

Model assisted Cox regression



Shoubhik Mondal, Sundarraman Subramanian*

Center for Applied Mathematics and Statistics, Department of Mathematical Sciences, New Jersey Institute of Technology, USA

ARTICLE INFO

Article history:

Received 26 December 2012

Available online 11 October 2013

AMS subject classifications:

62E29

62H10

62N01

62N02

62N03

Keywords:

Empirical coverage

Event-time hazard

Gaussian process

Loewner ordering

Mean integrated squared error

Missing at random

ABSTRACT

Semiparametric random censorship (SRC) models (Dikta, 1998) [7], derive their rationale from their ability to utilize parametric ideas within the random censorship environment. An extension of this approach is developed for Cox regression, producing new estimators of the regression parameter and baseline cumulative hazard function. Under correct parametric specification, the proposed estimator of the regression parameter and the baseline cumulative hazard function are shown to be asymptotically as or more efficient than their standard Cox regression counterparts. Numerical studies are presented to showcase the efficacy of the proposed approach even under significant misspecification. Two real examples are provided. A further extension to the case of missing censoring indicators is also developed and an illustration with pseudo-real data is provided.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The SRC framework to survival function estimation for a homogeneous population, introduced by Dikta [7], operates as follows: specify a good-fitting parametric model for $m(x)$, the conditional expectation of the censoring indicator given the observed (possibly censored) event time, and replace the censoring indicators with the estimated model thereafter. With correct parametric specification of $m(x)$, this leads to an estimator which is asymptotically more efficient than the Kaplan–Meier estimator. Subramanian [21] employed this idea to construct likelihood ratio based confidence intervals for survival functions and reported good performance even in the face of considerable misspecification. The SRC approach is more flexible than the nonparametric approach in the sense that it applies even when there are missing censoring indicators (MCIs). In fact, when the MCIs are missing at random (MAR), no additional effort need be expended to address estimation [22]. Here, we propose and implement an extension that incorporates the SRC approach into Cox regression.

Analogous to the homogeneous case, for Cox proportional hazards (PH) regression we propose to replace the censoring indicator with any good-fitting parametric model for the aforementioned conditional expectation, which, in addition to its dependence on the observed event time, may now also depend on a set of covariates \mathbf{Z} . In order to understand the rationale for tying SRC models to Cox regression, note that, under conditional independence of failure and censoring variables given the covariate \mathbf{Z} , $m(x, \mathbf{z}) = P(\delta = 1 | X = x, \mathbf{Z} = \mathbf{z})$ is the ratio of the conditional event-time hazard to the conditional total hazard [7,27]. Specifically, for the Cox PH regression model, the conditional censoring hazard is linked to the event-time hazard through the multiplicative factor $\exp(-\text{logit}(m))$, which is a smooth function of the conditional odds of non-censoring given X and \mathbf{Z} . The standard Cox analysis ignores this relationship by leaving the conditional censoring

* Corresponding author.

E-mail address: sundars@njit.edu (S. Subramanian).

hazard unspecified. With SRC models, however, we exploit the link, using a model for m . Although the relationship explicitly calls for employing the logit, other choices such as the complementary log–log, generalized proportional hazards (GPH) and the Cauchy link may also be explored for m . In Section 3, the logit and Cauchy links are shown to provide improved estimator performance over standard Cox PH regression, with the Cauchy performing better than the logit in the sensitivity study. Furthermore, the SRC framework adapts to MCIs readily, unlike standard Cox analysis. We expect that in practice the added flexibility and improved performance would justify the additional effort required in the search for a good-fitting model for m .

Yuan [27] extended the Koziol–Green model [12] to the subject-specific setting implicit in Cox regression, which subsumes an earlier approach [23] as well. In terms of finite sample performance, both the proposed and Yuan [27] estimators perform equally well, see Section 3. Our proposed method, however, offers a more attractive alternative for the following reasons. Yuan [27] developed a log profile likelihood function which, however, involves the censoring indicator δ and hence would be inapplicable when there are MCIs. Furthermore, his approach requires simultaneous estimation of the finite dimensional components β and θ , compromising to some extent the simplicity of standard Cox regression analysis. Indeed, for a logistic model, Yuan's [27] approach will not be able to take advantage of the available logit function in statistical software. Our proposed method retains the simplicity of standard Cox regression and applies readily even when the MCIs are MAR. We show that $\hat{\beta}$ and $\hat{\Lambda}(t)$, the proposed estimators of β and $\Lambda_0(t)$ respectively, are each asymptotically as or more efficient than the standard Cox regression estimators.

Liu and Wang [14] proposed two estimators of β to account for the MCIs. They only focused on estimation of β , which is a limitation. Their first estimator denoted $\hat{\beta}_{LW}$, was based on a mixture, that reduces to the Cox partial likelihood estimator when there are no MCIs – and therefore *less efficient* than our proposed estimator when there are no MCIs. Their second estimator requires computation of kernel estimates which would be inefficient due to curse of dimensionality and the need for data-based optimal bandwidths. Numerical studies reported in Section 3 reveal that $\hat{\beta}$ performs as well as or better than $\hat{\beta}_{LW}$.

The article is organized as follows. In Section 2 we present our proposed estimators and provide theoretical comparisons with standard Cox regression estimators. We then present our proposed extension when there are MCIs. In Section 3, we present the results of simulation studies comparing the proposed and other approaches under discussion. We also provide illustrations using data from a heart transplant study and a study on recidivism, and another illustration using pseudo-real data. Technical details are given in the [Appendix](#).

2. Proposed estimators and large sample results

Let $\mathcal{N}_p(\mu, \Sigma)$ denote a p -variate normal distribution with mean vector μ and variance–covariance matrix Σ . When there are no MCIs, we observe n independent and identically distributed triplets $(X_i, \delta_i, \mathbf{Z}_i)$, $i = 1, \dots, n$, where $X = \min(T, C)$ is the minimum of the failure and censoring times, δ is the censoring indicator (1 when uncensored and 0 when censored), and \mathbf{Z} denotes a $p \times 1$ vector of covariates. The conditional hazard function of the failure time given \mathbf{Z} takes the form $\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\beta^T \mathbf{Z}}$, where β is the $p \times 1$ regression parameter and $\lambda_0(t)$ is a baseline hazard function. Writing $N_i(t) = I(X_i \leq t)$ and $Y_i(t) = I(X_i \geq t)$, $i = 1, \dots, n$, the Cox partial likelihood estimator of β , denoted by $\hat{\beta}_C$, solves $S_C(\beta) = 0$, where

$$S_C(\beta) = \sum_{i=1}^n \int_0^\infty \delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n Y_j(t) e^{\beta^T \mathbf{Z}_j}}{\sum_{j=1}^n Y_j(t) e^{\beta^T \mathbf{Z}_j}} \right] dN_i(t). \quad (2.1)$$

Breslow's [4] estimator of the baseline cumulative hazard function is given by

$$\hat{\Lambda}_{0C}(t) \equiv \hat{\Lambda}_0(t, \hat{\beta}_C) = \sum_{i=1}^n \int_0^t \frac{\delta_i}{\sum_{j=1}^n Y_j(s) e^{\hat{\beta}_C^T \mathbf{Z}_j}} dN_i(s). \quad (2.2)$$

Andersen and Gill [1] proved that $\hat{\beta}_C \xrightarrow{P} \beta_0$ and $n^{1/2}(\hat{\beta}_C - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \Sigma_C^{-1})$, where Eq. (A.5) defines Σ_C . They also derived the weak convergence of $n^{1/2}(\hat{\Lambda}_{0C}(t) - \Lambda_0(t))$, with the limiting variance function given by the first two terms of Eq. (A.45); see also [24].

2.1. Censoring indicators always observed

To tie the SRC framework to Cox regression, we write $m(X, \mathbf{Z}, \theta_0) = P(\delta = 1|X, \mathbf{Z})$. The unknown $\theta \in \mathbb{R}^k$, whose true value is θ_0 , can be estimated by maximizing the quantity

$$\prod_{i=1}^n \{m(X_i, \mathbf{Z}_i, \theta)\}^{\delta_i} \{1 - m(X_i, \mathbf{Z}_i, \theta)\}^{1-\delta_i}. \quad (2.3)$$

Let $\hat{\theta}$ denote the maximizer of Eq. (2.3). Then $\hat{\theta} \xrightarrow{P} \theta_0$ [13]. Our proposed estimator $\hat{\beta}$ is obtained by solving the equation $S_n(\beta, \hat{\theta}) = 0$, where

$$S_n(\beta, \hat{\theta}) = \sum_{i=1}^n \int_0^\infty m(t, \mathbf{Z}_i, \hat{\theta}) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n Y_j(t) e^{\beta^T \mathbf{Z}_j} \mathbf{Z}_j}{\sum_{j=1}^n Y_j(t) e^{\beta^T \mathbf{Z}_j}} \right] dN_i(t). \quad (2.4)$$

Note that $\hat{\beta}$ maximizes the adjusted partial log-likelihood function

$$\check{l}_n(\beta, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty m(t, \mathbf{Z}_i, \hat{\theta}) \left\{ \beta^T \mathbf{Z}_i - \log \left(\sum_{j=1}^n Y_j(t) e^{\beta^T \mathbf{Z}_j} \right) \right\} dN_i(t). \quad (2.5)$$

Our estimator of the baseline cumulative hazard function is given by

$$\hat{\Lambda}_0(t, \hat{\beta}, \hat{\theta}) = \sum_{i=1}^n \int_0^t \frac{m(s, \mathbf{Z}_i, \hat{\theta})}{\sum_{j=1}^n Y_j(s) e^{\hat{\beta}^T \mathbf{Z}_j}} dN_i(s). \quad (2.6)$$

Theorem 1 and **Proposition 1** give the large sample results of our proposed estimators.

