Supporting Information for "A Semiparametric Cox-Aalen Transformation Model with Censored Data" by

Xi Ning[1], Yinghao Pan[1], Yanqing Sun[1,*], and Peter Gilbert[2,3]

[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, North Carolina

[2]Department of Biostatistics, University of Washington, Seattle, Washington

[3]Vaccine and Infectious Disease and Public Health Sciences Divisions, Fred Hutchinson Cancer Center, Seattle, Washington

**email*: yasun@charlotte.edu

In this supplementary material, we report additional simulation results and regression results for the HIV-1 example. We also provide detailed proofs of Theorems 1 and 2 of the main paper.

Web Appendix A: The proposed NPMLE under a special case

In Section 3.2 of the main paper, we proposed an NPMLE for the Cox-Aalen transformation model. More specifically, the NPMLE can be obtained via an EM-type algorithm where we solve the following set of equations in the M-step:

$$\sum_{i=1}^{n}\left(\Delta_i I(\widetilde{T}_i = t_k)\frac{X_{ik}}{X_{ik}^\top a_k} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})X_{ik}\right) = 0, \text{ for } k = 1, \ldots, m \quad \text{(S.1)}$$

and

$$\sum_{i=1}^{n}\sum_{k=1}^{m}\left\{\Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik})\right\} Z_{ik} = 0. \quad \text{(S.2)}$$

It is difficult to solve (S.1) and (S.2) jointly due to the curse of dimensionality. However, for a special when X is a vector of design variables for categories, there exist explicit formulae for calculating the high-dimensional parameters a_k $(k = 1, \ldots, m)$. Here, we give some further illustrations.

Let D be a categorical variable with q levels. Without loss of generality, we assume that D takes values in $\{1, \ldots, q\}$. Let $X = (1, X_2, \ldots, X_q)$ where $X_2, \ldots, X_q$ are group indicators, i.e., $X_2 = I(D = 2), \ldots, X_q = I(D = q)$. Here, $D = 1$ is considered as the reference group. We propose the following Gauss-Seidel method to jointly solve (S.1) and (S.2). Start with some initial values of the unknown parameters.

Step 1. Fix β, we update a_k, $(k = 1, \ldots, m)$ by solving (S.1). Note that for a fixed k, (S.1) can be written as

$$\begin{cases} \sum_{i=1}^{n} I(D_i = 1)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0 \\ \sum_{i=1}^{n} I(D_i = 2)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k} + a_{2k}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0 \\ \qquad\qquad\qquad\cdots \\ \sum_{i=1}^{n} I(D_i = q)\left\{\frac{\Delta_i I(\widetilde{T}_i = t_k)}{a_{1k} + a_{qk}} - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})\right\} = 0. \end{cases}$$

Hence, we obtain that

$$\begin{cases} a_{1k} = \frac{\sum_{i=1}^{n} I(D_i=1)\Delta_i I(\widetilde{T}_i=t_k)}{\sum_{i=1}^{n} I(D_i=1)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} \\ a_{2k} = \frac{\sum_{i=1}^{n} I(D_i=2)\Delta_i I(\widetilde{T}_i=t_k)}{\sum_{i=1}^{n} I(D_i=2)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} - a_{1k} \\ \qquad\qquad\qquad\qquad \cdots \\ a_{qk} = \frac{\sum_{i=1}^{n} I(D_i=q)\Delta_i I(\widetilde{T}_i=t_k)}{\sum_{i=1}^{n} I(D_i=q)I(\widetilde{T}_i\geq t_k)\hat{E}(\xi_i)\exp(\beta^\top Z_{ik})} - a_{1k}. \end{cases} \tag{S.3}$$

Step 2. Fix $a_1, \ldots, a_m$, we update β by solving (S.2) using the Newton-Raphson method.

We iterate between steps 1 and 2 until convergence.

Web Appendix B: Equivalence between the proposed ES and EM estimator under a special case

In this section, we show that when X is a vector of design variables for categories, the ES algorithm proposed in Section 3.3 coincides with the EM algorithm proposed in Section 3.2. To show this, we only need to show that for fixed β, equations (8) and (13) in the main paper share the same solution in terms of a_k $(k = 1, \ldots, m)$.

Let D be a categorical variable with q levels. Without loss of generality, we assume that D takes values in $\{1, \ldots, q\}$. Let $X = (1, X_2, \ldots, X_q)$ where $X_2, \ldots, X_q$ are group indicators, i.e., $X_2 = I(D = 2), \ldots, X_q = I(D = q)$. Here, $D = 1$ is considered as the reference group.

Note that for a fixed k, (13) can be written as

$$\sum_{i=1}^{n} \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k)\hat{E}(\xi_i)(X_{ik}^\top a_k)\exp(\beta^\top Z_{ik}) \right\} X_{ik} = 0,$$

which under this special case, is equivalent to

$$
\begin{cases}
\sum_{i=1}^n I(D_i = 1) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) a_{1k} \right\} = 0 \\
\sum_{i=1}^n I(D_i = 2) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) (a_{1k} + a_{2k}) \right\} = 0 \\
\qquad \qquad \qquad \ldots \\
\qquad \qquad \qquad \ldots \\
\sum_{i=1}^n I(D_i = q) \left\{ \Delta_i I(\widetilde{T}_i = t_k) - I(\widetilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) (a_{1k} + a_{qk}) \right\} = 0.
\end{cases}
\tag{S.4}
$$

It is easy to notice that (S.3) is the unique solution to (S.4) and thus also the unique solution to (13). In addition, we already showed in Web Appendix A that (S.3) is the unique solution to (8) under this special case. Thus, the ES and EM estimator coincide with each other when X is a vector of design variables for categories.

Web Appendix C: Proofs of Theorems 1 and 2

To establish the asymptotic properties of the proposed estimators, we assume the following regularity conditions:

Condition 1. With probability one, $X(\cdot)$ and $Z(\cdot)$ have bounded total variation in $[0, \tau]$.

Condition 2. Let $\mathcal{B}$ be a compact set of $\mathbb{R}^d$ and $BV[0, \tau]$ be the class of functions with bound variation over $[0, \tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0, \tau]$ with β_0 an interior point of $\mathcal{B}$ and $A_0(t) = (A_{01}(t), \ldots, A_{0q}(t))^\top$ is continuous over $[0, \tau]$ with $A_0(0) = 0$. Here $BV^q[0, \tau]$ denotes the product space $BV[0, \tau] \times \cdots \times BV[0, \tau]$.

Condition 3. With probability one, there exists a positive constant a such that $P(Y(\tau) = 1 \mid Z(\cdot), X(\cdot)) > a$ and $PN^2(\tau) < \infty$. If there exists a vector γ and a deterministic function $\gamma_0(t)$ such that $\gamma_0(t) + \gamma^\top X(t) = 0$ with probability one, then $\gamma = 0$ and $\gamma_0(t) = 0$.

Condition 4. The transformation function G is thrice continuously differentiable on $[0, \infty)$ with $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5. The map $\dot{\mathbf{\Psi}}_{\theta_0}$ defined in (S.8) is invertible, where $\theta_0 = (\beta_0, A_0)$.

C.1 Consistency

Theorem 1. *Under Conditions* $1 - 5$*, the proposed ES estimator* $(\hat{\beta}, \hat{A})$ *is strongly consistent to* (β_0, A_0).

Proof. Let $\phi(t) = G'(t)$, $\psi(t) = G''(t)/G'(t)$ and

$$\rho(t; \beta, A) = \int_0^t Y(s) e^{\beta^\top Z(s)} X^\top(s) dA(s).$$

Hence, the posterior mean of ξ can be written as

$$g(\tau; \beta, A) = \phi(\rho(\tau; \beta, A)) - \Delta\psi(\rho(\tau; \beta, A)).$$

Let P denote the true probability measure and $\mathbb{P}_n$ denote the empirical measure. Let $\theta = (\beta, A)$ and $\theta_0 = (\beta_0, A_0)$. Then the proposed ES estimator $\hat{\theta} = (\hat{\beta}, \hat{A})$ is essentially a Z-estimator solving the following observed-data estimating equation

$$\mathbb{P}_n \Phi(\beta, A)(t) \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \\ \Phi_2(\beta, A)(t) \end{pmatrix} = 0, \tag{S.5}$$

for $0 \le t \le \tau$, where

$$\Phi_1(\beta, A) = \int_0^\tau \left\{ Z(t) dN(t) - Y(t) e^{\beta^\top Z(t)} g(\tau; \beta, A) Z(t) X^\top(t) dA(t) \right\}, \tag{S.6}$$

and

$$\Phi_2(\beta, A)(t) = X(t) dN(t) - Y(t) e^{\beta^\top Z(t)} g(\tau; \beta, A) X(t) X^\top(t) dA(t).$$

Let h be a function in $BV_1[0,\tau]$, where $BV_1[0,\tau]$ denotes the set of functions with total variation bounded by 1 on $[0,\tau]$. Define

$$\Phi_2(\beta, A)[h] = \int_0^\tau h(t)\left\{X(t)dN(t) - Y(t)e^{\beta^\top Z(t)}g(\tau;\beta,A)X(t)X^\top(t)dA(t)\right\}. \tag{S.7}$$

Similar to Gao et al. (2017) and van der Vaart and Wellner (1996, Section 3.3.1), the proposed ES estimator $(\hat{\beta}, \hat{A})$ is equivalent to the root of the estimating equation

$$\mathbb{P}_n\Phi(\beta, A)[h] \equiv \mathbb{P}_n\begin{pmatrix}\Phi_1(\beta, A)\\ \\ \Phi_2(\beta, A)[h]\end{pmatrix} = 0,$$

for all $h \in BV_1[0,\tau]$. From (S.5), $\hat{A}$ is a step function with jumps at the observed failure time points t_k $(k = 1,\ldots,m)$. Write $\widetilde{h}(t) = \sum_{k=1}^m h(t_k)I(t_{k-1} < t \le t_k)$. Then the step function $\widetilde{h}$ can be written as a finite sum of simple functions, denoted as $\widetilde{h}(t) = \sum_{k=1}^m \alpha_k I(t_{k-1} < t \le t_k)$, where $\alpha_k = h(t_k)$. It is easy to see that $(\hat{\beta}, \hat{A})$ solves

$$\mathbb{P}_n\Phi_2(\hat{\beta}, \hat{A})[\widetilde{h}] = \sum_{k=1}^m \alpha_k \mathbb{P}_n\Phi_2(\hat{\beta}, \hat{A})(t_k) = 0.$$

