782	0.138	0.094	0.072	2.644	2.782	0.138	0.093	0.071
4	3.404	3.589	0.185	0.145	0.102	3.404	3.590	0.186	0.146	0.103
5	4.116	4.352	0.236	0.182	0.098	4.116	4.353	0.247	0.182	0.099

Age). Firstly, we apply the Stable method and the Unstable method to compare the restricted mean lifetime for $L = 1$ to 5 years. The results are summarized in Table 6, where $\hat{\mu}_0$ and $\hat{\mu}_1$ are the restricted mean lifetime estimators of the placebo and D-penicil groups, $\hat{\delta}$ is the estimator of δ, and SE($\hat{\delta}$) is the standard error estimate of $\hat{\delta}$. Form Table 6, we can see that the two methods yield similar results for all estimates. This may be because for this data, the dependent censoring rate is lower and thus the weighting estimator $\hat{\rho}_{ij}(t)$ is close to 1.

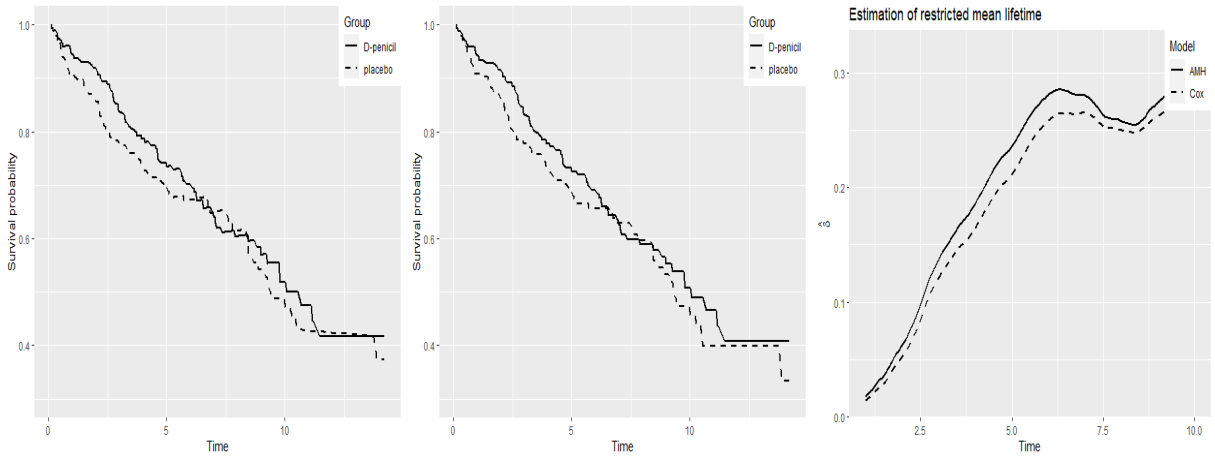
Form Table 6, we can further see that there is slightly significant difference between μ_0 and μ_1 at $L = 3$ (p-value ≈ 0.07). Thus, we conclude that the average restricted mean lifetime for the D-penicil group is slightly greater than the placebo group during the first 3 years.

We also compare our methods with the existing methods proposed by (Zhang and Schaubel, 2011), which replace the AMH models in (12) and (13) by Cox models, respectively. To the end, we plot the estimated survival curves for two treatment groups, define by $S_1(t) \triangleq E_{\mathbf{H}}\{P(T > t|A = 1, \mathbf{H})\}$ and $S_0(t) \triangleq E_{\mathbf{H}}\{P(T > t|A = 0, \mathbf{H})\}$. The results are displayed in Figures 1 (a)-(b). To further explore the difference between the two methods, we plot the estimated curve of $\delta(L) \triangleq \int_0^L [S_1(t) - S_0(t)]dt$ against L in Figure 1 (c). From Figure 1 (c), an interesting finding is that the proposed curve based on the AMH models is higher than that for Cox models (Zhang and Schaubel, 2011). It suggests that our method seems to outperform the Cox models when distinguishing the treatment effect of the drug D-penicillamine in this dataset.

6 Discussion

In this paper, we estimate treatment effect in restricted mean lifetime with observational data subject to two censoring mechanisms under the additive-multiplicative hazards model. A new estimating equation approach based on inverse probability weighting is proposed. The consistency and asymptotic normality of the resulting estimators are established. The simulation results show that the proposed estimation procedure has reasonable finite sample performance.

In many applications, it is not enough to only use the Cox or additive hazards model.