Theorem 1. When the parametric model for $m(\mathbf{x}, \mathbf{z})$ is correctly specified and when conditions A.1–A.6 and D.1 hold, (i) $\hat{\beta} \xrightarrow{P} \beta_0$ and $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \Sigma)$, as $n \rightarrow \infty$, where Σ is given by Eq. (A.10); and (ii) $n^{1/2}(\hat{\Lambda}_0(\cdot, \hat{\beta}, \hat{\theta}) - \Lambda_0(\cdot)) \xrightarrow{D} \mathbb{U}(\cdot)$ in $D[0, \tau]$, where \mathbb{U} is a zero-mean Gaussian process with covariance function $\sigma(\cdot, \cdot)$ given by Eq. (A.17).

Proposition 1. When the parametric model for $m(\mathbf{x}, \mathbf{z})$ is correctly specified, the estimators $\hat{\beta}$ and $\hat{\Lambda}_0(t, \hat{\beta}, \hat{\theta})$ are asymptotically as or more efficient than $\hat{\beta}_C$ and $\hat{\Lambda}_{0C}(t)$ respectively.

Proof. Write $\langle \cdot, \cdot \rangle$ for the inner product in the Euclidean space. Note that \mathbf{B}_0 is given by Eq. (A.9) and $D_r(m(\mathbf{x}, \mathbf{z}, \theta_0)) = \partial m(\mathbf{x}, \mathbf{z}, \theta) / \partial \theta_r |_{\theta=\theta_0}$, $r = 1, \dots, k$. From Eqs. (A.5) and (A.33), we have

$$\Sigma - \Sigma_C^{-1} = \Sigma_C^{-1} [\mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T - E \{ m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} I(X \leq \tau) \}] \Sigma_C^{-1}.$$

Applying Loewner's ordering [20] it suffices to prove for any $\mathbf{a} = (a_1, \dots, a_p)^T$ that

$$\mathbf{a}^T [\mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T - E \{ m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} I(X \leq \tau) \}] \mathbf{a} \leq 0. \quad (2.7)$$

We follow [7], who uses the following result in [17]: for any $\mathbf{b} \in \mathbb{R}^k$,

$$\sup_{\mathbf{h} \in \mathbb{R}^k - \{0\}} \frac{\langle \mathbf{h}, \mathbf{b} \rangle^2}{\langle \mathbf{h}, \mathbf{I}_0 \mathbf{h} \rangle} = \langle \mathbf{b}, \mathbf{I}_0^{-1} \mathbf{b} \rangle. \quad (2.8)$$

Note that $\mathbf{b} = (b_1, \dots, b_k)^T = \mathbf{B}_0^T \mathbf{a}$ links the first term of (2.7) with the right hand side of Eq. (2.8): $\langle \mathbf{b}, \mathbf{I}_0^{-1} \mathbf{b} \rangle = \mathbf{a}^T \mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T \mathbf{a}$. It then suffices to show that for all $\mathbf{h} \in \mathbb{R}^k - \{0\}$

$$\frac{\langle \mathbf{h}, \mathbf{b} \rangle^2}{\langle \mathbf{h}, \mathbf{I}_0 \mathbf{h} \rangle} \leq \mathbf{a}^T E [m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} I(X \leq \tau)] \mathbf{a}. \quad (2.9)$$

Let $G(\mathbf{x}, \mathbf{z})$ denote the joint cumulative distribution function of (X, \mathbf{Z}) and $\Gamma_t = [0, t] \times \mathbb{R}^p$. Let $G^1(t, \mathbf{z}) = \int_{\Gamma_t} m(u, \mathbf{z}, \theta_0) \bar{m}(u, \mathbf{z}, \theta_0) dG(u, \mathbf{z})$. For fixed $\mathbf{h} \in \mathbb{R}^k - \{0\}$,

$$\begin{aligned} \langle \mathbf{h}, \mathbf{b} \rangle^2 &= \left[E \left(\sum_{i=1}^k h_i D_i(m(X, \mathbf{Z}, \theta_0)) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, t))^T I(X \leq \tau) \right) \mathbf{a} \right]^2 \\ &= \left[\int_{\Gamma_\tau} \sum_{i=1}^k \frac{h_i D_i(m(t, \mathbf{z}, \theta_0)) (\mathbf{z} - \bar{\mathbf{z}}(\beta_0, t))^T}{m(t, \mathbf{z}, \theta_0) \bar{m}(t, \mathbf{z}, \theta_0)} dG^1(t, \mathbf{z}) \mathbf{a} \right]^2. \end{aligned}$$

Applying the Cauchy–Schwarz inequality,

$$\langle \mathbf{h}, \mathbf{b} \rangle^2 \leq \int_{\Gamma_\tau} (\mathbf{a}^T (\mathbf{z} - \bar{\mathbf{z}}(\beta_0, t))^{\otimes 2} \mathbf{a}) dG^1(t, \mathbf{z}) \int_{\Gamma_\tau} \left(\sum_{i=1}^k \frac{h_i D_i(m(t, \mathbf{z}, \theta_0))}{m(t, \mathbf{z}, \theta_0) \bar{m}(t, \mathbf{z}, \theta_0)} \right)^2 dG^1(t, \mathbf{z}).$$

The second integral on the right hand side above equals $\langle \mathbf{h}, I(\theta_0) \mathbf{h} \rangle$. The first integral is just the right hand side of inequality (2.9), proving the asymptotic efficiency of $\hat{\beta}$.

The efficiency of $\hat{\lambda}_0(t, \hat{\beta}, \hat{\theta})$ is proved likewise. Note that the sum of the first two terms on the right side of Eq. (A.45) gives the variance of $\hat{\lambda}_0(t, \hat{\beta}_C)$. It remains to prove that

$$[\mathbf{d}_0(t)]^T I_0^{-1} \mathbf{d}_0(t) \leq E \left[m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\alpha(t, X, \mathbf{Z}, \beta_0))^2 I(X \leq \tau) \right],$$

where $\mathbf{d}_0(t)$ and $\alpha(t, X, \mathbf{Z}, \beta_0)$ are given by Eqs. (A.14) and (A.15) respectively. Proceeding as before, with $\mathbf{b} = \mathbf{d}_0(t)$, it suffices to show that for all $\mathbf{h} = (h_1, \dots, h_k) \in \mathbb{R}^k - \{0\}$,

$$\frac{(\mathbf{h}, \mathbf{b})^2}{\langle \mathbf{h}, I(\theta_0) \mathbf{h} \rangle} \leq E \left[m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\alpha(t, X, \mathbf{Z}, \beta_0))^2 I(X \leq \tau) \right]. \quad (2.10)$$

Indeed, for fixed $\mathbf{h} \in \mathbb{R}^k - \{0\}$, apply the Cauchy–Schwarz inequality to obtain

$$\begin{aligned} (\mathbf{h}, \mathbf{b})^2 &= \left[E \left\{ \sum_{i=1}^k h_i D_i(m(X, \mathbf{Z}, \theta_0)) \alpha(t, X, \mathbf{Z}, \beta_0) I(X \leq \tau) \right\} \right]^2 \\ &= \left[\int_{\Gamma_\tau} \left(\sum_{i=1}^k \frac{h_i D_i(m(s, \mathbf{z}, \theta_0))}{m(s, \mathbf{z}, \theta_0) \bar{m}(s, \mathbf{z}, \theta_0)} \right) \alpha(t, s, \mathbf{z}, \beta_0) dG^1(s, \mathbf{z}) \right]^2 \\ &\leq \int_{\Gamma_\tau} (\alpha(t, s, \mathbf{z}, \beta_0))^2 dG^1(s, \mathbf{z}) \times \int_{\Gamma_\tau} \left(\sum_{i=1}^k \frac{h_i D_i(m(s, \mathbf{z}, \theta_0))}{m(s, \mathbf{z}, \theta_0) \bar{m}(s, \mathbf{z}, \theta_0)} \right)^2 dG^1(s, \mathbf{z}). \end{aligned}$$

The first integral above is readily seen to be the right side of inequality (2.10). The second integral above is just $\langle \mathbf{h}, I(\theta_0) \mathbf{h} \rangle$. This completes the proof.

2.2. Censoring indicators missing at random

The data are $(X_i, \xi_i, \sigma_i, \mathbf{Z}_i)$, $i = 1, \dots, n$, where ξ is binary and indicates δ 's non-missingness status, and $\sigma = \xi \delta$. MAR implies that, conditional on X and \mathbf{Z} , the indicators ξ and δ are independent: $E(\sigma | X, \mathbf{Z}) = m(X, \mathbf{Z}) \pi(X, \mathbf{Z})$, where $\pi(X, \mathbf{Z}) = E(\xi | X, \mathbf{Z})$. Under MAR, $\hat{\theta}$, the estimator of θ may be obtained by maximizing the complete-case likelihood function

$$\prod_{i=1}^n \{m(X_i, \mathbf{Z}_i, \theta)\}^{\sigma_i} \{1 - m(X_i, \mathbf{Z}_i, \theta)\}^{\xi_i - \sigma_i}. \quad (2.11)$$

With this modification, we continue denoting the resulting estimators by $\hat{\beta}$ and $\hat{\lambda}$. Theorem 2 gives the large sample results of our proposed estimators for this case.