The parameter of interest is $\theta = (\beta, A)$, where $A = (A_1,\ldots,A_q)^\top$. Let $\overline{\text{lin}}(BV_1[0,\tau])$ be the closed linear span for linear functionals of $BV_1[0,\tau]$. For each j $(j = 1,\ldots,q)$, A_j is contained in the Banach space $\overline{\text{lin}}(BV_1[0,\tau])$, where $A_j[h] = \int h(t)dA_j(t)$ for $h \in BV_1[0,\tau]$. The corresponding norm is defined as $\|A_j\|_\rho = \sup_{\|h\|_{BV}\le 1}|\int h(t)dA_j(t)|$, where $\|\cdot\|_{BV}$ is the bounded variation norm. Thus, $A = (A_1,\ldots,A_q)^\top$ is contained in the Banach space $\overline{\text{lin}}^q(BV_1[0,\tau])$ and we define $A[h] = \int h(t)dA(t) = (\int h(t)dA_1(t),\ldots,\int h(t)dA_q(t))^\top$ for $h \in BV_1[0,\tau]$. Here, $\overline{\text{lin}}^q(BV_1[0,\tau])$ stands for the product space $\overline{\text{lin}}(BV_1[0,\tau]) \times \cdots \times \overline{\text{lin}}(BV_1[0,\tau])$. Furthermore, the norm for A is defined as $\|A\|_{\mathcal{H}} = \sum_{j=1}^q \|A_j\|_\rho$ and the norm for θ is defined as $\|\theta\|_{\mathcal{V}} = \|\beta\|_d + \|A\|_{\mathcal{H}}$, where $\|\cdot\|_d$ is the Euclidean norm in $\mathbb{R}^d$ space. Hence, the function $\mathbb{P}_n\Phi(\beta, A)[h]$ is a map from $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$ to $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$.

Define $B_\delta(\beta_0, A_0) = \{(\beta, A) : \|\beta - \beta_0\|_d + \|A - A_0\|_{\mathcal{H}} < \delta\}$. We first show that the class of functions $\{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0, \tau]\}$ is P-Donsker for some fixed $\delta > 0$. Since $Y(t)$ and $N(t)$ are either cadlag or caglad functions in $l^\infty[0, \tau]$, they are both Donsker by Lemma 4.1 in Kosorok (2008). Trivially, Conditions 1 and 2 indicate that $\{\beta \in \mathcal{B}\}$, $\{Z(t), t \in [0, \tau]\}$ and $\{X(t), t \in [0, \tau]\}$ are all Donsker classes, and therefore so is $\{\beta^\top Z(t), \beta \in \mathcal{B}, t \in [0, \tau]\}$ since the products of bounded Donsker classes are Donsker. The class $\{e^{\beta^\top Z(t)}, \beta \in \mathcal{B}, t \in [0, \tau]\}$ is also Donsker since exponentiation is Lipschitz continuous on compacts. On the other hand, we rewrite

$$\rho(\tau; \beta, A) = \int_0^\tau Y(s) e^{\beta^\top Z(s)} X^\top(s) dA(s) = \sum_{j=1}^q \int_0^\tau Y(s) e^{\beta^\top Z(s)} X_j(s) dA_j(s).$$

Following Zeng et al. (2016), for any $j = 1, \dots, q$, if A_j is a monotone function, the class $\{\int_0^\tau Y(s) e^{\beta^\top Z(s)} X_j(s) dA_j(s) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is a Donsker class because it is a convex hull of functions $\{Y(s) \exp\{\beta^\top Z(s)\} X_j(s)\}$. By Condition 2, A_j ($j = 1, \dots, q$) can be expressed as the difference of pairs of monotonely increasing functions since it has bounded total variation over $[0, \tau]$. Thus, $\{\int_0^\tau Y(s) e^{\beta^\top Z(s)} X_j(s) dA_j(s) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is Donsker because the sums of bounded Donsker classes are also Donsker from Example 2.10.7 in van der Vaart and Wellner (1996). It follows immediately that $\{\rho(\tau; \beta, A) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is also a Donsker class. Similarly, the class $\left\{\int_0^\tau Y(t) e^{\beta^\top Z(t)} Z(t) X^\top(t) dA(t) : (\beta, A) \in B_\delta(\beta_0, A_0)\right\}$ is a Donsker class. By Condition 4, $G(x)$ is thrice continuously differentiable on $[0, \infty)$ and $G'(x) > 0$ for any $x \in [0, \infty)$, then

$$g(\tau; \beta, A) = G'(\rho(\tau; \beta, A)) - \Delta G''(\rho(\tau; \beta, A))/G'(\rho(\tau; \beta, A)),$$

is bounded for $(\beta, A) \in B_\delta(\beta_0, A_0)$. Moreover, $\{g(\tau; \beta, A) : (\beta, A) \in B_\delta(\beta_0, A_0)\}$ is a Donsker class due to the fact that any continuously differentiable function is locally Lipschitz and the preservation of the Donsker property under Lipschitz-continuous transformations by Theorem 9.31 in

Kosorok (2008). Notice that $\{h(\cdot) : h \in BV_1[0,\tau]\}$, $\{\int_0^\tau h(t)X(t)dN(t) : h \in BV_1[0,\tau]\}$ and

$$\left\{\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t): \quad (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\right\},$$

are all Donsker classes. This follows because the class of functions with an upper bound of their total variations is Donsker by Example 19.11 and Theorem 19.5 of van der Vaart (1998). Under Conditions $1-4$, now it is clear that the following classes

$$\begin{aligned}
&\left\{\int_0^\tau Z(t)dN(t)\right\}, \left\{\int_0^\tau h(t)X(t)dN(t): \quad h \in BV_1[0,\tau]\right\},\\
&\left\{g(\tau;\beta,A)\int_0^\tau Y(t)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t): \quad (\beta, A) \in B_\delta(\beta_0, A_0)\right\},\\
&\left\{g(\tau;\beta,A)\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t): \quad (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\right\},
\end{aligned}$$

are all Donsker classes. Therefore, the class of function

$$\{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\},$$

is P-Donsker as the sums of bounded Donsker classes are also Donsker.

To prove the local consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$, we use Theorem 1.20 (the implicit function theorem) in Schwartz (1969). For any $\theta = (\beta, A)$ in $B_\delta(\beta_0, A_0)$, write $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta) = \mathbb{P}_n\Phi(\beta, A)[h]$. Note that $\mathbf{\Psi}(\theta)$ and $\mathbf{\Psi}_n(\theta)$ are actually h-dependent. Rigorously speaking, we should write $\mathbf{\Psi}(\theta)[h] = P\Phi(\beta, A)[h]$ and $\mathbf{\Psi}_n(\theta)[h] = \mathbb{P}_n\Phi(\beta, A)[h]$, but in the rest of the article, we suppress the letter h in both $\mathbf{\Psi}(\theta)[h]$ and $\mathbf{\Psi}_n(\theta)[h]$ when there is no confusion. The Fréchet derivative of $\mathbf{\Psi}(\theta)$ with respect to θ at $\theta = \theta_0$ can be derived using (S.6) and (S.7). In particular, the Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)$ can be easily computed based on the weaker form

$$\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0) = \left.\frac{d\mathbf{\Psi}(\theta_0 + \eta(\theta - \theta_0))}{d\eta}\right|_{\eta=0} = \begin{pmatrix} C_{11}(\beta - \beta_0) + C_{12}(A - A_0) \\ \\ C_{21}(\beta - \beta_0) + C_{22}(A - A_0) \end{pmatrix}, \qquad \text{(S.8)}$$

where

$$
\begin{aligned}
&C_{11}(\beta-\beta_0)=B_1(\beta-\beta_0),\\
&C_{12}(A-A_0)=\int_0^\tau B_2(t)d(A-A_0),\\
&C_{21}(\beta-\beta_0)[h]=B_3[h](\beta-\beta_0),\\
&C_{22}(A-A_0)[h]=\int_0^\tau B_4[h](t)d(A-A_0).
\end{aligned}
\tag{S.9}
$$

Specifically,

$$
\begin{aligned}
B_1=&-E\left[g(\tau,\beta_0,A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)Z^\top(t)X^\top(t)dA_0(t)\right]\\
&-E\left[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\left\{\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right\}^{\bigotimes 2}\right],\\
B_2(t)=&-E\left[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right.\\
&\left.\times Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\right]\\
&-E\left\{g(\tau;\beta_0,A_0)Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right\},\\
B_3[h]=&-E\left[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)h(t)dA_0(t)\right.\\
&\left.\times\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z^\top(t)X^\top(t)dA_0(t)\right]\\
&-E\left[g(\tau;\beta_0,A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}X(t)Z^\top(t)X^\top(t)h(t)dA_0(t)\right],\\
B_4[h](t)=&-E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0))-\Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\
&\times\left\{\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)dA_0(t)\right\}Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg]\\
&-E\left\{h(t)Y(t)e^{\beta_0^\top Z(t)}g(\tau;\beta_0,A_0)X(t)X^\top(t)\right\}.
\end{aligned}
$$