(a) Survival curves for AMH model (b) Survival curves for Cox model (c) Estimated curves of $\delta(L)$

Figure 1: Comparisons for PBC data

For example, when analysing PBC data, we find that Histologic has both multiplicative and additive effects. Thus, for PBC data, the Cox or additive hazards model mat not be suitable. In summary, we think that our proposed method based on the AMH model is a useful tool to compare effects of different treatments in epidemiologic studies.

In practice, we may encounter that the estimated directions from the additive and multiplicative components are inconsistent. However, we think that such inconsistency may not be a bad thing. In the real world, there indeed exists the situation that some covariates may have several different effects, which may be inconsistent. It is challenging to analyse this phenomenon, which needs further research.

Acknowledgement

The authors would like to thank the editor, associate editor and reviewers for their constructive comments. Jicai Liu's research was supported by the National Science Foundation of China (11426156, 11501372, 11971018) and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education (KLATASDS2001). Riquan Zhang's research was supported by the National Natural Science Foundation of China (1197117111831008).

Appendix A: Proof of Theorem 1

For simplicity, we first introduce some notations as follows:

$$\tilde{a}_0^C(s) = E\{Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))\}, \qquad a_0^C(s) = n^{-1}\sum_{i=1}^n Y_{ij}(s)h(\hat{\gamma}_j^T\mathbf{X}_i(s)),$$

$$\tilde{a}_{h'}^C(s) = \frac{E\{Y_{ij}(s)h'(\gamma_j^T\mathbf{X}_i(s))\mathbf{X}_i^T(s)\}}{E\{Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))\}}, \qquad a_{h'}^C(s) = \frac{n^{-1}\sum Y_{ij}(s)h'(\gamma_j^T\mathbf{X}_i(s))\mathbf{X}_i^T(s)}{n^{-1}\sum_{i=1}^n Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))},$$

$$\tilde{a}_{g'}^C(s) = \frac{E\{Y_{ij}(s)g'(\beta_j^T\mathbf{W}_i(s))\mathbf{W}_i^T(s)\}}{E\{Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))\}}, \qquad a_{g'}^C(s) = \frac{n^{-1}\sum Y_{ij}(s)g'(\beta_j^T\mathbf{W}_i(s))\mathbf{W}_i^T(s)}{n^{-1}\sum_{i=1}^n Y_{ij}(s)h(\gamma_j^T\mathbf{X}_i(s))},$$

$$\tilde{a}_0(s) = E\{\rho_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)\}, \qquad a_0(s) = n^{-1}\sum_{i=1}^n \rho_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i),$$

$$\tilde{a}_{h'}(s) = \frac{E\{\rho_{ij}(s)Y_{ij}(s)h'(\alpha_j^T\mathbf{Z}_i)\mathbf{Z}_i^T\}}{E\{\rho_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)\}}, \qquad a_{h'}(s) = \frac{n^{-1}\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)h'(\alpha_j^T\mathbf{Z}_i)\mathbf{Z}_i^T}{n^{-1}\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)},$$

$$\tilde{a}_{g'}(s) = \frac{E\{\rho_{ij}(s)Y_{ij}(s)g'(\eta_j^T\mathbf{G}_i)\mathbf{G}_i^T\}}{E\{\rho_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)\}}, \qquad a_{g'}(s) = \frac{n^{-1}\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)g'(\eta_j^T\mathbf{G}_i)\mathbf{G}_i^T}{n^{-1}\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)}.$$

By the weak law of large numbers, $a_0^C(s)$, $a_{h'}^C(s)$, $a_{g'}^C(s)$, $a_0(s)$, $a_{h'}(s)$ and $a_{g'}(s)$ converge in probability to $\tilde{a}_0^C(s)$, $\tilde{a}_{h'}^C(s)$, $\tilde{a}_{g'}^C(s)$, $a_0(s)$, $\tilde{a}_{h'}(s)$ and $\tilde{a}_{g'}(s)$, respectively. To prove Theorem 1, we need the following lemmas.