Theorem 2. When the model for $m(x, \mathbf{z})$ is correctly specified, when the MCIs are MAR, and conditions A.1–A.6 and D.1 hold, then (i) $\hat{\beta} \xrightarrow{P} \beta_0$ and $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \Sigma_M)$, as $n \rightarrow \infty$, where Σ_M is given by Eq. (A.11); (ii) $n^{1/2}(\hat{\lambda}_0(\cdot, \hat{\beta}, \hat{\theta}) - \Lambda_0(\cdot)) \xrightarrow{D} \mathbb{V}(\cdot)$ in $D[0, \tau]$, where \mathbb{V} is a zero-mean Gaussian process with covariance function $\sigma_M(\cdot, \cdot)$ given by Eq. (A.19).

Remark. When $\pi \equiv 1$ (no MCIs), then $\tilde{\mathbf{I}}_0 = \mathbf{I}_0$ and Σ_M reduces to Σ ; furthermore, $\sigma_M(t_1, t_2)$ reduces to $\sigma(t_1, t_2)$.

3. Numerical results

Here, we first report the results of comparison studies that we carried out when the censoring indicators are fully observed. These studies were conducted to assess performance (1) when the fitted model for m was the same as that used to generate the censoring indicators (no model misspecification); and (2) when the fitted model was different from that which generated the indicators. Comparisons between the estimators were based on the mean squared error (MSE) for $\hat{\beta}$, and the mean integrated squared error (MISE) for $\hat{\lambda}$ and the subject specific survival function estimators. We also compared the empirical coverage probabilities (ECPs) and estimated mean lengths (EMLs) of the 95% confidence intervals for $\hat{\beta}$ and $\hat{\beta}_C$, both in the absence and presence of misspecification. We present two illustrations. We then report results for the MCI scenario, including an illustration using pseudo-real data, where MCIs are artificially introduced. The estimators of [27] and [14] are denoted by $\hat{\beta}_Y$ and $\hat{\beta}_{LW}$ respectively. Note that $\hat{\beta}_Y$ applies only in the absence of MCIs and comparison with $\hat{\beta}_{LW}$ applies only when there are MCIs.

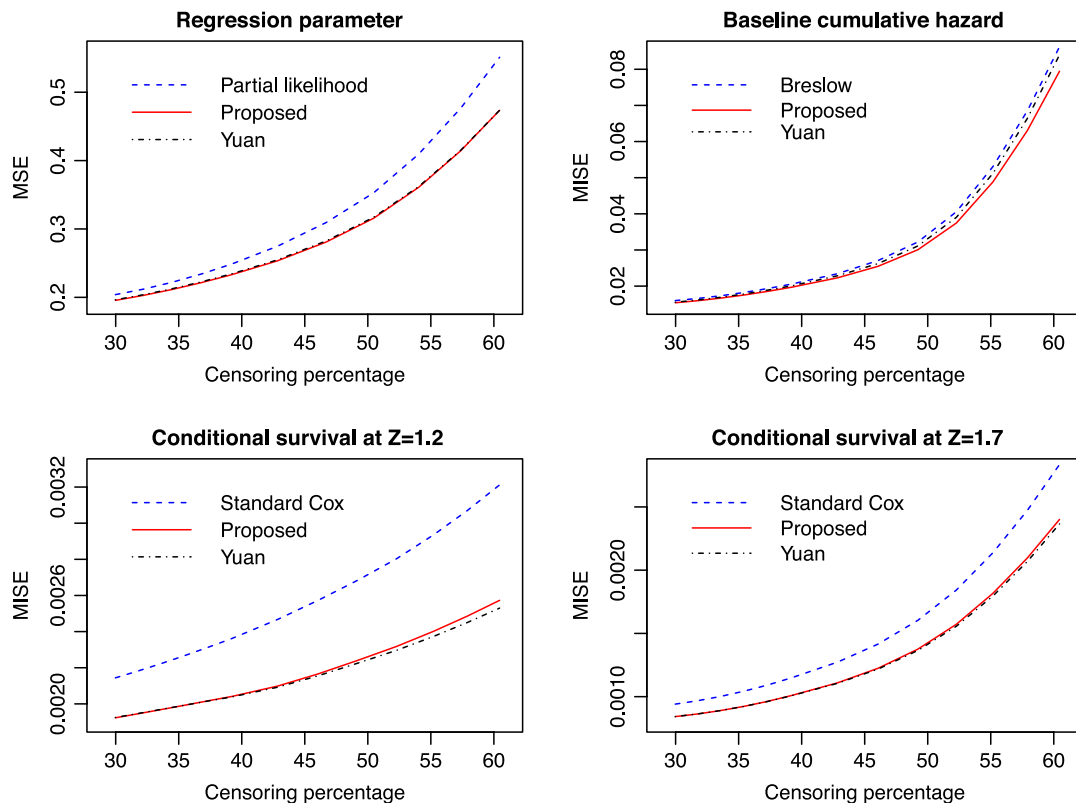


Fig. 1. Comparison of estimators when m is correctly specified. The mean squared error (MSE) and the mean integrated squared error (MISE) are plotted against censoring rate.

3.1. Censoring indicators always observed

We provide comparisons between the proposed, standard Cox, and Yuan [27] estimators.

3.1.1. No misspecification

The covariate was one dimensional and uniformly distributed over $[1, 2]$. The conditional event-time and censoring hazards given Z were $\exp(Z)$ and $\exp(\gamma(Z - 1))$ respectively, where γ is the censoring parameter selected to give censoring rates (CRs) between 30% and 60%. The baseline hazard was taken to be unity. The true model for the conditional probability $m(x, z) = P(\delta = 1|X = x, Z = z)$ is then the logit model given by $m(X, Z, \theta) = 1/(1 + \exp(-\theta_0 - \theta_1 Z))$. Sample size was 100 and all the estimates were averaged over 10,000 replications. The MISEs of the baseline cumulative hazard/subject-specific survival function estimators were calculated over $[0, \tau]$ where τ represented the 80th percentile of the marginal/conditional distributions of the event time. The results are shown in Fig. 1.

Our proposed estimator $\hat{\beta}$ outperforms $\hat{\beta}_C$ significantly, while being comparable to $\hat{\beta}_Y$. The relative improvement $[100 \times (\text{Cox} - \text{proposed})/\text{Cox}]$ over standard Cox varied from 4% to 15% in terms of MSE. For baseline cumulative hazard, the relative improvement of \hat{A} over the Breslow estimator in terms of MISE varied from 3% to 9%. The proposed estimator performed better than Yuan's [27] estimator as well. For subject specific survival with two covariate levels $Z = 1.2$ and $Z = 1.7$, the relative improvement of the proposed and Yuan [27] estimators over standard Cox varied between 10% and 20%.

Asymptotic confidence intervals based on $\hat{\beta}_Y$ require the bootstrap, see [27]. In Fig. 2 the ECP and EML of the 95% asymptotic confidence intervals for β , based on the large sample distributions of $\hat{\beta}$ and $\hat{\beta}_C$, are shown. The ECPs are close to the nominal 95%. However, the proposed approach offers a reduction of about 6.5% in EML over standard Cox.

3.1.2. Model for m misspecified

As in the first study, the event-time hazard was Z , where Z was one dimensional, having uniform distribution over $(1, 2)$. The baseline hazard was unity and the conditional censoring time was uniform over $(0, \gamma Z)$, where γ was calibrated to provide CRs between 10% and 60%. The true model for m is $m(t, z) = z(\gamma z - t)/[z(\gamma z - t) + 1]$. Misspecification was introduced by fitting the Cauchy link $m(X, Z, \theta) = 0.5 + \frac{1}{\pi} \tan^{-1}(\theta_0 + \theta_1 X + \theta_2 Z)$ to the generated censoring indicators. Sample size was 100 and each estimate was averaged over 1,000 replications. The results are shown in Fig. 3.

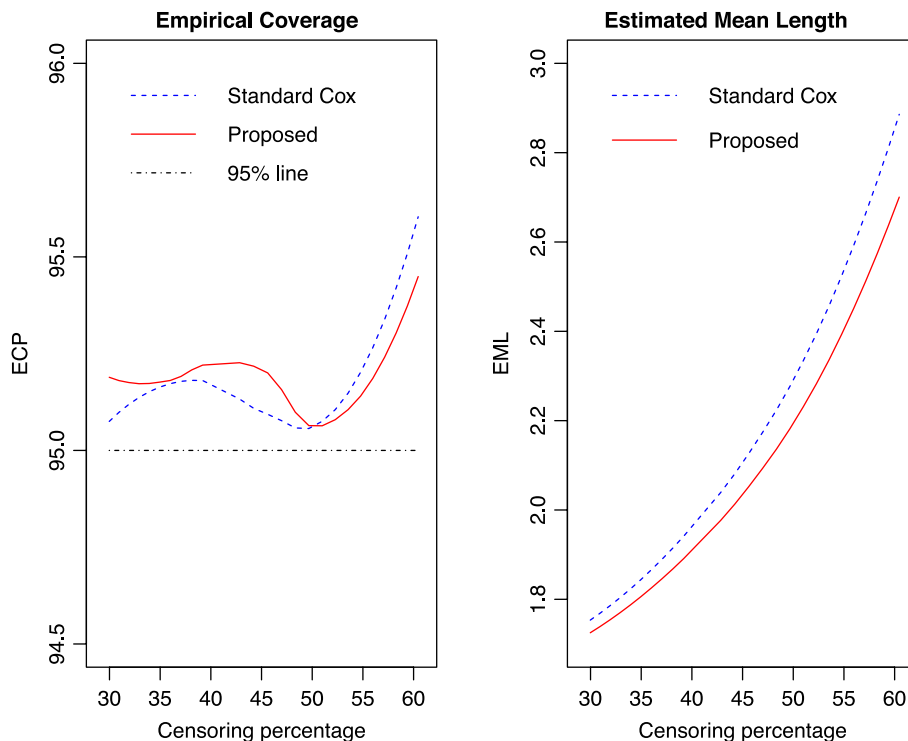


Fig. 2. Empirical coverage probability (ECP) and estimated average length (EML) of 95% confidence intervals are plotted for various censoring rates when m is correctly specified.