Here, $a^{\bigotimes 2}=aa^\top$ for any column vector a. It can be shown that $||\boldsymbol{\Psi}(\theta)-\boldsymbol{\Psi}(\theta_0)-\dot{\boldsymbol{\Psi}}_{\theta_0}(\theta-\theta_0)||=o(||\theta-\theta_0||)$ as $\theta\to\theta_0$. Hence, $\boldsymbol{\Psi}(\theta)$ is Fréchet-differentiable at θ_0. The detailed calculations of the derivatives are given in Web Appendix C.3. Here we just present the corresponding results. Similarly, the Fréchet derivative of $\boldsymbol{\Psi}_n(\theta)=\mathbb{P}_n\Phi(\beta,A)[h]$ with respect to θ at $\theta=\theta_0$ can be derived and we use $\dot{\mathbf{\Psi}}_{\theta_0,n}$ to denote the corresponding derivative map. In particular, $\dot{\mathbf{\Psi}}_{\theta_0,n}$ can be obtained by replacing $\mathbf{\Psi}$ with $\mathbf{\Psi}_n$ in (S.8) and the expectations E in the terms B_1, $B_2(t)$, $B_3[h]$ and $B_4[h](t)$ with the empirical measure $\mathbb{P}_n$. Then one can obtain that $\|\mathbf{\Psi}_n(\theta) - \mathbf{\Psi}_n(\theta_0) - \dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|)$ as $\theta \to \theta_0$. Hence, $\mathbf{\Psi}_n(\theta)$ is also Fréchet-differentiable at θ_0. Clearly, the maps $\dot{\mathbf{\Psi}}_\theta$ and $\dot{\mathbf{\Psi}}_{\theta,n}$ depend continuously on θ in $B_\delta(\beta_0, A_0)$.

Next, we show that $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. Following the previous Donsker theory arguments, it can be shown that $\dot{\mathbf{\Psi}}_\theta(\theta^*)[h] - \dot{\mathbf{\Psi}}_{\theta,n}(\theta^*)[h] = o_p(1)$ uniformly in (θ, θ^*, h) in $B_\delta(\beta_0, A_0) \times \mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau]) \times BV_1[0,\tau]$ for some $\delta > 0$. By Condition 5, we know that $\dot{\mathbf{\Psi}}_{\theta_0}$ is invertible. Thus, there exists a constant $c_1 > 0$ such that $\|\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)\| \geq c_1\|\theta - \theta_0\|$ for any θ in $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$ by Lemma 6.16 in Kosorok (2008). Notice that there exists a positive constant c_2 such that

$$\left\| \frac{\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)}{\|\theta - \theta_0\|} \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0,n}\left(\frac{\theta - \theta_0}{\|\theta - \theta_0\|} \right) \right\| = \left\| \dot{\mathbf{\Psi}}_{\theta_0}\left(\frac{\theta - \theta_0}{\|\theta - \theta_0\|} \right) + o_p(1) \right\| \geq c_1 + o_p(1) \geq c_2,$$

as $n \to \infty$ for any θ in $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$. Thus, $\|\dot{\mathbf{\Psi}}_{\theta_0,n}(\theta - \theta_0)\| \geq c_2\|\theta - \theta_0\|$ as $n \to \infty$ for any θ in $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$. Hence, $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. In brief, we verified the three conditions, i.e., $\mathbf{\Psi}_n(\theta)$ is Fréchet-differentiable at θ_0, $\dot{\mathbf{\Psi}}_{\theta,n}$ depends continuously on θ in $B_\delta(\beta_0, A_0)$ and $\dot{\mathbf{\Psi}}_{\theta_0,n}$ is invertible for larger enough n. The implicit function theorem yields that for a sufficiently small $\delta > 0$, the map $\mathbf{\Psi}_n(\theta)$ is one-to-one from $B_\delta(\beta_0, A_0)$ onto a neighborhood of zero for large n.

Finally, we notice that $\{\Phi(\beta_0, A_0)[h] : h \in BV_1[0,\tau]\}$ is a Donsker class because

$$\{\Phi(\beta_0, A_0)[h] : h \in BV_1[0,\tau]\} \subset \{\Phi(\beta, A)[h] : (\beta, A) \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}.$$

Hence, $\mathbb{P}_n\Phi(\beta_0, A_0)[h] - P\Phi(\beta_0, A_0)[h] = o_p(1)$, or equivalently, $\mathbf{\Psi}_n(\theta_0) - \mathbf{\Psi}(\theta_0) = o_p(1)$. The martingale properties and double expectations yield

$$\mathbf{\Psi}(\theta_0) = P\Phi(\beta_0, A_0)[h] = P \begin{pmatrix} \Phi_1(\beta_0, A_0) \\ \\ \Phi_2(\beta_0, A_0)[h] \end{pmatrix} = 0.$$

Therefore, $\mathbf{\Psi}_n(\theta_0) = o_p(1)$. For an arbitrary small $\delta > 0$ and large n, by the implicit function theorem (Schwartz, 1969), there exists $\hat{\theta} = (\hat{\beta}, \hat{A})$ with $(\|\hat{\beta} - \beta_0\|_d + \|\hat{A} - A_0\|_{\mathcal{H}}) < \delta$ and $\mathbf{\Psi}_n(\hat{\theta}) = \mathbb{P}_n\Phi(\hat{\beta}, \hat{A})[h] = 0$ for any $h \in BV_1[0, \tau]$. This proves the consistency of $\hat{\theta} = (\hat{\beta}, \hat{A})$. $\square$

C.2 Asymptotic Normality

Theorem 2. *Under the Conditions* $1 - 5$*,* $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ *converges weakly to a zero-mean Gaussian process in the metric space* $\mathbb{R}^d \times \overline{lin}^q(BV_1[0, \tau])$.

Proof. We appeal to verify the conditions in Theorem 3.3.1 and Lemma 3.3.5 of van der Vaart and Wellner (1996). Write

$$\begin{aligned} \mathbb{G}_n\Phi(\theta)[h] &= n^{1/2}\left\{\mathbb{P}_n\Phi(\theta)[h] - P\Phi(\theta)[h]\right\} \\ &= n^{1/2}\left\{\mathbf{\Psi}_n(\theta)[h] - \mathbf{\Psi}(\theta)[h]\right\}, \end{aligned}$$

where $\mathbf{\Psi}_n(\theta)[h] = \mathbb{P}_n\Phi(\theta)[h]$ and $\mathbf{\Psi}(\theta)[h] = P\Phi(\theta)[h]$. We prove the asymptotic normality of the proposed ES estimator by the following four steps:

(1) Show that $\mathbb{G}_n\Phi(\theta_0)[h] = n^{1/2}\left\{\mathbf{\Psi}_n(\theta_0)[h] - \mathbf{\Psi}(\theta_0)[h]\right\}$ converges in distribution to a tight random element $\mathbf{W}$ in $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0, \tau])$.

Under Conditions $1 - 4$, we have

$$\sup_{h \in BV_1[0,\tau]} \|\mathbf{\Psi}(\theta_0)[h]\| < \infty.$$

Because $\{\Phi(\theta_0)[h] : h \in BV_1[0,\tau]\}$ is a Donsker class, $\mathbb{G}_n\Phi(\theta_0)[h]$ converges weakly to a Gaussian process $\mathbf{W}$ in $\mathbb{R}^d \times \overline{\text{lin}}^q(BV_1[0,\tau])$.

(2) Verify $\mathbf{\Psi}(\theta)$ is Fréchet differentiable as a function of θ at $\theta = \theta_0$.

The Fréchet-differentiablility of $\mathbf{\Psi}(\theta)$ can be checked directly. In particular, the Fréchet derivative $\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0)$ can be easily computed based on the weaker form

$$\dot{\mathbf{\Psi}}_{\theta_0}(\theta - \theta_0) = \frac{d\mathbf{\Psi}(\theta_0 + \eta(\theta - \theta_0))}{d\eta}\bigg|_{\eta=0} = \begin{pmatrix} C_{11}(\beta - \beta_0) + C_{12}(A - A_0) \\ \\ C_{21}(\beta - \beta_0) + C_{22}(A - A_0) \end{pmatrix},$$

where each of the components is given in (S.9). The detailed calculations are shown in Web Appendix C.3.

(3) To verify the condition (3.3.4) in van der Vaart and Wellner (1996), it's sufficient to verify the conditions in Lemma 3.3.5 of van der Vaart and Wellner (1996).

From Conditions $1-4$ and the results derived in the proof of Theorem 1, we have shown that $\{\Phi(\theta)[h] : \theta \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}$ and $\{\Phi(\theta_0)[h] : h \in BV_1[0,\tau]\}$ both are Donsker classes. Thus,

$$\{\Phi(\theta)[h] - \Phi(\theta_0)[h] : \theta \in B_\delta(\beta_0, A_0), h \in BV_1[0,\tau]\}$$

is also a Donsker class for some $\delta > 0$. In view of the dominated convergence theorem, to show

$$\sup_{h \in BV_1[0,\tau]} P\{\Phi(\theta)[h] - \Phi(\theta_0)[h]\}^2 \to 0,$$

as $\theta \to \theta_0$, it is valid to show that $\Phi(\theta)[h]$ converges to $\Phi(\theta_0)[h]$ pointwise, uniformly in h. This condition is satisfied because $h(t)$ has bounded total variation over $[0,\tau]$ and $\Phi(\theta)[h]$ is a continuous function over θ under Conditions $1-4$. Because $\hat{\theta}$ converges to θ_0 almost surely, it follows from Lemma 3.3.5 of van der Vaart and Wellner (1996) that

$$\|\mathbb{G}_n(\Phi(\hat{\theta}) - \Phi(\theta_0))\| = o_{p^*}(1 + n^{1/2}\|\hat{\theta} - \theta_0\|), \tag{S.10}$$

where $o_{p^*}(1)$ denotes converging to zero in outer probability.