Lemma 1 *Under Conditions* $(C1)-(C6)$, $\hat{\theta}_j^C$ *is consistent to* θ_j^C, *and* $\hat{\Lambda}_{0j}^C(t)$ *converges in probability to* $\Lambda_{0j}^C(t)$ *uniformly in* $t \in [0,\tau]$. *Furthermore,*

$$n^{1/2}(\hat{\theta}_j^C - \theta_j^C) = A_j(\theta_j^C)^{-1}n^{-1/2}\sum_{i=1}^n U_{ij}^C(\theta_j^C) + o_p(1), \tag{A.1}$$

$$n^{1/2}(\hat{\Lambda}_{0j}^C(t) - \Lambda_j^C(t)) = n^{-1/2}\sum_{i=1}^n \Phi_{ij}^C + o_p(1), \tag{A.2}$$

where

$$U_{ij}^C(\theta_j^C) = \int_0^\tau \{D_i(t) - \bar{d}_j(t,\theta_j^C)\}dM_{ij}^C(t), \quad \tilde{a}^C(t) = -\begin{bmatrix} \int_0^t \tilde{a}_{g'}^C(s)ds \\ \int_0^t \tilde{a}_{h'}^C(s)\lambda_j^C(s)ds \end{bmatrix},$$

$$\Phi_{ij}^C = \int_0^t \frac{dM_{ij}^C(s)}{\tilde{a}_0^C(s)} + [\tilde{a}^C(t)]^T A(\theta_j^C)^{-1}U_{ij}^C(\theta_j^C).$$

Proof It follows from Theorem 2.2 and Theorem 3.1 of Lin and Ying (1995) that $\hat{\theta}_j^C$ is consistent to θ_j^C, and $n^{1/2}(\hat{\Lambda}_{0j}^C(t) - \Lambda_j^C(t))$ converges weakly, $t \in [0,\tau]$, to a zero-mean gaussian process. By Taylor expansion of estimating equation in (6), we obtain (A.1),

where $U_{ij}^C(\theta_j^C)$, are i.i.d and zero-mean. By Taylor expansion at $\hat{\theta}_j = \theta_j$, we have that

$$\begin{aligned} n^{1/2}(\hat{\Lambda}_{0j}^C(t) - \Lambda_{0j}^C(t)) &= n^{1/2}(\hat{\Lambda}_{0j}^C(t;\hat{\theta}_j^C) - \hat{\Lambda}_{0j}^C(t;\theta_j^C)) + n^{1/2}(\hat{\Lambda}_{0j}^C(t;\theta_j^C) - \Lambda_{0j}^C(t)) \\ &= n^{-1/2}\sum_{i=1}^n \Big\{ \int_0^t \frac{dM_{ij}^C(t)}{\tilde{a}_0^C(s)} + n^{-1/2}[\tilde{a}^C(t)]^T A(\theta_j^C)^{-1} U_{ij}^C(\theta_j^C) \Big\} + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^n \Phi_{ij}^C + o_p(1), \end{aligned}$$

where $U_{ij}^C(\theta_j^C)$ and $dM_{ij}^C(t)$ are i.i.d and mean-zero. □

Lemma 2 *Under Conditions* $(C1) - (C6)$*, we obtain that* $\hat{\theta}_j$ *is consistent to* θ_j*, and*

$$n^{1/2}(\hat{\theta}_j - \theta_j) \xrightarrow{d} \mathcal{N}(0, A_j^{-1}(\theta_j) V_j \{A_j(\theta_j)^{-1}\}^T),$$

where $V_j = E\{U_{ij}(\theta_j)\}^{\otimes 2}$ *with* $U_{ij}(\theta_j)$ *defined in* (A.6).

Proof By Taylor expansion at $\hat{\theta}_j = \theta_j$ of $U(\tau;\hat{\theta}_j)$, we obtain

$$n^{1/2}(\hat{\theta}_j - \theta_j) = n^{-1/2} A_j^{-1}(\theta_j) \sum_{i=1}^n V_{ij}(\theta_j, \hat{\rho}) + o_p(1),$$

where $V_{ij}(\theta_j, \hat{\rho}) = \int_0^\tau \{D_i - \bar{D}_j(\theta, \hat{\rho})\}\hat{\rho}_{ij}(t) dM_{ij}(t)$. Consider the following decomposition

$$n^{-1/2}\sum_{i=1}^n V_{ij}(\theta_j, \hat{\rho}) = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{D_i - \bar{D}_j(t;\theta_j,\rho)\}\rho_{ij}(t) dM_{ij}(t) \tag{A.3}$$

$$- n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\bar{D}_j(t;\theta_j,\hat{\rho}) - \bar{D}_j(t;\theta_j,\rho)\}\hat{\rho}_{ij}(t) dM_{ij}(t) \tag{A.4}$$

$$+ n^{-1/2}\sum_{i=1}^n \int_0^\tau \{D_i - \bar{D}_j(t;\theta_j,\rho)\}(\hat{\rho}_{ij}(t) - \rho_{ij}(t)) dM_{ij}(t). \tag{A.5}$$