Table 1
Regression parameter estimates from analysis of SHT data.

Method	Estimates		Standard error	
	Age	Age-square	Age	Age-square
Cox	0.0434	0.0020	0.0106	0.0007
Proposed	0.0424	0.0017	0.0103	0.0006

The proposed estimator $\hat{\beta}$ outperformed $\hat{\beta}_C$, with the reduction in MSE relative to standard Cox being as high as 6%. Our baseline cumulative hazard estimator performed best for moderate CRs, with reduction up to 8%. For subject specific survival with two covariate levels $Z = 1.2$ and $Z = 1.7$, the relative improvement of the proposed and Yuan [27] estimators over standard Cox varied between 2% and 45%. Thus, even under significant model misspecification, our proposed estimators outperformed the standard Cox estimators.

In Fig. 4, the ECP and EML of the 95% asymptotic confidence intervals for β , based on $\hat{\beta}$ and $\hat{\beta}_C$, are shown. Even under significant misspecification the proposed method provides ECP close to the nominal 95%, with a reduction of about 1.5% in EML over standard Cox.

3.1.3. Illustration using the Stanford heart transplant (SHT) data

We illustrate our method using the well-known SHT data. For each of 184 transplant cases, the survival time was recorded in days from the date of transplant. There were 71 censored. The Cox PH model with the two-dimensional covariate $Z = (Z_1, Z_2)^T$, where $Z_1 = \text{age} - 41.7$ and $Z_2 = Z_1^2$ has been reported to fit the SHT data well (e.g. [8]). We employed a logistic model for m with covariates X (time) and Z_1 that were selected by a stepwise regression procedure: $\text{logit}(m(X, Z, \theta)) = \theta_0 + \theta_1 X + \theta_2 Z_1$. The results are presented in Table 1. Point estimates for both approaches are comparable, although the proposed standard errors are always lower than their standard Cox counterparts.

Subject-specific curves for two levels of “age”, namely 42.7 and 51.7, are plotted in Fig. 6 (along with two other curves, see Section 3.2.2). A faster decline in survival is seen for patients who are older, more so for the standard Cox estimate of the subject-specific survival, which may be reasoned as follows. The negative estimate of θ_1 indicates that the odds of censoring increases with “time”, which supports a basic notion that a longer surviving patient may be more likely to drop out of study, hence be censored. Our proposed method, by incorporating the censoring information through the model-based estimate, is able to address the underestimation of survival rates evidenced by standard Cox.

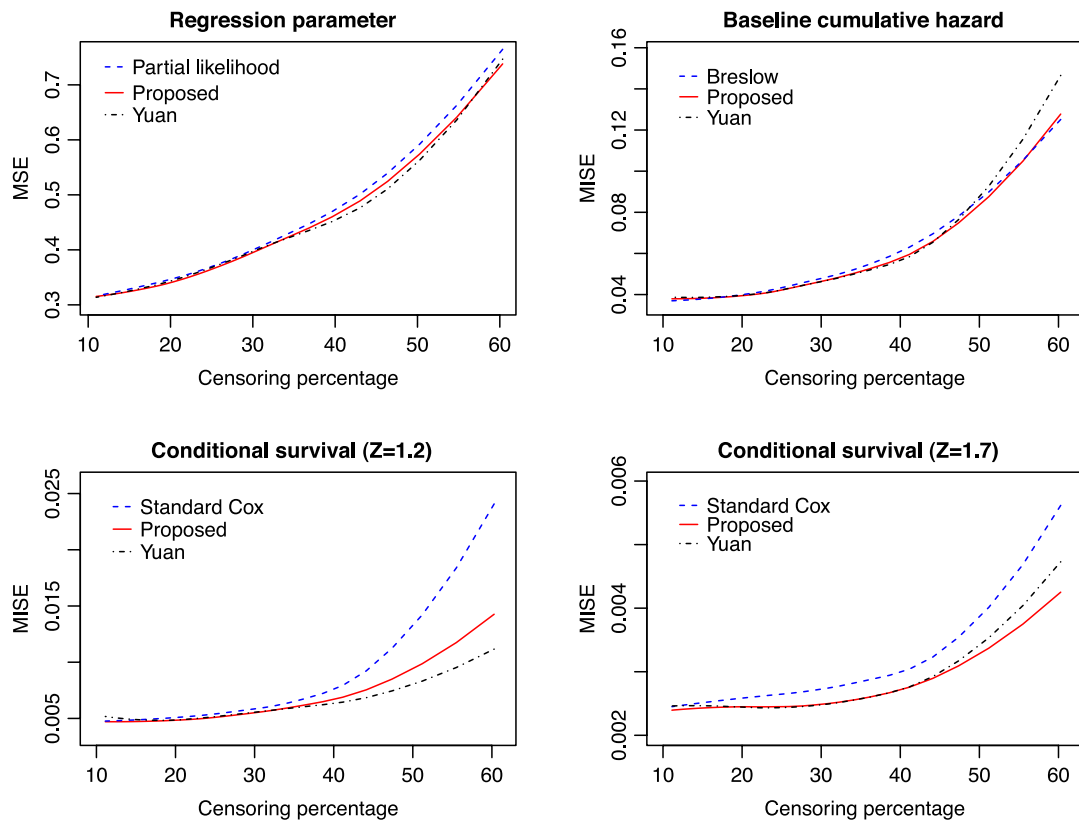


Fig. 3. Comparison of estimators when m is misspecified. The mean squared error (MSE) and the mean integrated squared error (MISE) are plotted against censoring rate.

Table 2

Standard Cox and proposed estimates after first-stage analysis of recidivism data.

Variable	Standard Cox			Proposed		
	Estimate	Std. Err.	P-value	Estimate	Std. Err.	P-value
fin	−0.379	0.191	0.047	−0.112	0.103	0.279
age	−0.057	0.022	0.009	−0.063	0.011	1.77e−08
race	0.314	0.308	0.308	0.085	0.158	0.592
workexp	−0.15	0.212	0.48	−0.001	0.115	0.991
mar	−0.434	0.382	0.256	−0.079	0.164	0.63
parole	−0.085	0.196	0.665	−0.028	0.108	0.799
prior	0.092	0.029	0.001	0.115	0.026	1.07e−05

3.1.4. Another illustration using recidivism data

For our second illustration, we present an analysis of data pertaining to an experimental study of recidivism of male prisoners [18]. The observed time (week) is the number of weeks for first arrest after release and the censoring indicator (arrest) equals 1 for those arrested during the period of the study (one year) and 0 otherwise. Fox [10] provided an analysis of these data using standard Cox. We considered the following seven covariates: financial aid status (fin), age at the time of release (age), race (race), full time work experience status before going to jail (workexp), marital status (mar), parole status (parole), and number of prior convictions (prior). We fitted the logistic model for m given by

$$\text{logit}(m(t, \mathbf{Z}, \boldsymbol{\theta})) = \theta_1 \text{age} + \theta_2 \text{prior}. \quad (3.12)$$

Among several choices investigated, Eq. (3.12) was optimal in terms of goodness-of-fit.

In Table 2, we present our results for standard Cox and proposed estimators. Both methods indicated that “age” and “prior” are significant factors. While standard Cox analysis finds “fin” marginally significant, we found it not significant, a conclusion supported by the second-stage analysis, where only the potentially significant variables “age”, “prior” and “fin” were considered. The same model was fitted for m , see Eq. (3.12). Table 3 gives the results for standard Cox and proposed estimators. Note that, unlike standard Cox, our proposed method was able to discard all insignificant factors in the first analysis itself.