(4) Equation (S.10) can be written as

$$n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\hat{\theta}) - n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) = o_{p^*}(1 + n^{1/2}\|\hat{\theta} - \theta_0\|).$$

By the definition of θ_0 and $\hat{\theta}$, $\mathbf{\Psi}(\theta_0) = 0$ and $\mathbf{\Psi}_n(\hat{\theta}) = 0$. It follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that

$$n^{1/2}\dot{\mathbf{\Psi}}_{\theta_0}(\hat{\theta} - \theta_0) = -n^{1/2}(\mathbf{\Psi}_n - \mathbf{\Psi})(\theta_0) + o_{p^*}(1).$$

Finally, Condition 5 and the continuous mapping theorem give

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow -\dot{\mathbf{\Psi}}_{\theta_0}^{-1}\mathbf{W}.$$

□

C.3 The Fréchet Derivative Map

This subsection provides details on the calculation of the Fréchet derivative of $\mathbf{\Psi}(\theta) = P\Phi(\beta, A)$. The Fréchet derivative of $P\Phi(\beta, A)$ at (β_0, A_0) is given by the map

$$(\beta - \beta_0, A - A_0) \rightarrow \begin{pmatrix} C_{11} & C_{12} \\ \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} \beta - \beta_0 \\ \\ A - A_0 \end{pmatrix},$$

where

$$\begin{aligned}
C_{11}(\beta - \beta_0) &= B_1(\beta - \beta_0), \\
C_{12}(A - A_0) &= \int_0^\tau B_2(t) d(A - A_0), \\
C_{21}(\beta - \beta_0)[h] &= B_3[h](\beta - \beta_0), \\
C_{22}(A - A_0)[h] &= \int_0^\tau B_4[h](t) d(A - A_0).
\end{aligned}$$

Specifically,

$$
\begin{aligned}
B_1 &= \frac{\partial P\Phi_1(\beta, A)}{\partial \beta}\bigg|_{\beta=\beta_0, A=A_0} \\
&= -E\left[g(\tau, \beta_0, A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)Z^\top(t)X^\top(t)dA_0(t)\right] \\
&\quad - E\left[\{\phi'(\rho(\tau;\beta_0, A_0)) - \Delta\psi'(\rho(\tau;\beta_0, A_0))\}\left\{\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right\}^{\otimes 2}\right].
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{\partial g(\tau;\beta, A+\eta A^*)}{\partial \eta}\bigg|_{\eta=0} &= \{\phi'(\rho(\tau;\beta, A+\eta A^*)) - \Delta\psi'(\rho(\tau;\beta, A+\eta A^*))\}\,\rho(\tau;\beta, A^*)\bigg|_{\eta=0} \\
&= \{\phi'(\rho(\tau;\beta, A)) - \Delta\psi'(\rho(\tau;\beta, A))\}\,\rho(\tau;\beta, A^*)
\end{aligned}
$$

where $\phi'(\cdot)$ and $\psi'(\cdot)$ are the first derivative of $\phi(\cdot)$ and $\psi(\cdot)$, respectively. Then,

$$
\begin{aligned}
\frac{\partial \Phi_1(\beta, A+\eta A^*)}{\partial \eta}\bigg|_{\eta=0} &= -\int_0^\tau Y(t)e^{\beta^\top Z(t)}\frac{\partial g(\tau;\beta, A+\eta A^*)}{\partial \eta}Z(t)X^\top(t)d(A+\eta A^*)(t)\bigg|_{\eta=0} \\
&\quad - \int_0^\tau Y(t)e^{\beta^\top Z(t)}g(\tau;\beta, A+\eta A^*)Z(t)X^\top(t)dA^*(t)\bigg|_{\eta=0} \\
&= -\{\phi'(\rho(\tau;\beta, A)) - \Delta\psi'(\rho(\tau;\beta, A))\}\int_0^\tau Y(t)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA(t) \\
&\quad \times \int_0^\tau Y(t)e^{\beta^\top Z(t)}X^\top(t)dA^*(t) \\
&\quad - g(\tau;\beta, A)\int_0^\tau Y(t)e^{\beta^\top Z(t)}Z(t)X^\top(t)dA^*(t).
\end{aligned}
$$

Thus, we can obtain

$$
\begin{aligned}
&\frac{\partial P\Phi_1(\beta, A+\eta A^*)}{\partial \eta}\bigg|_{\eta=0, A^*=A-A_0, \beta=\beta_0, A=A_0} \\
&= -E\left[\{\phi'(\rho(\tau;\beta_0, A_0)) - \Delta\psi'(\rho(\tau;\beta_0, A_0))\}\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right. \\
&\quad \left.\times \int_0^\tau Y(t)e^{\beta_0^\top Z(t)}X^\top(t)d(A-A_0)(t)\right] \\
&\quad - E\left[g(\tau;\beta_0, A_0)\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)d(A-A_0)(t)\right].
\end{aligned}
$$

Hence,

$$\begin{aligned}B_2(t) = &- E\left[\{\phi'(\rho(\tau;\beta_0,A_0)) - \Delta\psi'(\rho(\tau;\beta_0,A_0))\}\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)dA_0(t)\right.\\ &\left.\times Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\right]\\ &- E\left\{g(\tau;\beta_0,A_0)Y(t)e^{\beta_0^\top Z(t)}Z(t)X^\top(t)\right\}.\end{aligned}$$

Similarly,

$$\begin{aligned}B_3[h] &= \left.\frac{\partial P\Phi_2(\beta,A)[h]}{\partial\beta}\right|_{\beta=\beta_0,A=A_0}\\ &= -E\left[\{\phi'(\rho(\tau;\beta_0,A_0)) - \Delta\psi'(\rho(\tau;\beta_0,A_0))\}\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)dA_0(t)\right.\\ &\quad\left.\times\int_0^\tau Y(t)e^{\beta_0^\top Z(t)}Z^\top(t)X^\top(t)dA_0(t)\right]\\ &\quad - E\left[g(\tau;\beta_0,A_0)\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)Z^\top(t)X^\top(t)dA_0(t)\right].\end{aligned}$$

Lastly,

$$\begin{aligned}\left.\frac{\partial\Phi_2(\beta,A+\eta A^*)[h]}{\partial\eta}\right|_{\eta=0} = &- \int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}\frac{\partial g(\tau;\beta,A+\eta A^*)}{\partial\eta}X(t)X^\top(t)d(A+\eta A^*)(t)\Big|_{\eta=0}\\ &- \int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}g(\tau;\beta,A+\eta A^*)X(t)X^\top(t)dA^*(t)\Big|_{\eta=0}\\ = &- \{\phi'(\rho(\tau;\beta,A)) - \Delta\psi'(\rho(\tau;\beta,A))\}\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA(t)\\ &\times\int_0^\tau Y(t)e^{\beta^\top Z(t)}X^\top(t)dA^*(t)\\ &- g(\tau;\beta,A)\int_0^\tau h(t)Y(t)e^{\beta^\top Z(t)}X(t)X^\top(t)dA^*(t).\end{aligned}$$

Hence,

$$\left.\frac{\partial P\Phi_2(\beta,A+\eta A^*)[h]}{\partial\eta}\right|_{\eta=0,A^*=A-A_0,\beta=\beta_0,A=A_0} = \int_0^\tau B_4[h](t)d(A-A_0)(t),$$

where

$$\begin{aligned}B_4[h](t) = &- E\Bigg[\{\phi'(\rho(\tau;\beta_0,A_0)) - \Delta\psi'(\rho(\tau;\beta_0,A_0))\}\\ &\times\left\{\int_0^\tau h(t)Y(t)e^{\beta_0^\top Z(t)}X(t)X^\top(t)dA_0(t)\right\}Y(t)e^{\beta_0^\top Z(t)}X^\top(t)\Bigg]\\ &- E\left\{h(t)Y(t)e^{\beta_0^\top Z(t)}g(\tau;\beta_0,A_0)X(t)X^\top(t)\right\}.\end{aligned}$$

Web Appendix D: Additional simulation results

In this section, we provide additional simulation results for Scenarios $1-4$ of the main paper, and consider Scenario 5 to show the advantages of the proposed model over Zeng and Lin's model when there exist additive covariate effects. Recall that Scenarios $1-4$ are:

Scenario 1. $X = (1, X_2)^\top$ with $X_2 \sim$ Ber(0.4), $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 2. $X = (1, X_2)^\top$ with $X_2 \sim$ Unif$(0, 1)$, $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 3. $X(t) = (1, X_2(t))^\top$ with $X_2(t) = B_3 + B_4 t$, where $B_3 \sim$ Unif$(1, 2)$ and $B_4 \sim$ Unif$(0.1, 0.5)$, $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 4. Let D be a categorical variable that takes values in $\{1, 2, 3\}$ with equal probability. $X = (1, X_2, X_3)^\top$, where $X_2 = I(D = 2)$, $X_3 = I(D = 3)$, $A_1(t) = \log(1 + t/4)$, $A_2(t) = 0.1t$ and $A_3(t) = 0.05t$.

Note that throughout all four scenarios, we chose bandwidth $h = 0.1$ for the kernel estimator $\hat{\alpha}(t)$. Web Figure 1 shows the estimation results for $\alpha(\cdot)$ under Scenario 1. Web Figure 2 and 3 give the estimation results for $A(\cdot)$ and $\alpha(\cdot)$ under Scenario 2, respectively. Similarly, Web Figure 4 and 5 present the estimation results for $A(\cdot)$ and $\alpha(\cdot)$ under Scenario 4, respectively. For each figure, Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the standard error estimator, and empirical coverage percentage of the 95% confidence interval. Web Table 1 reports the parameter estimation results under Scenario 3 with r misspecified as 0 while the data is generated from $r = r_{true}$. Here, r_{true} can be any value from $\{0, 0.5, 1, 1.5, 2, 2.5, 3\}$. It is easy to see that the misspecification of r values led to biased estimates and lower coverage probabilities than the nominal levels, even though the proposed variance estimators can accurately reflect the true variations.