By the theory of empirical processes (Lin et al., 2000), it can be shown that

$$(A.3) = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{D_i - \bar{d}_j(t;\theta_j,\rho)\}\rho_{ij}(t) dM_{ij}(t) + o_p(1).$$

By the Functional Delta Method, (A.4) can be shown to converge in probability to 0. Additionally,

$$\begin{aligned} (A.5) &= \hat{H}_j n^{-1/2}\sum_{l=1}^n A(\theta_j)^{-1} U_{ij}^C(\theta_j) + n^{-1/2}\sum_{l=1}^n \int_0^\tau \hat{G}(s,\tau)(\tilde{a}_0^C(s))^{-1} dM_{lj}^C(s) + o_p(1) \\ &= H_j n^{-1/2}\sum_{l=1}^n A(\theta_j)^{-1} U_{ij}^C(\theta_j) + n^{-1/2}\sum_{l=1}^n \int_0^\tau G(s,\tau)(\tilde{a}_0^C(s))^{-1} dM_{lj}^C(s) + o_p(1), \end{aligned}$$

where

$$\begin{aligned}
\hat{H}_j &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \{D_i - \bar{D}_j(t;\theta_j,\rho)\}\rho_{ij}(t)I_{ij}^T(t)dM_{ij}(t),\\
\hat{G}(s,\tau) &= n^{-1}\sum_{i=1}^{n} Y_{ij}(s)h(\gamma^T\mathbf{X}_i(s))\int_s^\tau \{D_i - \bar{D}_j(t;\theta_j,\rho)\}\rho_{ij}(t)dM_{ij}(t),\\
H_j &= E\Big\{\int_0^\tau \{D_i - \bar{d}_j(t;\theta_j,\rho)\}\rho_{ij}(t)I_{ij}^T(t)dM_{ij}(t)\Big\},\\
G_j(s,\tau) &= E\Big\{Y_{ij}(s)h(\gamma^T\mathbf{X}_i(s))\int_s^\tau \{D_i - \bar{D}_j(t;\theta_j,\rho)\}\rho_{ij}(t)dM_{ij}(t)\Big\},
\end{aligned}$$

with

$$I_{ij}(t) = \begin{bmatrix} \int_0^t Y_{ij}(s)(g'\{\beta_j^T\mathbf{W}_i(s)\}\mathbf{W}_i^T(s) - h(\gamma_j^T\mathbf{X}_i(s))\tilde{a}_{g'}^C(s))ds \\ \int_0^t Y_{ij}(s)(h'(\gamma_j^T\mathbf{X}_i(s))\mathbf{X}_i^T(s) - h(\gamma_j^T\mathbf{X}_i(s))\tilde{a}_{h'}^C(s))d\Lambda_{0j}^C(s) \end{bmatrix}^T.$$

Using (A.3)-(A.5), we have

$$n^{1/2}(\hat{\theta}_j - \theta_j) = n^{-1/2}A(\theta_j)^{-1}\sum_{i=1}^{n} U_{ij}(\theta_j) + o_p(1),$$

where

$$\begin{aligned}
U_{ij}(\theta_j) &= \int_0^\tau \{D_i - \bar{d}_j(t;\theta_j,\rho)\}\rho_{ij}(t)dM_{ij}(t) + H_jA(\theta_j^C)^{-1}U_{ij}^C(\theta_j^C)\\
&\quad + \int_0^\tau G_j(s,\tau)(\tilde{a}_0^C(s))^{-1}dM_{ij}^C(s). \qquad \text{(A.6)}
\end{aligned}$$

The weak law of large numbers yields that $n^{-1}\sum_{i=1}^n U_{ij}(\theta_j) = o_p(1)$. By the central limit theorem, we have that

$$n^{1/2}(\hat{\theta}_j - \theta_j) \xrightarrow{d} \mathcal{N}(0, A_j^{-1}(\theta_j)V_j\{A_j(\theta_j)^{-1}\}^T). \qquad \square$$

Proof *of Theorem* 1. Following similar arguments as in the proof of Lemma 2, we obtain that $\hat{\Lambda}_{0j}(t)$ converges in probability to $\Lambda_{0j}(t)$ uniformly in $t \in [0,\tau]$ and $\hat{\delta}$ converges in probability to δ.