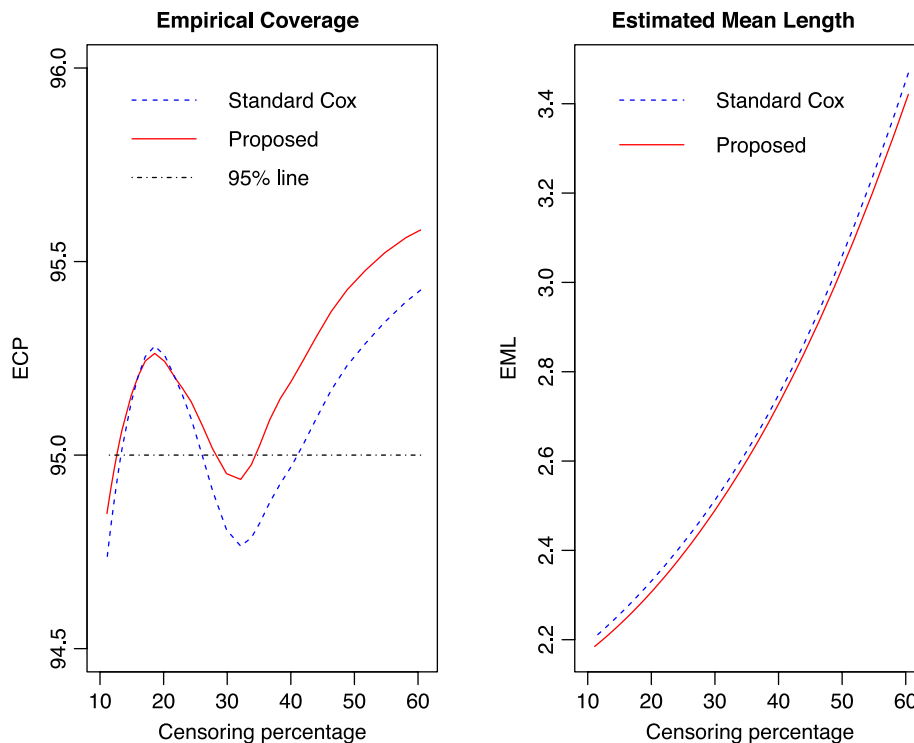


Fig. 4. Empirical coverage probability (ECP) and estimated average length (EML) of 95% confidence intervals plotted for various censoring rates when the model for m is misspecified.

Table 3

Standard Cox and proposed estimates after second-stage analysis of recidivism data.

Variable	Standard Cox			Proposed		
	Estimate	Std. Err.	P-value	Estimate	Std. Err.	P-value
fin	−0.347	0.190	0.0681	−0.104	0.1027	0.312
age	−0.067	0.021	0.0013	−0.064	0.0106	1.8e−09
prior	0.097	0.027	0.0004	0.115	0.0258	8.027e−06

3.2. Censoring indicators missing at random

We provide comparisons between $\hat{\beta}$ and $\hat{\beta}_{\text{LW}}$ and then an illustration using synthetic data.

3.2.1. Simulation study in the absence and presence of misspecification

Since Liu and Wang [14] did not investigate estimation of Λ_0 , here we only provide comparisons between $\hat{\beta}$ and $\hat{\beta}_{\text{LW}}$. Data were simulated according to the designs in Section 3.1.1 (no misspecification) and Section 3.1.2 (model for m misspecified). We imputed missingness via the logit model $\pi(t, Z) = 1/(1 + e^{-\alpha t})$, where α was chosen to give missingness rates (MRs) of about 20% and 44%. The MSEs were based on 10,000 replications, and the sample size was 100. Results are shown in Fig. 5. When m was correctly specified, $\hat{\beta}$ offered a reduction of up to 8% in MSE over $\hat{\beta}_W$. When m was misspecified, $\hat{\beta}$ offered a reduction of up to 7% over $\hat{\beta}_{\text{LW}}$ when the MR was 20% and up to 1.14% for 44% MR.

3.2.2. Illustration using synthetic data

We imputed missingness in the SHT data through the model $P(\xi = 1|t, Z) = e^t/(1 + e^t)$. It turned out that 44 observations had MCIs. The estimates, given in Table 4, are quite close to the complete data estimates (cf. Table 1), although MCIs inflate the standard errors.

Subject-specific survival curves for the age levels 42.7 and 52.7, obtained using the proposed method, are plotted in Fig. 6. Also plotted for comparison are the corresponding curves obtained using standard Cox based on *complete cases* (where observations with MCIs are ignored), full-data proposed, and full-data standard Cox. The proposed survival curves with or without MCIs show good agreement, indicating that our extension works well in practice.

Table 4
Estimates of Stanford Heart data under MCI.

Method	Estimates		Standard error	
	Age	Age-square	Age	Age-square
Proposed	0.0413	0.0017	0.01150	0.00061
Liu–Wang	0.0410	0.0018	0.00974	0.00063

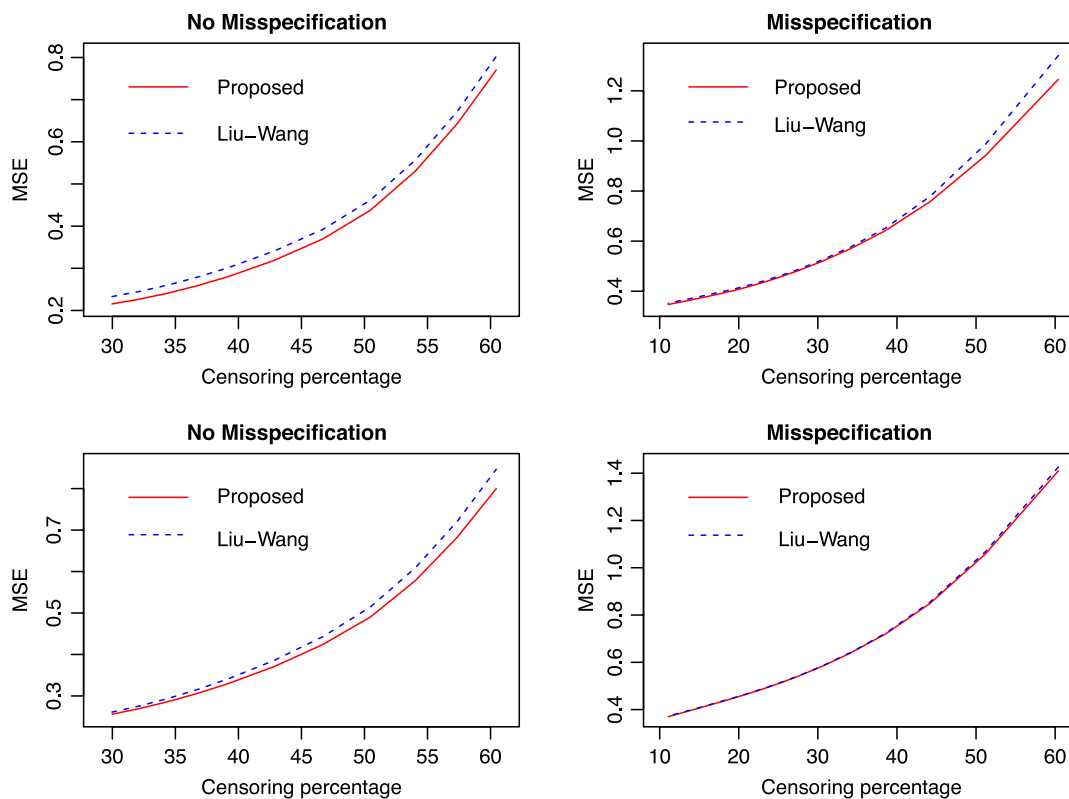


Fig. 5. Comparison of estimators when censoring indicators are missing at random. The mean squared error of $\hat{\beta}$ and $\hat{\beta}_{LW}$ are plotted for various censoring rates.

4. Concluding discussion

In this article, we have proposed and developed a novel model-based approach to standard Cox PH regression. We have derived the asymptotic properties of the new estimators and shown that, when the model for the conditional probability $m(x)$ is correctly specified, the new estimators are asymptotically as or more efficient than their standard Cox PH regression counterparts. Our numerical studies show that, under correct model specification, the proposed method produces better estimates of regression coefficients, baseline hazard, and the subject-specific survival function. Even under significant misspecification, our approach gave better parameter estimates. Our results are comparable to that of [27], whose method, however, cannot be applied when the censoring indicators are missing at random. Under MCIs, our method performed better than Liu and Wang's [14] procedure, who, however, did not provide any estimates for baseline hazard. We have given a unified approach that can handle both cases of absence and presence of MCIs without any extra effort. With the implementation of the proposed approach, it would be possible to reinforce, or modify in borderline cases, past conclusions by investigators of several cancer and other studies. This aspect was illustrated very well in the analysis of the recidivism data set.

A referee pointed out that a parametric assumption on m , together with the Cox model, imposes a new semiparametric model and wondered whether a direct maximum likelihood approach is possible. Writing $r(x) = e^x$ and denoting $k(\mathbf{z})$ as the density function of \mathbf{Z} , the adjusted likelihood, after factoring in the model for m , takes the form [see also [27]]

$$\prod_{i=1}^n \left(\frac{1 - m(X_i, \mathbf{Z}_i, \theta)}{m(X_i, \mathbf{Z}_i, \theta)} \right)^{1 - m(X_i, \mathbf{Z}_i, \theta)} \lambda_0(X_i) r(\beta' \mathbf{Z}_i) \exp \left[-r(\beta' \mathbf{Z}_i) \int_0^\tau \frac{Y_i(u) \lambda_0(u)}{m(u, \mathbf{Z}_i, \theta)} du \right] k(\mathbf{Z}_i).$$

The maximum of the above likelihood may not exist if $\Lambda_0(t)$, the baseline cumulative hazard, is restricted to be absolutely continuous. Therefore, allowing Λ_0 to be discrete, λ_0 may be replaced with the jump size of Λ_0 to obtain a modified