In addition, we considered Scenario 5 as shown below, and generated the failure time T from the following Cox-Aalen transformation model

$$\Lambda(t \mid X, Z) = G\Big\{ \int_0^t \exp(\beta_1 Z_1) X^\top dA(s) \Big\}. \tag{S.11}$$

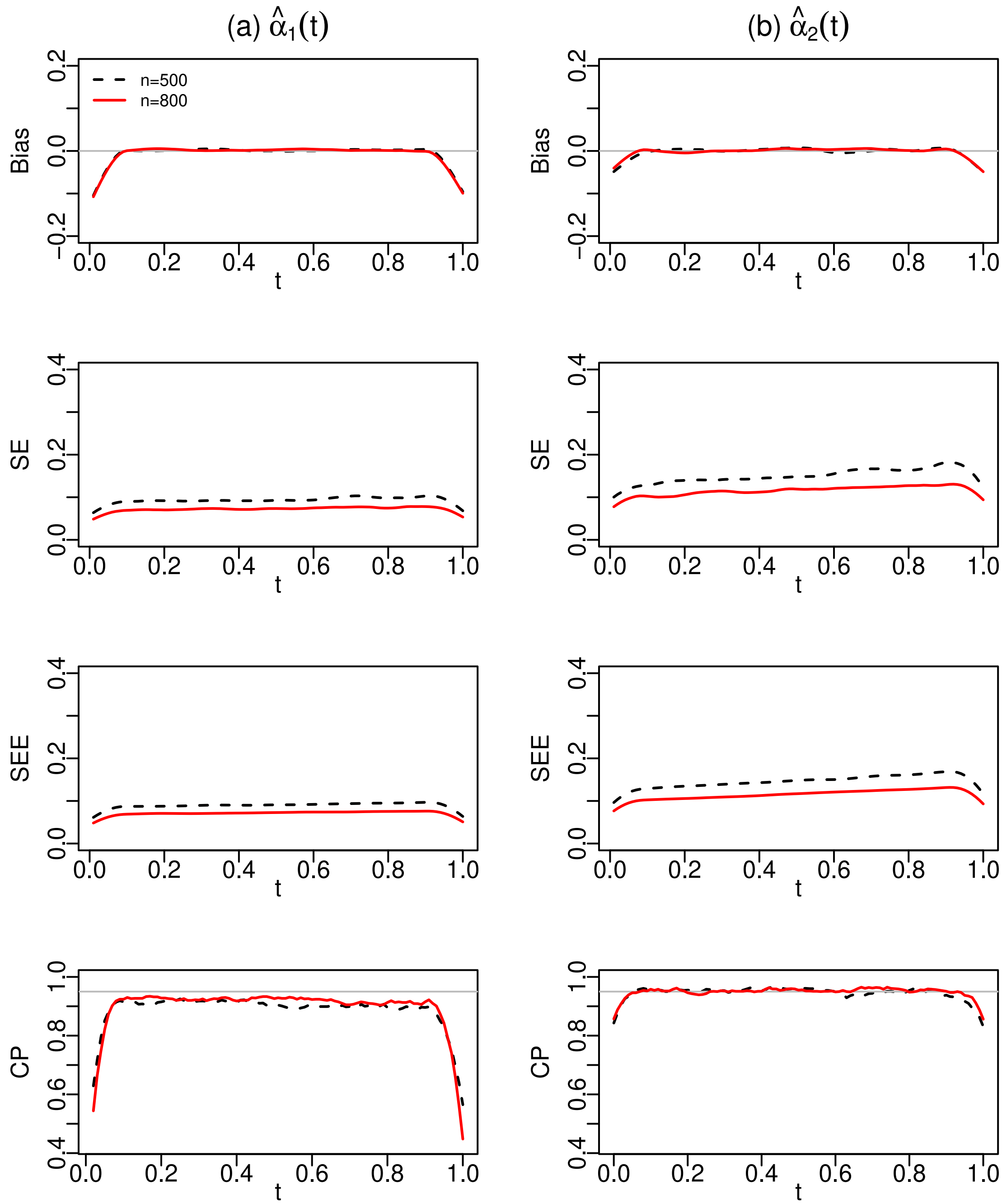
Scenario 5. $\beta_1 = 1$, $Z_1 \sim$ Ber(0.5) and $X(t) = (1, X_2)^\top$ with $X_2 \sim$ Ber(0.4). $A_1(t) = 0.1t + t^4/2$ and $A_2(t) = -2t^3/3 + t^2/2$.

Clearly, in model (S.11), Z_1 has a multiplicative effect while X_2 has an additive effect. If we naively treat all covariate effects as multiplicative and fit Zeng and Lin's model, we will obtain biased survival probability predictions. Web Figure 6 illustrates this bias. We also compared the predicted cumulative hazards from the proposed model and Zeng and Lin's model, displayed in Web Figure 7. When $Z_1 = 0$, the cumulative hazards for groups $X_2 = 0$ and $X_2 = 1$ intersect, indicating that the cumulative hazard in $X_2 = 1$ group is initially larger than the $X_2 = 0$ group, but becomes smaller later in the study. However, using Zeng and Lin's model, the cumulative hazard in $X_2 = 0$ is consistently larger than the group $X_2 = 1$. Hence, the proposed model can more accurately capture the complexity of the cumulative hazards when there exist additive covariate effects.

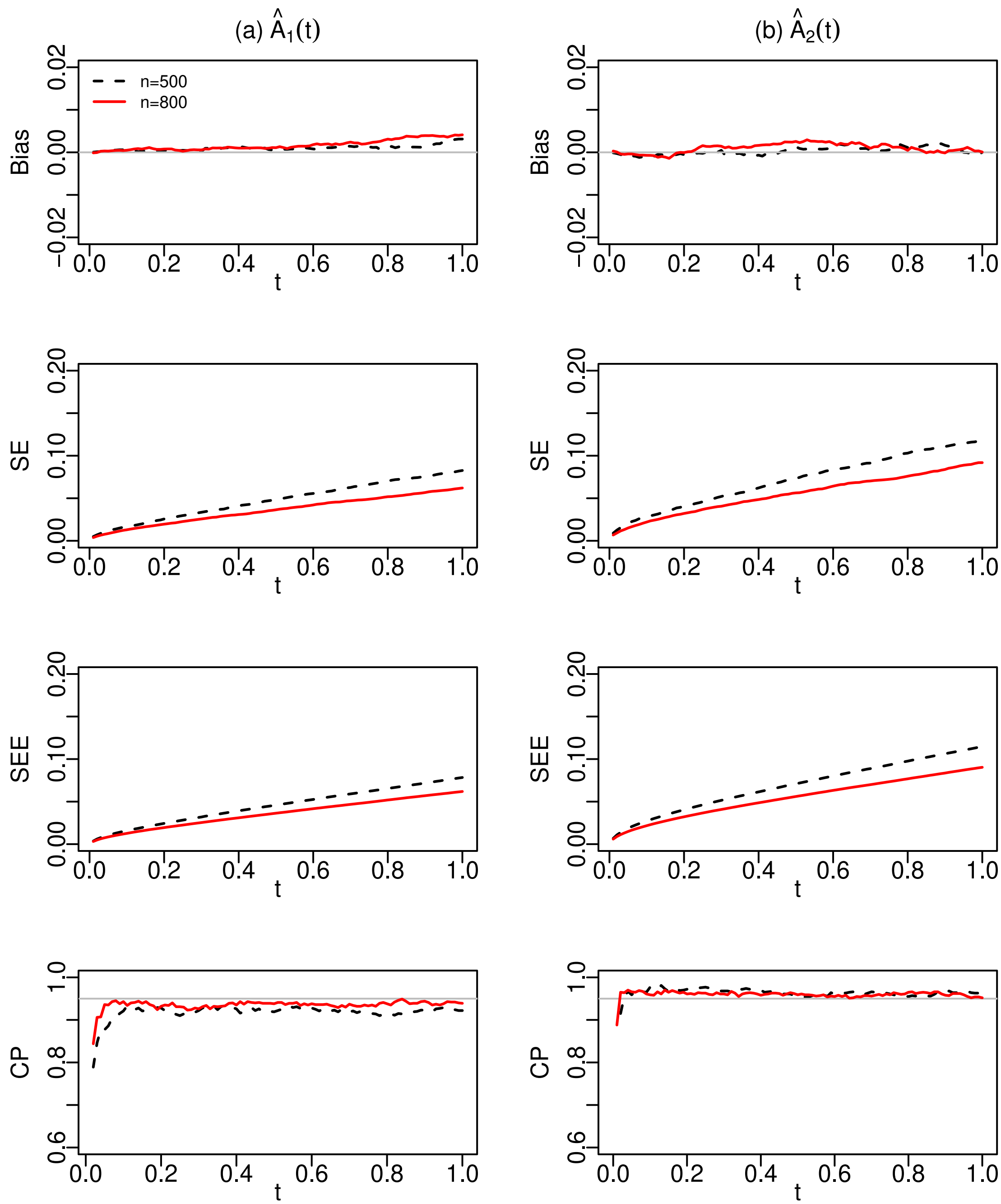
Web Table 1: Simulation results for estimation of the regression parameters with a misspecified $r = 0$ under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$