Firstly, we consider $\hat{\Lambda}_{0j}(t)$. Note that

$$\begin{aligned}
n^{1/2}\{\hat{\Lambda}_{0j}(t) - \Lambda_{0j}(t)\} &= n^{1/2}\{\hat{\Lambda}_{0j}(t;\hat{\rho},\hat{\theta}_j,a_0(s;\hat{\rho})) - \hat{\Lambda}_{0j}(t;\hat{\rho},\theta_j,a_0(s;\hat{\rho}))\} \qquad \text{(A.7)}\\
&\quad + n^{1/2}\{\hat{\Lambda}_{0j}(t;\hat{\rho},\theta_j,a_0(s;\hat{\rho})) - \hat{\Lambda}_{0j}(t;\rho,\theta_j,a_0(s;\hat{\rho}))\} \qquad \text{(A.8)}\\
&\quad + n^{1/2}\{\hat{\Lambda}_{0j}(t;\rho,\theta_j,a_0(s;\hat{\rho})) - \hat{\Lambda}_{0j}(t;\rho,\theta_j,a_0(s;\rho))\} \qquad \text{(A.9)}\\
&\quad + n^{1/2}\{\hat{\Lambda}_{0j}(t;\rho,\theta_j,a_0(s;\rho)) - \Lambda_{0j}(t)\}. \qquad \text{(A.10)}
\end{aligned}$$

For (A.7), we obtain that

$$
\begin{aligned}
(A.7) &= n^{1/2}\begin{bmatrix} -\int_0^t \frac{\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)g'(\eta_j^T\mathbf{G}_i)ds}{\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)} \\ -\int_0^t a_{h'}(s)\frac{\sum_{i=1}^n \hat{\rho}_{ij}(s)[dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds]}{\sum_{i=1}^n \hat{\rho}_{ij}(s)Y_{ij}(s)h(\alpha_j^T\mathbf{Z}_i)} \end{bmatrix}^T (\hat{\theta}_j - \theta_j) + o_p(1) \\
&= n^{1/2}\begin{bmatrix} -\int_0^t a_{g'}(s)ds \\ -\int_0^t a_{h'}(s)d\hat{\Lambda}_{0j}(s) \end{bmatrix}^T (\hat{\theta}_j - \theta_j) + o_p(1) \\
&= n^{1/2}\begin{bmatrix} -\int_0^t \tilde{a}_{g'}(s)ds \\ -\int_0^t \tilde{a}_{h'}(s)d\hat{\Lambda}_{0j}(s) \end{bmatrix}^T (\hat{\theta}_j - \theta_j) + o_p(1) \\
&= n^{1/2}\tilde{a}^T(t)(\hat{\theta}_j - \theta_j) + o_p(1),
\end{aligned}
$$

where $\tilde{a}(t) = \begin{bmatrix} -\int_0^t \tilde{a}_{g'}(s)ds \\ -\int_0^t \tilde{a}_{h'}(s)d\hat{\Lambda}_{0j}(s) \end{bmatrix}$. For (A.8), we obtain that

$$
\begin{aligned}
(A.8) &= n^{-1/2}\sum_{i=1}^n \int_0^t \{\hat{\rho}_{ij}(s) - \rho_{ij}(s)\}a_0^{-1}(s;\hat{\rho})(dN_{ij}(s) - Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds) + o_p(1) \\
&= n^{-1/2}\hat{B}_j(t)A(\theta_j^C)^{-1}\sum_{l=1}^n U_{ij}^C(\theta_j^C) + n^{-1/2}\sum_{l=1}^n \int_0^t \hat{K}(u,t)(\tilde{a}_0^C(u))^{-1}dM_{lj}^C(u) + o_p(1) \\
&= n^{-1/2}B_j(t)A(\theta_j^C)^{-1}\sum_{l=1}^n U_{ij}^C(\theta_j^C) + n^{-1/2}\sum_{l=1}^n \int_0^t K(u,t)(\tilde{a}_0^C(u))^{-1}dM_{lj}^C(u) + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{B}_j(t) &= n^{-1}\sum_{i=1}^n \int_0^t \rho_{ij}(s)a_0^{-1}(s;\hat{\rho})I_{ij}^T(s)(dN_{ij}(s) - Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds), \\
\hat{K}(u,t) &= n^{-1}\sum_{i=1}^n Y_{ij}(u)h(\gamma_j^T\mathbf{X}_i(u))\int_u^t \rho_{ij}(s)a_0^{-1}(s;\hat{\rho})(dN_{ij}(s) - Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds), \\
B_j(t) &= E(\int_0^t \rho_{ij}(s)a_0^{-1}(s;\rho)I_{ij}^T(s)(dN_{ij}(s) - Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)), \\
K(u,t) &= E(Y_{ij}(u)h(\gamma_j^T\mathbf{X}_i(u))\int_u^t \rho_{ij}(s)a_0^{-1}(s;\rho)(dN_{ij}(s) - Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)).
\end{aligned}
$$