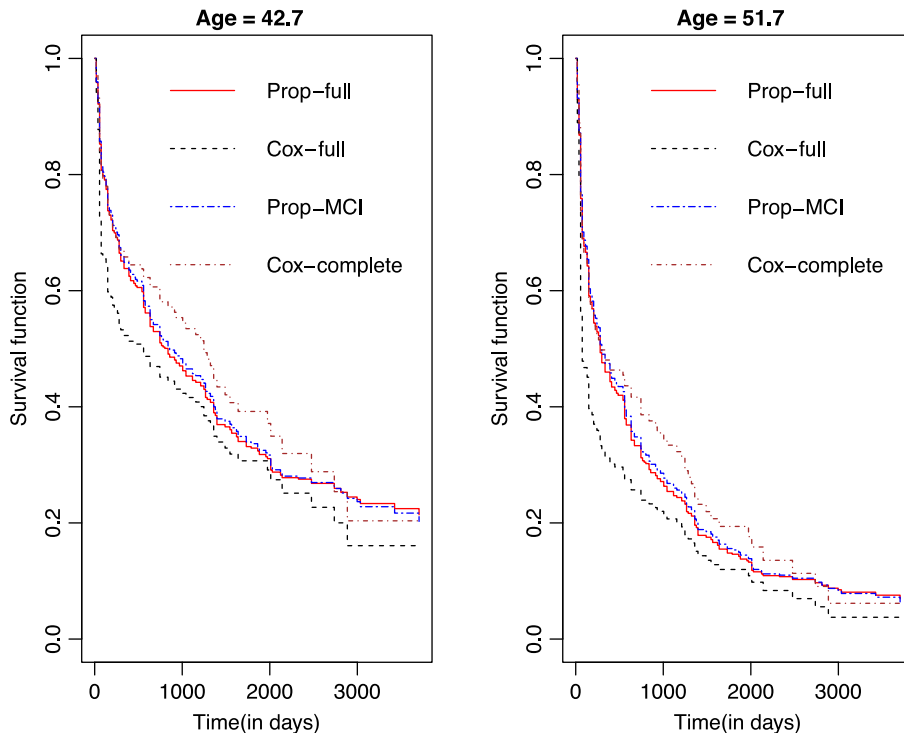


Fig. 6. Subject specific survival function when censoring indicators are missing at random.

likelihood, whose maximizer can be found numerically, see [15], which follows such an approach for the proportional hazards cure model. For standard Cox, the nonparametric maximum likelihood estimators found in this way are identical to $\hat{\beta}_C$ and the Breslow estimator Λ_{0C} , see Lin and Zeng [28]. It would be a worthwhile direction for further research to investigate whether such direct maximization of the likelihood would yield improved estimators of β and $\Lambda_0(t)$, without compromising the simplicity of analysis as well. Yuan [27] obtained a likelihood by *profiling* out λ_0 in the above likelihood, which he then maximized to obtain his estimators of β and $\Lambda_0(t)$. However, our proposed approach offers an attractive alternative and performs as well as his estimators.

Our incorporation of binary regression models into standard Cox regression raises the issue of finding good-fitting models for $m(x)$. A number of choices such as the logit, probit, complementary log–log, generalized proportional hazards, and the Cauchy link may be explored to arrive at a good-fitting model for m . They have been found to be mostly adequate for modeling binary responses, see for example [6]. In Section 3, the logit and Cauchy links were shown to provide improved estimator performance over standard Cox PH regression, with the Cauchy performing better than the logit in the sensitivity study.

Acknowledgments

The authors extend their sincere thanks to the Editor-in-Chief and an Editor for the opportunity to submit revised versions. They are very grateful to a reviewer for comments that resulted in a significantly improved final version.

Appendix

Before we present proofs of our theorems, we will need some preliminaries, namely setting out notation and recalling some existing results that we will employ. We shall first focus our proofs on the interval $[0, \tau]$, where $\tau < \tau_H$, and $\tau_H = \sup\{t : P(X > t) > 0\}$ is the right end-point of the support of the distribution of X . We then provide an extension of our asymptotic normality proof that applies to the entire interval $[0, \tau_H]$, that uses “all of the data”, following the method of proof given in Theorem 8.4.3 of [9].

Let $\mathbf{W} = (X, \mathbf{Z})$ and $\bar{m}(\mathbf{w}, \theta) = 1 - m(\mathbf{w}, \theta)$. Let $D_r(m(\mathbf{w}, \theta))$ denote the partial derivative of $m(\mathbf{w}, \theta)$ with respect to θ_r . Write $\text{Grad}(m(\mathbf{w}, \theta)) = [D_1(m(\mathbf{w}, \theta)), \dots, D_k(m(\mathbf{w}, \theta))]^T$ and let $\mathbf{J}_\theta(t, \mathbf{z}) = [(\text{Grad}(m(t, \mathbf{z}, \theta)))^{\otimes 2}]$, where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Vectors and matrices will be in bold. The information in the absence of MCIs [see Eq. (4) on page 257 of [7]] and in the presence of MCIs [see Eq. (3.11) on page 134 of [22]] is given by

$$\mathbf{I}(\theta_0) \equiv \mathbf{I}_0 = E \left(\frac{\mathbf{J}_{\theta_0}(\mathbf{W})}{m(\mathbf{W}, \theta_0)\bar{m}(\mathbf{W}, \theta_0)} \right); \quad \tilde{\mathbf{I}}(\theta_0) \equiv \tilde{\mathbf{I}}_0 = E \left(\frac{\pi(\mathbf{W})\mathbf{J}_{\theta_0}(\mathbf{W})}{m(\mathbf{W}, \theta_0)\bar{m}(\mathbf{W}, \theta_0)} \right). \quad (\text{A.1})$$

Writing $Y(t) = I(X \geq t)$, we define the following quantities; see [1]:

$$\mathbf{s}^{(m)}(\boldsymbol{\beta}, t) = \frac{1}{n} \sum_{i=1}^n [Y_i(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_i} \mathbf{Z}_i^{\otimes m}], \quad \mathbf{s}^{(m)}(\boldsymbol{\beta}, t) = E[Y(t) e^{\boldsymbol{\beta}^T \mathbf{Z}} \mathbf{Z}^{\otimes m}]; \quad m = 0, 1, 2;$$

$$\bar{\mathbf{Z}}(\boldsymbol{\beta}, t) = \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}, t)}; \quad \bar{\mathbf{z}}(\boldsymbol{\beta}, t) = \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}, t)}; \quad \mathbf{v}(\boldsymbol{\beta}, t) = \frac{\mathbf{s}^{(2)}(\boldsymbol{\beta}, t)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}, t)} \right)^{\otimes 2}.$$

Note that $\Lambda_{X|Z}(t)$, the conditional cumulative hazard function of X given \mathbf{Z} , is the sum of the conditional cumulative hazards of T and C , and that $m(t, \mathbf{z})$ is the ratio of the event-time hazard to the total hazard. It follows that $\Lambda_{X|Z}(t) = \int_0^t e^{\boldsymbol{\beta}_0^T \mathbf{z}} \lambda_0(u) du / m(u, \mathbf{z}, \boldsymbol{\theta}_0)$. See also page 497 of [27], where $\gamma(x, z, \boldsymbol{\theta})$ satisfies the relation $1 + \gamma(x, z, \boldsymbol{\theta}) = 1/m(x, z, \boldsymbol{\theta})$. For each $i = 1, \dots, n$, the counting process $N_i(t) = I(X_i \leq t)$, conditional on \mathbf{Z}_i , has a compensator $\int_0^t Y_i(s) d\Lambda_{X_i|Z_i}(s)$ so that

$$M_i(t) = N_i(t) - \int_0^t \frac{Y_i(u) e^{\boldsymbol{\beta}_0^T \mathbf{Z}_i} \lambda_0(u)}{m(u, \mathbf{Z}_i, \boldsymbol{\theta}_0)} du \quad (\text{A.2})$$

is a martingale with respect to the sigma-field $\mathcal{F}_t = \sigma\{\mathbf{Z}_i, I(X_i \leq s), i = 1, \dots, n : s \leq t\}$, see page 81 of [3]. Furthermore, the counting process $N_i^u(t) = I(X_i \leq t, \delta_i = 1)$, conditional on \mathbf{Z}_i , has a compensator $\int_0^t Y_i(s) e^{\boldsymbol{\beta}_0^T \mathbf{Z}_i} \lambda_0(s) ds$ so that

$$M_i^u(t) = N_i^u(t) - \int_0^t Y_i(u) e^{\boldsymbol{\beta}_0^T \mathbf{Z}_i} \lambda_0(u) du \quad (\text{A.3})$$

is a martingale with respect to the sigma-field $\check{\mathcal{F}}_t = \sigma\{\mathbf{Z}_i, N_i^u(s), Y_i(s+), i = 1, \dots, n : s \leq t\}$; see page 128 of [9]. The corresponding predictable covariation processes are given by

$$\langle M^u, M^u \rangle(t) = \int_0^t Y(u) e^{\boldsymbol{\beta}_0^T \mathbf{Z}} \lambda_0(u) du; \quad \langle M, M \rangle(t) = \int_0^t \frac{Y(u) e^{\boldsymbol{\beta}_0^T \mathbf{Z}} \lambda_0(u)}{m(u, \mathbf{Z}, \boldsymbol{\theta}_0)} du. \quad (\text{A.4})$$

See Eqs. (2.1)–(2.3) of [1] for $\langle M^u, M^u \rangle(t)$. Theorem 2.5.2 of [9] yields the expression for $\langle M, M \rangle(t)$.