		$\beta_1 = 0.5$				$\beta_2 = -0.5$			
n	r_{true}	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					Scenario 3				
500	0	0.002	0.180	0.180	0.949	-0.004	0.305	0.305	0.953
	0.5	-0.032	0.186	0.185	0.950	0.034	0.318	0.315	0.951
	1	-0.059	0.192	0.192	0.943	0.067	0.329	0.326	0.947
	1.5	-0.082	0.199	0.197	0.933	0.086	0.332	0.335	0.950
	2	-0.098	0.205	0.201	0.917	0.101	0.344	0.343	0.946
	2.5	-0.118	0.212	0.206	0.906	0.117	0.357	0.352	0.932
	3	-0.133	0.216	0.210	0.888	0.129	0.363	0.359	0.936
2000	0	0.001	0.088	0.089	0.951	-0.006	0.149	0.152	0.965
	0.5	-0.031	0.094	0.092	0.933	0.033	0.165	0.157	0.938
	1	-0.058	0.098	0.095	0.904	0.058	0.165	0.162	0.941
	1.5	-0.078	0.097	0.098	0.872	0.084	0.167	0.167	0.929
	2	-0.094	0.100	0.100	0.839	0.103	0.168	0.171	0.907
	2.5	-0.110	0.102	0.102	0.812	0.117	0.171	0.175	0.904
	3	-0.122	0.105	0.105	0.786	0.130	0.175	0.179	0.895

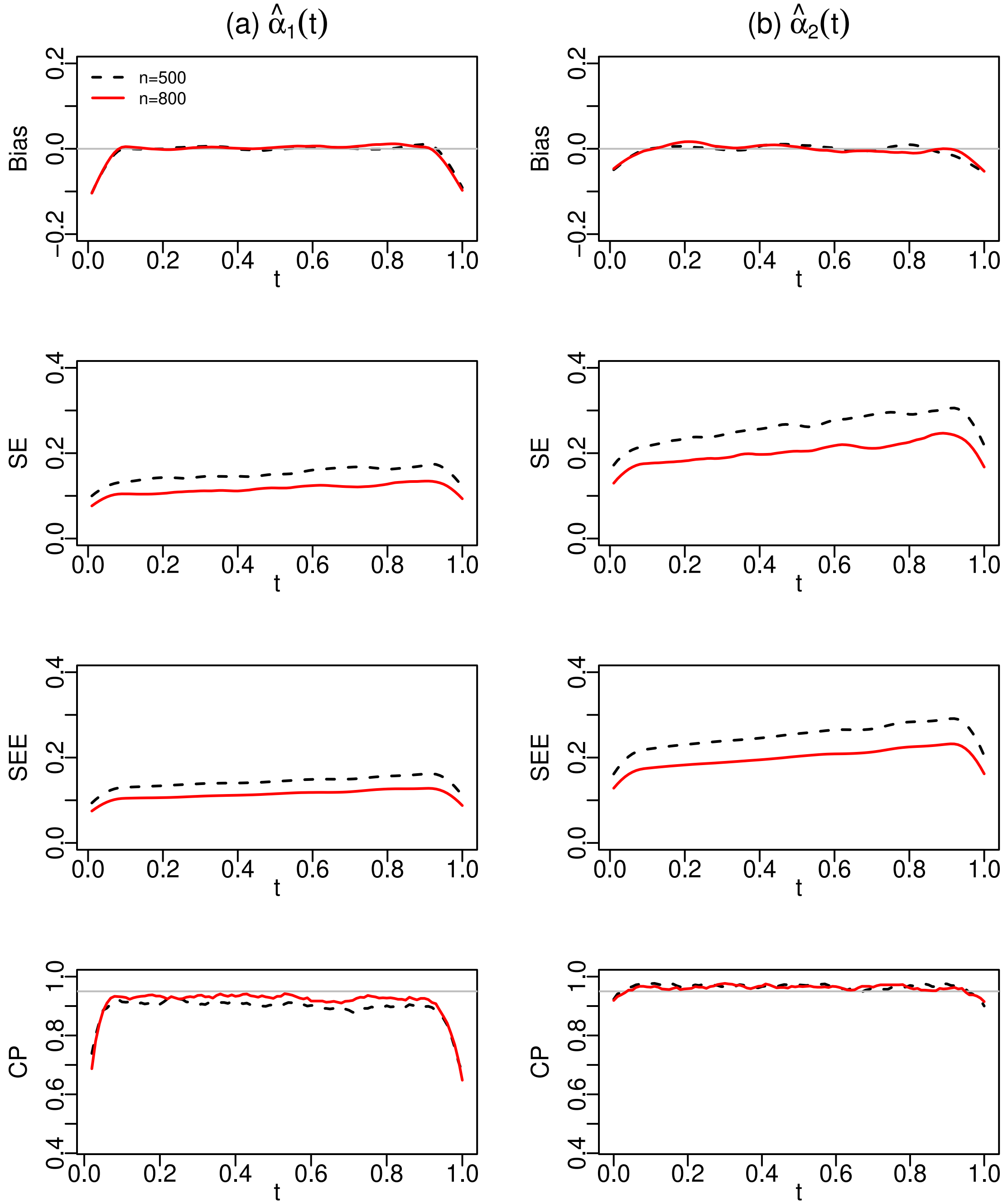
Note: Bias, bias of the parameter estimator; SE, empirical standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, empirical coverage percentage of the 95% confidence interval.



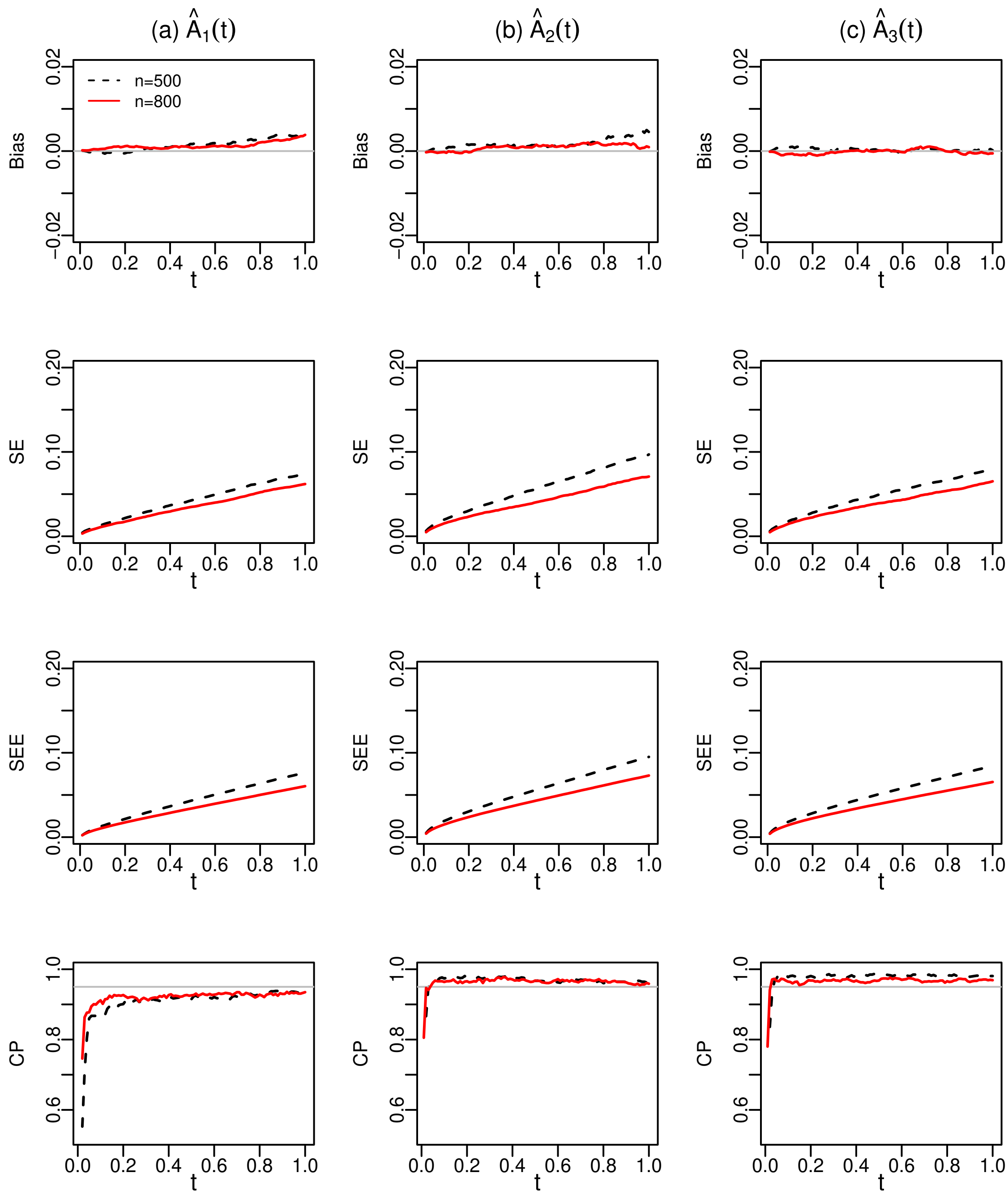
Web Figure 1: Estimation results for (a) $\alpha_1(t) = 1/(4+t)$ and (b) $\alpha_2(t) = 0.1$ in Scenario 1, under logarithmic transformation $G(x) = r^{-1}\log(1+rx)$ with $r = 0$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.