It is easy to see that $\hat{B}_j(t)$ and $\hat{K}(u,t)$ converge to $B_j(t)$ and $K(u,t)$ in probabilaity. We

can rewrite expression (A.9) as follows

$$
\begin{aligned}
(A.9) =& n^{-1/2}\sum_{i=1}^{n}\int_0^t \rho_{ij}(s)\{a_0^{-1}(s;\hat{\rho})-a_0^{-1}(s;\rho)\}(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)\\
=& -n^{-1}\sum_{i=1}^{n}\int_0^t \rho_{ij}(s)a_0^{-2}(s;\hat{\rho})\{n^{-1}\sum_{l=1}^{n}Y_{lj}(s)h(\alpha_j^T\mathbf{Z}_l)n^{1/2}(\hat{\rho}_{lj}(s)-\rho_{lj}(s))\}\\
&\times(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)+o_p(1)\\
=& n^{-1/2}\hat{E}_j(t)\sum_{k=1}^{n}A(\theta_j^C)^{-1}U_{kj}^C+n^{-1/2}\sum_{k=1}^{n}\int_0^t \hat{P}_j(u,t)(\tilde{a}_0^C(u))^{-1}dM_{kj}^C(u)+o_p(1)\\
=& n^{-1/2}E_j(t)\sum_{k=1}^{n}A(\theta_j^C)^{-1}U_{kj}^C+n^{-1/2}\sum_{k=1}^{n}\int_0^t P_j(u,t)(\tilde{a}_0^C(u))^{-1}dM_{kj}^C(u)+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{F}_j(s) &= -n^{-1}\sum_{l=1}^{n}Y_{lj}(s)h(\alpha_j^T\mathbf{Z}_l)\rho_{lj}(s)I_{lj}^T(s),\\
\hat{Q}_j(u,s) &= -n^{-1}\sum_{l=1}^{n}Y_{lj}(s)h(\alpha_j^T\mathbf{Z}_l)\rho_{lj}(s)h(\gamma_j^T\mathbf{X}_l(u)),\\
\hat{E}_j(t) &= n^{-1}\sum_{i=1}^{n}\int_0^t \rho_{ij}(s)a_0^{-2}(s;\hat{\rho})\{F_j(s)\}(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds),\\
\hat{P}_j(u,t) &= n^{-1}\sum_{i=1}^{n}\int_u^t \rho_{ij}(s)a_0^{-2}(s;\hat{\rho})Q(u,s)(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds),
\end{aligned}
$$

with their limiting values, given by

$$
\begin{aligned}
F_j(s) &= E\{Y_{lj}(s)h(\alpha_j^T\mathbf{Z}_l)\rho_{lj}(s)I_{lj}^T(s)\},\\
Q_j(u,s)_j &= E\{Y_{lj}(s)h(\alpha_j^T\mathbf{Z}_l)\rho_{lj}(s)h(\gamma_j^T\mathbf{X}_l(u))\},\\
E_j(t) &= E\Big\{\int_0^t \rho_{ij}(s)a_0^{-2}(s;\hat{\rho})\{F_j(s)\}(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)\Big\},\\
P_j(u,t) &= E\Big\{\int_u^t \rho_{ij}(s)a_0^{-2}(s;\hat{\rho})Q(u,s)(dN_{ij}(s)-Y_{ij}(s)g(\eta_j^T\mathbf{G}_i)ds)\Big\}.
\end{aligned}
$$

Finally, (A.10) can be expressed as

$$
(A.10) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \rho_{ij}(s)\tilde{a}_0^{-1}(s)dM_{ij}(s)+o_p(1).
$$