We note several identities involving Σ_C , the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0)$:

$$\Sigma_C = \int_0^\tau \mathbf{v}(\boldsymbol{\beta}_0, t) \mathbf{s}^{(0)}(\boldsymbol{\beta}_0, t) \lambda_0(t) dt, \quad (\text{A.5})$$

$$\Sigma_C = E \left[\int_0^\tau m(t, \mathbf{Z}, \boldsymbol{\theta}_0) \mathbf{v}(\boldsymbol{\beta}_0, t) dN(t) \right], \quad (\text{A.6})$$

$$\Sigma_C = E \left[m(X, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right]. \quad (\text{A.7})$$

Andersen and Gill [1] derived Eq. (A.5). Furthermore, following the proof of Theorem 3.2 of [1], it can be shown that the right hand side of Eq. (A.5) also equals $E \left[\int_0^\tau (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, s))^{\otimes 2} dN^u(s) \right]$. Apart from the fact that this term is the right hand side of Eq. (A.7), a simple conditional argument applied to this term also establishes Eq. (A.6). We will also need the following quantities:

$$\mathbf{V}_0 = E \left[(m(X, \mathbf{Z}, \boldsymbol{\theta}_0))^2 (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right], \quad (\text{A.8})$$

$$\mathbf{B}_0 = E \left[\int_0^\tau [\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t)] [\text{Grad}(m(\mathbf{W}, \boldsymbol{\theta}_0))]^T dN(t) \right], \quad (\text{A.9})$$

$$\Sigma = \Sigma_C^{-1} [\mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T + \mathbf{V}_0] \Sigma_C^{-1}, \quad (\text{A.10})$$

$$\Sigma_M = \Sigma_C^{-1} [\mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} \mathbf{B}_0^T + \mathbf{V}_0] \Sigma_C^{-1}, \quad (\text{A.11})$$

$$\mathbf{C}_0(t) = \int_0^t \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0, s)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s)} \lambda_0(s) ds, \quad (\text{A.12})$$

$$\mathbf{D}_0(t) = E \left[\int_0^t \frac{\text{Grad}(m(s, \mathbf{Z}, \boldsymbol{\theta}_0))}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s)} dN(s) \right], \quad (\text{A.13})$$

$$\mathbf{d}_0(t) = \mathbf{D}_0(t) - \mathbf{B}_0^T \Sigma_C^{-1} \mathbf{C}_0(t). \quad (\text{A.14})$$

Also define

$$\alpha(t, X, \mathbf{Z}, \boldsymbol{\beta}_0) = \frac{I(X \leq t)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, X)} - [\mathbf{C}_0(t)]^T \Sigma_C^{-1} [\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X)], \quad (\text{A.15})$$

$$\gamma(t_1, t_2) = E \left[m^2(X, \mathbf{Z}, \boldsymbol{\theta}_0) \alpha(t_1, X, \mathbf{Z}, \boldsymbol{\beta}_0) \alpha(t_2, X, \mathbf{Z}, \boldsymbol{\beta}_0) I(X \leq \tau) \right], \quad (\text{A.16})$$

$$\sigma(t_1, t_2) = [\mathbf{d}_0(t_1)]^T \mathbf{I}_0^{-1} [\mathbf{d}_0(t_2)] + \gamma(t_1, t_2), \quad (\text{A.17})$$

$$\sigma(t, t) = [\mathbf{d}_0(t)]^T \mathbf{I}_0^{-1} [\mathbf{d}_0(t)] + \gamma(t, t), \quad (\text{A.18})$$

$$\sigma_M(t_1, t_2) = [\mathbf{d}_0(t_1)]^T \tilde{\mathbf{I}}_0^{-1} [\mathbf{d}_0(t_2)] + \gamma(t_1, t_2), \quad (\text{A.19})$$

$$\sigma_M(t, t) = [\mathbf{d}_0(t)]^T \tilde{\mathbf{I}}_0^{-1} [\mathbf{d}_0(t)] + \gamma(t, t). \quad (\text{A.20})$$

Throughout the paper we shall assume that T and C are conditionally independent given \mathbf{Z} . We shall also need the following conditions to prove [Theorems 1](#) and [2](#).

A. 1. The covariate \mathbf{Z} is bounded, that is, for $M_0 > 0$, $\mathbf{Z} \in [-M_0, M_0]^p$ almost surely.

A. 2. There exists a neighborhood \mathcal{B} of $\boldsymbol{\beta}_0$ such that, for $j = 0, 1, 2$,

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}, t \in [0, \tau]} \|\mathbf{s}^j(\boldsymbol{\beta}, t) - \mathbf{s}^j(\boldsymbol{\beta}_0, t)\| \xrightarrow{P} 0.$$

A. 3. The functions $\mathbf{s}^{(j)}$ are bounded and $\mathbf{s}^{(0)}$ is bounded away from 0 on $\mathcal{B} \times [0, \tau]$; for $j = 0, 1, 2$, the family of functions $\mathbf{s}^{(j)}(\cdot, t)$, $0 \leq t \leq \tau$, is an equicontinuous family at $\boldsymbol{\beta}_0$.

A. 4. The matrix $\boldsymbol{\Sigma}_C$ [cf. Eq. (A.5)–(A.7)] is positive definite.

A. 5. The matrices $\mathbf{I}(\boldsymbol{\theta}_0)$, $\tilde{\mathbf{I}}(\boldsymbol{\theta}_0)$ [cf. Eq. (A.1)] are positive definite.

A. 6. The function $m(x, \mathbf{z})$ is bounded away from zero in $\Gamma_{\tau_H} \equiv [0, \tau_H) \times [-M_0, M_0]^p$.

Condition A.1 is assumed in Theorem 4.2 of [1] and in Theorem 8.4.1. of [9], both for the iid case, and considered in this paper as well. Although an alternative set of weaker conditions involving the finiteness of the second moments of \mathbf{Z} and $M(X, \mathbf{Z})$ (defined in condition D.1) can be given, condition A.1 is necessary to prove our results over the entire interval $[0, \tau_H]$, see theorem 8.4.3 of [9]. The set of conditions A.2–A.4 was given by Andersen and Gill [1] for standard Cox and discussed well there; see also pages 289–290 of [9]. Condition A.5 is standard in parametric inference. Condition A.6 will be needed for proving asymptotic normality of our proposed estimator over the whole interval $[0, \tau_H]$. For binary regression models with logit, probit or Cauchy links, which would be our principal focus, condition A.1 implies the following condition.

D. 1. There exists a neighborhood $V(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ and a measurable function $M(\cdot, \cdot)$ of x and \mathbf{z} such that, for each $r = 1, \dots, k$, $|D_r(m(x, \mathbf{z}, \boldsymbol{\theta}))| \leq M(x, \mathbf{z})$ and $E(M(X, \mathbf{Z})) < \infty$.

For general m , however, condition D.1 will be needed for proving consistency of $\hat{\boldsymbol{\beta}}$; see also a precursor in Theorem 2.4 of [7], of which D.1 is an extension.

Note that, since $\hat{\boldsymbol{\theta}}$ is derived via maximum likelihood [cf. Eq. (2.3)], it is an M -estimator and therefore has the representation given by

$$n^{1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{-1/2} \sum_{i=1}^n \mathbf{I}_0^{-1} \frac{\delta_i - m(\mathbf{W}_i, \boldsymbol{\theta}_0)}{m(\mathbf{W}_i, \boldsymbol{\theta}_0) \bar{m}(\mathbf{W}_i, \boldsymbol{\theta}_0)} \text{Grad}(m(\mathbf{W}_i, \boldsymbol{\theta}_0)) + o_p(1). \quad (\text{A.21})$$

See page 241 of [25]. For binary regression, see Example 5.40 of [26]. For more general m , adaptations of conditions in [16] as carried out by Dikta [7], would be necessary. However, binary regression models for m would be our chief focus, since they are often used in survival analysis applications and they are readily available in all statistical software.

A.1. Proof of [Theorem 1](#)

A.1.1. Consistency of $\hat{\boldsymbol{\beta}}$

Since $\check{\mathbf{I}}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}})$, defined by Eq. (2.5), is free of $\boldsymbol{\beta}$, it follows that $\hat{\boldsymbol{\beta}}$ maximizes

$$\begin{aligned} \mathbf{I}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) &= \check{\mathbf{I}}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) - \check{\mathbf{I}}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}) \\ &\equiv \frac{1}{n} \sum_{i=1}^n \int_0^\tau m(t, \mathbf{Z}_i, \hat{\boldsymbol{\theta}}) \left\{ (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i - \log \left(\frac{S^{(0)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}_0, t)} \right) \right\} dN_i(t). \end{aligned}$$

In order to apply the methods of [1], we introduce the function

$$\tilde{l}(\boldsymbol{\beta}, \boldsymbol{\theta}_0) = E \int_0^\tau m(t, \mathbf{Z}, \boldsymbol{\theta}_0) \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z} - \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] \right] dN(t),$$

and show that it is concave. Indeed, applying Eq. (A.2) followed by condition A.6,

$$\begin{aligned} \tilde{l}(\boldsymbol{\beta}, \boldsymbol{\theta}_0) &= E \int_0^\tau m(t, \mathbf{Z}, \boldsymbol{\theta}_0) \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z} - \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] \right] dM(t) \\ &\quad + E \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z} - \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] \right] I(X \geq t) e^{\beta_0^T \mathbf{Z}} \lambda_0(t) dt \\ &= \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{s}^{(1)}(\boldsymbol{\beta}_0, t) - \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] s^{(0)}(\boldsymbol{\beta}_0, t) \right] \lambda_0(t) dt, \end{aligned}$$

coincides with Andersen and Gill's [1] concave limit function of $A(\boldsymbol{\beta}, \tau)$ (as well as of $X(\boldsymbol{\beta}, \tau)$), see p. 1106 of their paper, and, hence, $\tilde{l}(\boldsymbol{\beta}, \boldsymbol{\theta}_0)$ is a concave function of $\boldsymbol{\beta}$. Since $\mathbf{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ is a random concave function of $\boldsymbol{\beta}$ as well, consistency will follow by the arguments in the concluding part of Lemma 3.1 of [1], provided it can be shown that

$$\|\mathbf{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) - \tilde{l}(\boldsymbol{\beta}, \boldsymbol{\theta}_0)\| = o_p(1), \text{ for each } \boldsymbol{\beta} \in \mathcal{B}.$$