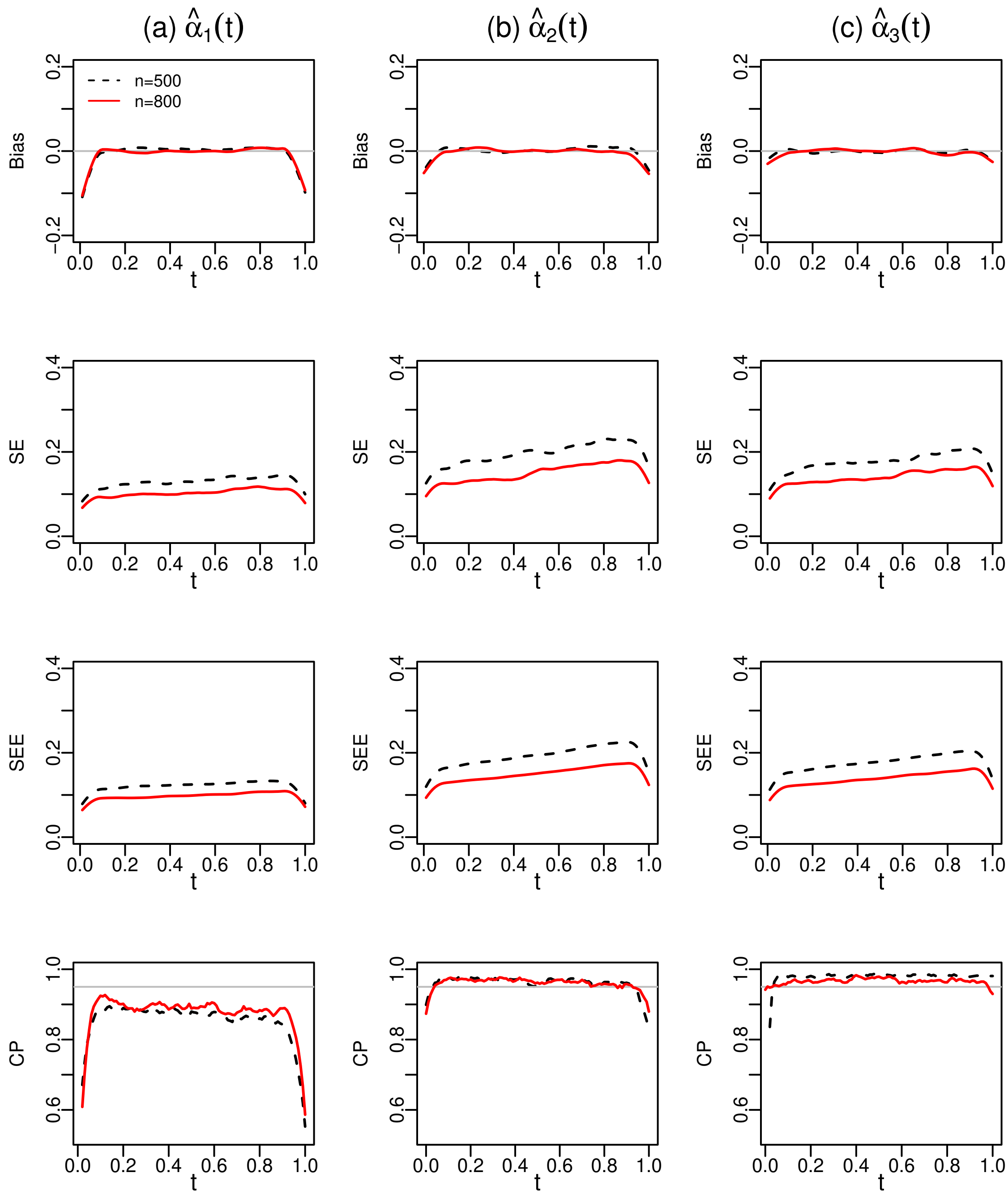
Web Figure 2: Estimation results for (a) $A_1(t) = \log(1 + t/4)$ and (b) $A_2(t) = 0.1t$ in Scenario 2, under logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 0.5$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.



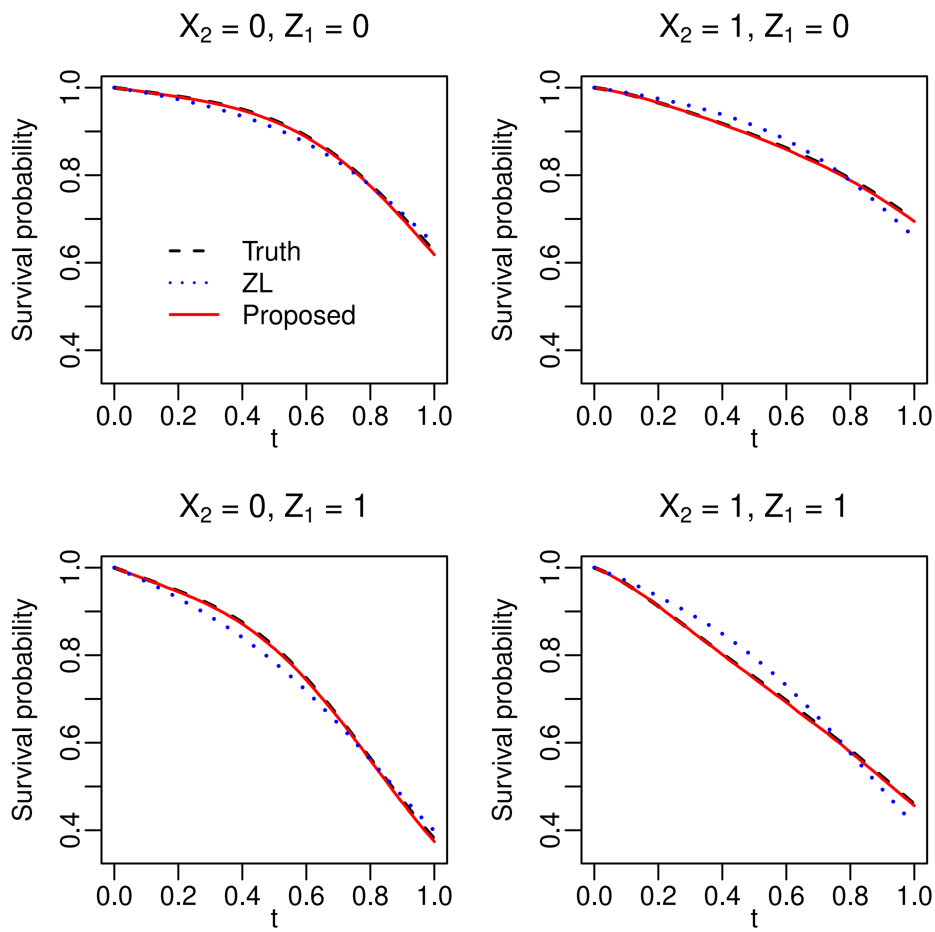
Web Figure 3: Estimation results for (a) $\alpha_1(t) = 1/(4+t)$ and (b) $\alpha_2(t) = 0.1$ in Scenario 2, under logarithmic transformation $G(x) = r^{-1}\log(1+rx)$ with $r = 0.5$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.



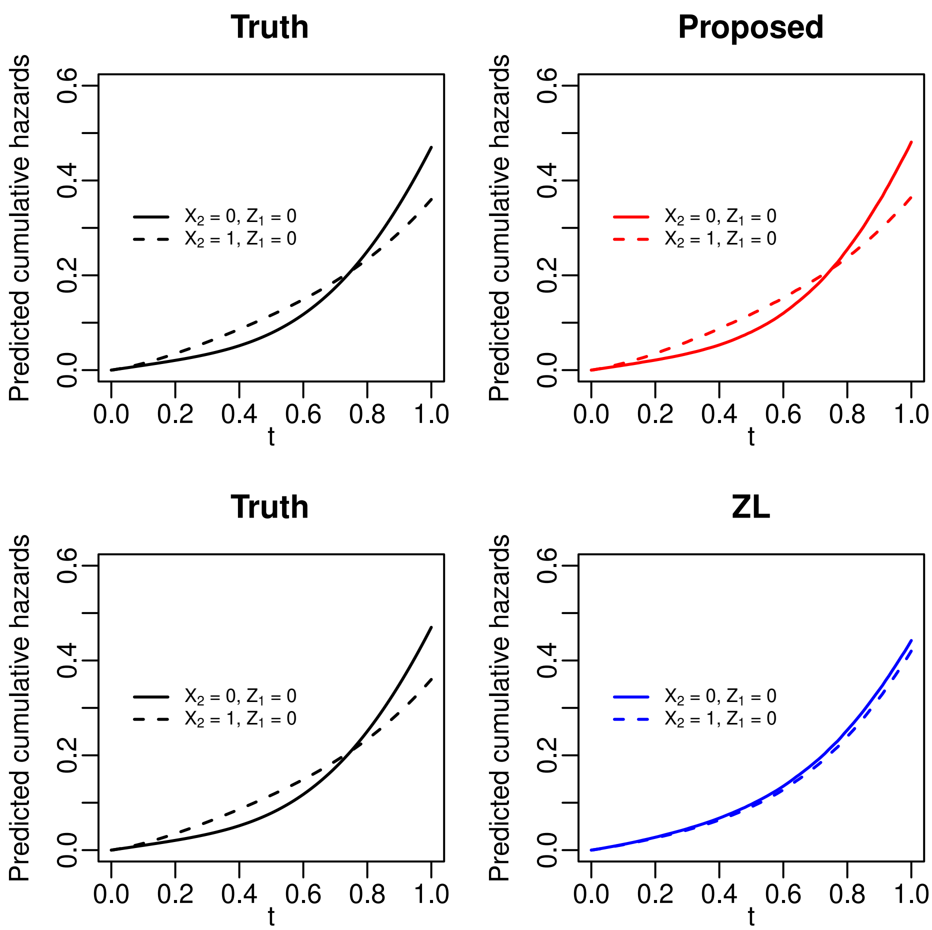
Web Figure 4: Estimation results for (a) $A_1(t) = \log(1 + t/4)$, (b) $A_2(t) = 0.1t$ and (c) $A_3(t) = 0.05t$ in Scenario 4, under logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 1$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.



Web Figure 5: Estimation results for (a) $\alpha_1(t) = 1/(4 + t)$, (b) $\alpha_2(t) = 0.1$ and (c) $\alpha_3(t) = 0.05$ in Scenario 4, under logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 1$. The dashed and solid lines represent $n = 500$ and $n = 800$, respectively.



Web Figure 6: Predicted survival probability under Scenario 5 with $n = 200$ and $r = 1$ based on the proposed model and Zeng and Lin's model. Here, "ZL" stands for Zeng and Lin's model.



Web Figure 7: Predicted cumulative hazards under Scenario 5 with $n = 200$ and $r = 1$ based on the proposed model and Zeng and Lin's model. Here, "ZL" stands for Zeng and Lin's model.