Thus, it follows from (A.7)-(A.10) that uniformly in $t\in[0,\tau]$,

$$
n^{1/2}(\hat{\Lambda}_{0j}(t)-\Lambda_{0j}(t)) = n^{-1/2}\sum_{i=1}^{n}\Phi_{ij}(t)+o_p(1),
$$

where

$$\Phi_{ij}(t) = \begin{bmatrix} -\int_0^t \tilde{a}_{g'}(s)ds \\ -\int_0^t \tilde{a}_{h'}(s)d\hat{\Lambda}_{0j}(s) \end{bmatrix}^T A(\theta_j)^{-1}U_{ij}(\theta_j) + B_j(t)A(\theta_j^C)^{-1}U_{ij}^C(\theta_j^C)$$
$$+ \int_0^t K(u,t)(\tilde{a}_0^C(u))^{-1}dM_{lj}^C(u) + E_j(t)A(\theta_j^C)^{-1}U_{ij}^C(\theta_j^C)$$
$$+ \int_0^t P_j(u,t)(\tilde{a}_0^C(u))^{-1}dM_{kj}^C(u) + \int_0^t \rho_{ij}(s)a_0^{-1}(s)dM_{ij}(s).$$

By the definition of $\hat{\delta} = \int_0^L (\hat{S}_1(t) - \hat{S}_0(t))dt$, we have that

$$\int_0^L n^{1/2}(\hat{S}_j(t) - S_j(t)) = n^{-1/2}\sum_{i=1}^n \int_0^L \{\hat{S}_{ij}(t) - S_{ij}(t)\}dt \tag{A.11}$$
$$+ n^{-1/2}\sum_{i=1}^n \int_0^L \{S_{ij}(t) - S_j(t)\}dt. \tag{A.12}$$

By the functional central limit theorem and Slutsky's Theorem, (A.11) equals

$$(A.11) = -n^{-1}\sum_{i=1}^n \Big\{ n^{-1/2}\int_0^L S_{ij}(t)\begin{bmatrix} \Lambda_{0j}(t)h'(\eta_j^T\mathbf{Z}_i)\mathbf{Z}_i^T \\ tg'(\alpha_j^T\mathbf{G}_i)\mathbf{G}_i^T \end{bmatrix}^T dtA(\theta_j)^{-1}\sum_{l=1}^n U_{lj}(\theta_j)$$
$$- n^{-1/2}\sum_{l=1}^n \int_0^L h(\eta_j^T\mathbf{Z}_i)(\mu_{ij}(L) - \mu_{ij}(s))d\Phi_{lj}(s)\Big\}$$
$$= -E\Big\{ \begin{matrix} \int_0^L S_{ij}(t)\Lambda_{0j}(t)h'(\eta_j^T\mathbf{Z}_i)\mathbf{Z}_i^T dt \\ \int_0^L S_{ij}(t)tg'(\alpha_j^T\mathbf{G}_i)\mathbf{G}_i^T dt \end{matrix} \Big\}^T A(\theta_j)^{-1}n^{-1/2}\sum_{l=1}^n U_{lj}(\theta_j)$$
$$- n^{-1/2}\sum_{l=1}^n \int_0^L E\{h(\eta_j^T\mathbf{Z}_i)(\mu_{ij}(L) - \mu_{ij}(s))\}d\Phi_{lj}(s) + o_p(1).$$

Combining (A.11) with (A.12), we have that

$$\int_0^L n^{1/2}(\hat{S}_j(t) - S_j(t)) = n^{-1/2}\sum_{i=1}^n \phi_{ij} + o_p(1)$$

where

$$\phi_{ij} = -E\Big\{ \begin{matrix} \int_0^L S_{ij}(t)\Lambda_{0j}(t)h'(\eta_j^T\mathbf{Z}_i)\mathbf{Z}_i^T dt \\ \int_0^L S_{ij}(t)tg'(\alpha_j^T\mathbf{G}_i)\mathbf{G}_i^T dt \end{matrix} \Big\}^T A(\theta_j)^{-1}U_{lj}(\theta_j)$$
$$- \int_0^L E\{h(\eta_j^T\mathbf{Z}_i)(\mu_{ij}(L) - \mu_{ij}(s))\}d\Phi_{lj}(s) + (u_{ij} - u_j).$$

Finally, we have that

$$n^{1/2}(\hat{\delta} - \delta) = n^{-1/2}\sum_{i=1}^n (\phi_{i1} - \phi_{i0}) + o_p(1). \tag{A.13}$$

□

Thus, $n^{1/2}(\hat{\delta} - \delta)$ is asymptotically normal with mean zero and variance $E(\phi_{i1} - \phi_{i0})^2$, with the empirical variance estimator $n^{-1}\sum_{i=1}^n (\hat{\phi}_{i1} - \hat{\phi}_{i0})$, where $\hat{\phi}_{ij}$ is calculated by substituting the sample analogs for the terms in (A.13).