Web Appendix E: Additional results for the HIV-1 trials

This section provides additional results in Section 5 of the main paper. Web Table 2 summarizes the number of participants across different treatment groups and regions. The corresponding HIV-1 infections are provided in parentheses. Web Figure 8 plots the log-likelihood function against r under the class of transformation functions $G(x) = r^{-1}\log(1 + rx)$. Web Table 3 and Web Figure 9, respectively, show the estimated regression coefficients and the estimated baseline cumulative hazard functions under different Cox-Aalen transformation models ($r = 0$, 1 and 2). Web Figure 10 shows that the effects of "region" cross within the participants in South Africa and other sub-Saharan African countries, but Zeng and Lin's model cannot fully represent such effects, which further confirms the flexibility of the proposed model.

Finally, we conduct a simulation study to verify that the log-likelihood values vary slightly over the r values under a high percentage censoring rate. Specifically, suppose that the failure time T follows the Cox-Aalen transformation model

$$\Lambda(t) = G\left\{\int_0^t \exp(\beta_1 Z_1) d\Lambda_X(s)\right\}, \tag{S.12}$$

where $\beta_1 = 1$, $Z_1 \sim$ Unif$(0, 1)$, $X = (1, X_2)^\top$ with $X_2 \sim$ Ber(0.5), $A_1(t) = t^2/2$ and $A_2 = 0.1t$. Let $\tau = 1$. We generate one censoring time $C \sim$ Exponential(b) such that $b = 0.3$ and $b = 7$ yield a censoring rate around 50% and 95%, respectively. Considering the logarithmic transformation function $G(x) = r^{-1}\log(1 + rx)$, we generate the data from the model (S.12) with $r = 0$ and then fit the generated data with r values ranging from 0 to 3 with an increment of 0.1. Note that $r = 0$ can be considered as the true model while other r values are misspecified. Web Figure 11 plots the average log-likelihood against r across 200 replicates. One can see that the true value $r = 0$ is indeed the one that maximizes the log-likelihood. However, under high censoring percentage, the log-likelihood changes very slowly with different values of r.

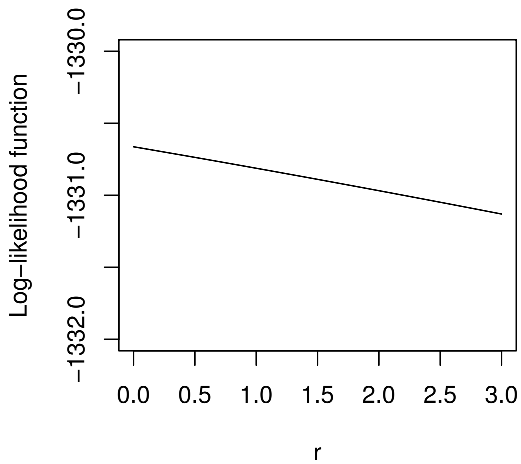
Web Table 2: Summary statistics for the number of participants across different treatment groups and regions. The number of HIV-1 infections is provided in parentheses.

Regions	Placebo	Low-dose	High-dose	Total
USA&Switzerland	467(9)	469(8)	465(6)	1401(23)
Brazil&Peru	420(29)	412(24)	417(22)	1249(75)
South Africa	335(16)	337(16)	337(11)	1009(43)
Other sub-Saharan African countries	297(13)	302(12)	301(8)	900(33)
Total	1519(67)	1520(60)	1520(47)	4559(174)

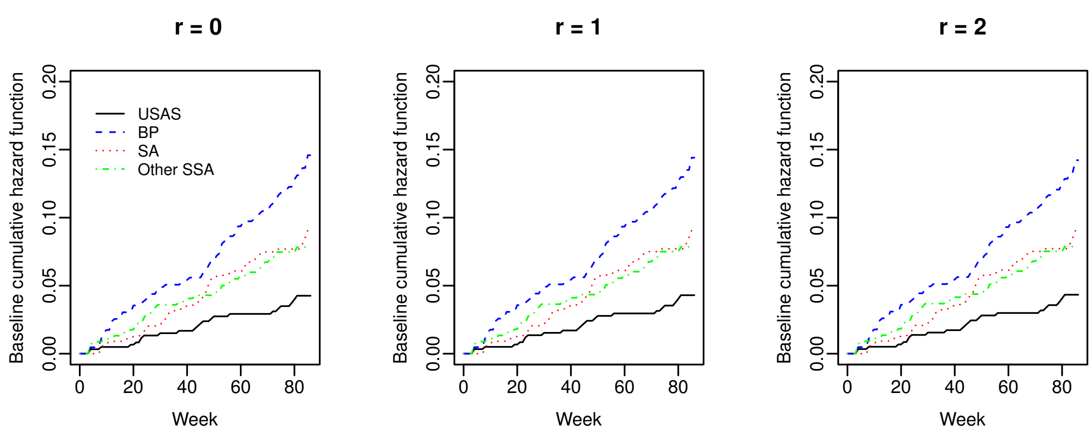
Web Table 3: Regression analysis results in the HIV-1 trials under the proposed model

	$r = 0$			$r = 1$			$r = 2$		
Covariates	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
Low-dose	-0.108	0.178	0.542	-0.118	0.183	0.519	-0.128	0.190	0.501
High-dose	-0.363	0.190	0.056	-0.379	0.195	0.052	-0.395	0.202	0.050
$21-30$	-0.429	0.187	0.022	-0.440	0.194	0.023	-0.450	0.204	0.027
$31-40$	-1.219	0.274	< 0.001	-1.246	0.280	< 0.001	-1.272	0.288	< 0.001
$41-52$	-1.989	0.721	0.006	-2.016	0.730	0.006	-2.042	0.740	0.006

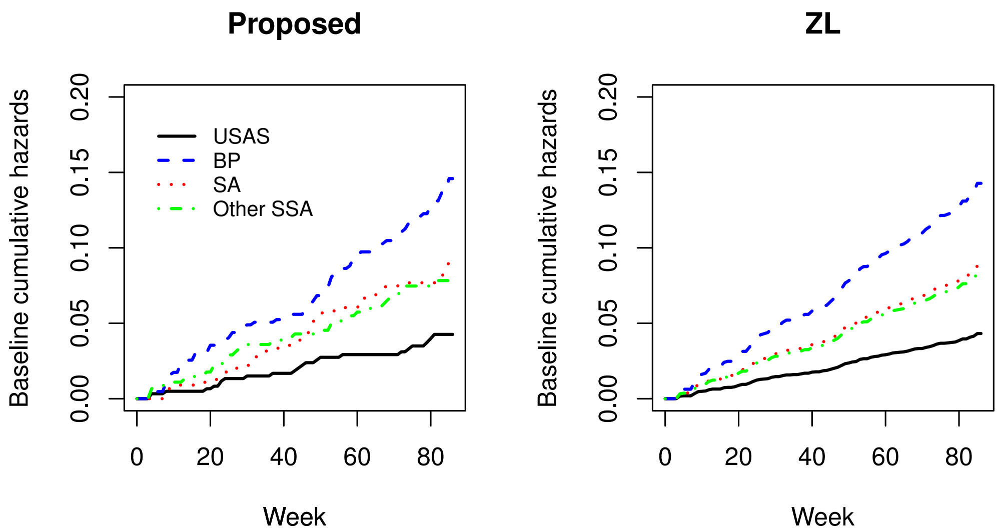
Note: Est and SE stand for the estimates of the regression parameters and the estimated standard errors, respectively. Here, we use the logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ with $r = 0, 1$ and 2.



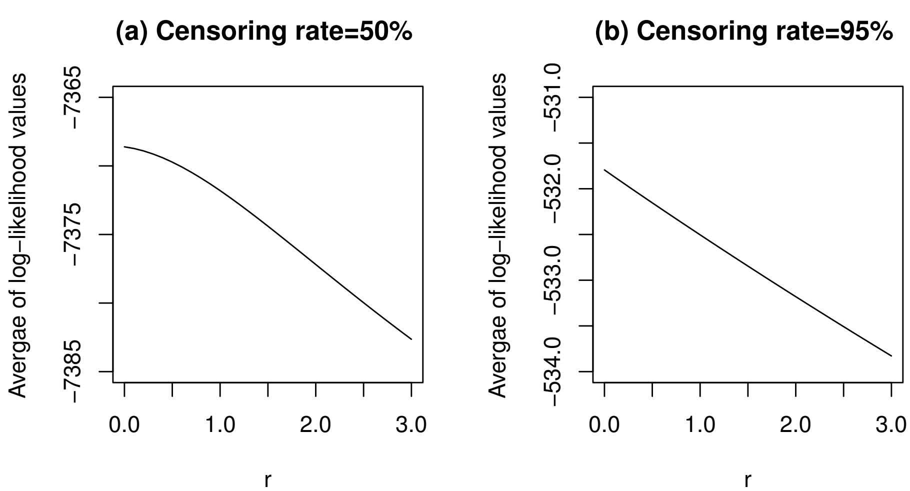
Web Figure 8: Log-likelihood function evaluated at the final parameter estimates when different values of r are considered in the Cox-Aalen transformation model with $G(x) = r^{-1}\log(1 + rx)$.



Web Figure 9: Estimated baseline cumulative hazard function for four regions under the logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ with $r = 0, 1$ and 2, respectively. Here, "USAS", "BP", "SA" and "Other SSA" represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.



Web Figure 10: Comparison with Zeng and Lin's model with $r = 0$ under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$



Web Figure 11: Average of log-likelihood function evaluated at the final parameter estimates when different values of r are considered in the Cox-Aalen transformation model (S.12) with $G(x) = r^{-1}\log(1 + rx)$.

References

Gao, F., Zeng, D., and Lin, D.-Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, 73(4):1161–1168.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Schwartz, J. T. (1969). *Nonlinear functional analysis*, volume 4. CRC Press.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3. Cambridge university press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271.