Appendix B

In this Appendix, we provide details on calculating the true δ in the Monte Carlo simulations. Here, we only consider Scenarios 1-3 and calculate the true δ as follows:

Step 1: By the setting of $Z \sim Uniform(0,1)$ and $G \sim Uniform(0,1)$, the joint density function $f_{\mathbf{H}}(\cdot,\cdot)$ of $\mathbf{H} = (Z,G)^T$ can be calculated as

$$f_{\mathbf{H}}(z,g) = \begin{cases} 1, 0 \le z \le 1, 0 \le g \le 1 \\ 0, else. \end{cases} \tag{B.1}$$

Step 2: Recall the setting that survival time T is generated by the following AMH model

$$T = \begin{cases} -\log(\epsilon_1)/(\exp(-1.6+1.3Z)+0.15G), & \text{if } A=1, \\ -\log(\epsilon_1)/(\exp(-2.6+1.1Z)+0.15G), & \text{if } A=0, \end{cases}$$

where $\epsilon_1 \sim$ Uniform$(0,1)$. This, together with (B.1), yields that the survival function is equal to

$$S_T(t|\mathbf{H}) = \begin{cases} \exp\{-t(\exp(-1.6+1.3Z)+0.15G)\}, & \text{if } A=0, \\ \exp\{-t(\exp(-2.6+1.1Z)+0.15G)\}, & \text{if } A=1. \end{cases}$$

Step 3: By the setting of $L = 10$, the true δ can be caculated

$$\begin{aligned} \delta &= \int_0^{10} E_{\mathbf{H}}\Big\{S_T(t|A=1,\mathbf{H}) - S_T(t|A=0,\mathbf{H})\Big\}dt \\ &= \int_0^{10}\int_0^1\int_0^1 f_{\mathbf{H}}(z,g)\{\exp[-t(\exp(-2.6+1.1z)+0.15g)] \\ &\quad - \exp[-t(\exp(-1.6+1.3z)+0.15g)]\}dzdgdt \\ &\approx 2.097. \end{aligned}$$

Similarly, we can calculate the true $\delta = 2.762$ for Scenario 4, $\delta = 1.854$ for Scenario 5 in **S1-S3** and $\delta = 1.854$ for Scenario 5 in **S4**.

References

Chen, P. and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*, 57(4):1030–1038.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*. John Wiley & Sons.

Hernán, M. A., Brumback, B., and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448.

Jiang, B., Li, J., and Fine, J. (2018). On two-step residual inclusion estimator for instrument variable additive hazards model. *Biostatistics & Epidemiology*, 2(1):47–60.

Karrison, T. (1987). Restricted mean life with adjustment for covariates. *Journal of the American Statistical Association*, 82(400):1169–1176.

Li, J., Fine, J., and Brookhart, A. (2015). Instrumental variable additive hazards models. *Biometrics*, 71(1):122–130.

Lin, D. Y., Wei, L.-J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730.

Lin, D. Y. and Ying, Z. L. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Annals of Statistics*, 23(5):1712–1734.

Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, volume 24, page 3. San Francisco CA.

Robins, J. M. and Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56(3):779–788.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pages 297–331. Springer.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Shows, J. H., Lu, W., and Zhang, H. H. (2010). Sparse estimation and inference for censored median regression. *Journal of statistical planning and inference*, 140(7):1903–1917.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine*, 16(4):385–395.

Tsiatis, A. A. (2006). *Semiparametric Theory And Missing Data*. Springer.

van der Laan, M. and Robins, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Spring-Verlag: New York.

Wey, A., Vock, D. M., Connett, J., and Rudser, K. (2016). Estimating restricted mean treatment effects with stacked survival models. *Statistics in Medicine*, 35(19):3319–3332.

Yang, S. (2013). Semiparametric inference on the absolute risk reduction and the restricted mean survival difference. *Lifetime Data Analysis*, 19(2):219–241.

Ying, A., Xu, R., and Murphy, J. (2019). Two-stage residual inclusion for survival data and competing risksan instrumental variable approach with application to seer-medicare linked data. *Statistics in Medicine*, 38(10):1775–1801.

Zhang, H. H. and Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika*, 94(3):691–703.

Zhang, M. and Schaubel, D. E. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*, 67(3):740–749.

Zhang, M. and Schaubel, D. E. (2012). Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics*, 68(4):999–1009.

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association*, 93(442):702–709.