By the triangle inequality, it suffices to introduce the intermediate function

$$\tilde{\mathbf{l}}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0) \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i - \log \left[\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right] \right] dN_i(t),$$

and instead show that the following two equations hold pointwise in $\boldsymbol{\beta} \in \mathcal{B}$,

$$\|\mathbf{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) - \tilde{\mathbf{l}}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0)\| = o_p(1) \quad (\text{A.22})$$

$$\|\tilde{\mathbf{l}}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0) - \tilde{l}(\boldsymbol{\beta}, \boldsymbol{\theta}_0)\| = o_p(1). \quad (\text{A.23})$$

The strong law of large numbers implies Eq. (A.23). To prove Eq. (A.22), Taylor's expansion of $\mathbf{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ yields

$$\mathbf{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) - \tilde{\mathbf{l}}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0) = [\mathbf{l}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0) - \tilde{\mathbf{l}}_n(\boldsymbol{\beta}, \boldsymbol{\theta}_0)] + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{l}_n(\boldsymbol{\beta}, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \quad (\text{A.24})$$

where $\boldsymbol{\theta}^*$ is an intermediate value between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. We apply conditions A.2 and A.3 to deduce that the first term of Eq. (A.24) is $o_p(1)$ (in fact, uniformly for $\boldsymbol{\beta} \in \mathcal{B}$). We show that the second term is also $o_p(1)$, assuming for simplicity that $p = 1$. Note that condition D.1 implies that $|\langle \text{Grad}_{\boldsymbol{\theta}}(m(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}^*)), (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rangle| \leq kM(\mathbf{x}, \mathbf{z}) \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$. Then,

$$\begin{aligned} &n \left| \langle \text{Grad}_{\boldsymbol{\theta}}(\mathbf{l}_n(\boldsymbol{\beta}, \boldsymbol{\theta}^*)), (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rangle \right| \\ &= \left| \left\langle \sum_{i=1}^n \text{Grad}_{\boldsymbol{\theta}}(m(X_i, \mathbf{Z}_i, \boldsymbol{\theta}^*)) \left\{ (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i - \log \left(\frac{s^{(0)}(\boldsymbol{\beta}, X_i)}{s^{(0)}(\boldsymbol{\beta}_0, X_i)} \right) \right\} I(X_i \leq \tau), (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right\rangle \right| \\ &\leq k \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \left\{ \text{constant} + \sup_{t \in [0, \tau], \boldsymbol{\beta} \in \mathcal{B}} \left| \log \left(\frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right) \right| \right\} \sum_{i=1}^n M(X_i, \mathbf{Z}_i) + o_p(n) = o_p(n), \end{aligned}$$

by A.1 and D.1, together with the strong law of large numbers and consistency of $\hat{\boldsymbol{\theta}}$. \square

Remark. The proof of consistency is exactly the same, whether there are MCIs or not.

A.1.2. Asymptotic normality of $\hat{\boldsymbol{\beta}}$

Defining three normalized sums of iid random variables by

$$\mathbf{U}_{n,1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = n^{-1/2} \sum_{i=1}^n \mathbf{B}_0 \mathbf{I}_0^{-1} \frac{\delta_i - m(\mathbf{W}_i, \boldsymbol{\theta}_0)}{m(\mathbf{W}_i, \boldsymbol{\theta}_0) \bar{m}(\mathbf{W}_i, \boldsymbol{\theta}_0)} \text{Grad}(m(\mathbf{W}_i, \boldsymbol{\theta}_0)),$$

$$\mathbf{U}_{n,2}(\boldsymbol{\beta}_0) = n^{-1/2} \sum_{i=1}^n \int_0^\tau [(\mathbf{Z}_i - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))] dM_i^u(t),$$

$$\mathbf{U}_{n,3}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = -n^{-1/2} \sum_{i=1}^n \int_0^\tau [(\delta_i - m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0))(\mathbf{Z}_i - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))] dN_i(t),$$

we develop the asymptotic representation given below for proving asymptotic normality.

$$n^{1/2}(\hat{\beta} - \beta_0) = \Sigma^{-1} [\mathbf{U}_{n,1}(\beta_0, \theta_0) + \mathbf{U}_{n,2}(\beta_0) + \mathbf{U}_{n,3}(\beta_0, \theta_0)] + o_p(1). \quad (\text{A.25})$$

Defining $n\mathbf{A}_n(\beta_0, \hat{\theta}) = \partial \mathbf{S}_n(\beta, \hat{\theta}) / \partial \beta|_{\beta=\beta_0}$ and $n\mathbf{B}_n(\beta_0, \theta_0) = \partial \mathbf{S}_n(\beta_0, \theta) / \partial \theta|_{\theta=\theta_0}$, and using Taylor's expansion and the consistency of $\hat{\beta}$ and $\hat{\theta}$, we can show that

$$\mathbf{S}_n(\hat{\beta}, \hat{\theta}) = \mathbf{S}_n(\beta_0, \theta_0) + n\mathbf{B}_n(\beta_0, \theta_0)(\hat{\theta} - \theta_0) + n\mathbf{A}_n(\beta_0, \hat{\theta})(\hat{\beta} - \beta_0) + o_p(n^{1/2}).$$

Then we utilize $\mathbf{S}_n(\hat{\beta}, \hat{\theta}) = 0$ to deduce from the above equation that

$$-\mathbf{A}_n(\beta_0, \hat{\theta}) \left(n^{1/2}(\hat{\beta} - \beta_0) \right) = n^{-1/2} \mathbf{S}_n(\beta_0, \theta_0) + \mathbf{B}_n(\beta_0, \theta_0) \left(n^{1/2}(\hat{\theta} - \theta_0) \right) + o_p(1). \quad (\text{A.26})$$

After some basic algebra, the first term on the right side of Eq. (A.26) can be expressed as

$$\begin{aligned} n^{-1/2} \mathbf{S}_n(\beta_0, \theta_0) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau (\mathbf{Z}_i - \bar{\mathbf{Z}}(\beta_0, t)) dN_i^u(t) + \mathbf{U}_{n,3}(\beta_0, \theta_0) \\ &\quad + n^{-1/2} \sum_{i=1}^n \int_0^\tau (\delta_i - m(t, \mathbf{Z}_i, \theta_0)) (\bar{\mathbf{Z}}(\beta_0, t) - \bar{\mathbf{z}}(\beta_0, t)) dN_i(t). \end{aligned} \quad (\text{A.27})$$

We carry out in detail, once, an application of Lemma 2 of [11] to prove that a remainder term is asymptotically negligible – the same argument will recur at other places below. The third term on the right side of Eq. (A.27) can be expressed as

$$n^{1/2} \int_0^\tau (\bar{\mathbf{Z}}(\beta_0, t) - \bar{\mathbf{z}}(\beta_0, t)) dQ_n(t),$$

where $nQ_n(t) = \sum_{i=1}^n \int_0^\tau (\delta_i - m(t, \mathbf{Z}_i, \theta_0)) dN_i(t)$. Note that $n^{1/2}Q_n(\cdot)$ converges weakly to a zero-mean Gaussian process on $D[0, \tau]$ and that $\sup_{t \in [0, \tau]} |\bar{\mathbf{Z}}(\beta_0, t) - \bar{\mathbf{z}}(\beta_0, t)| \xrightarrow{P} 0$. We can apply Lemma 2 of [11] to conclude that the third term is $o_p(1)$, provided that $\bar{\mathbf{Z}}(\beta_0, s)$ has total variation bounded in probability (TVBP) and $\bar{\mathbf{z}}(\beta_0, s)$ has bounded variation (BV). The BV property extends to a product of two functions of BV. It extends as well to a reciprocal of a function of BV that is bounded away from zero; see page 130 of [2]. Since $\bar{\mathbf{z}}(\beta_0, s)$ is a ratio of two monotonic functions the denominator being uniformly bounded away from zero (condition A.3), it is of BV. Since $\bar{\mathbf{Z}}(\beta_0, s)$ is a ratio of $\mathbf{S}^{(1)}(\beta_0, s)$ and $\mathbf{S}^{(0)}(\beta_0, s)$, we first prove that $\mathbf{S}^{(1)}(\beta_0, s)$ and $\mathbf{S}^{(0)}(\beta_0, s)$ each have TVBP. For simplicity, consider a one-dimensional covariate ($p = 1$) and let $0 = t_0 < t_1 \dots < t_{k-1} < t_k = \tau$ denote an arbitrary partition of the interval $[0, \tau]$. Defining $\Delta \mathbf{S}^{(1)}(\beta_0, t_j) = \mathbf{S}^{(1)}(\beta_0, t_j) - \mathbf{S}^{(1)}(\beta_0, t_{j-1})$, we have

$$\begin{aligned} \sum_{j=1}^k |\Delta \mathbf{S}^{(1)}(\beta_0, t_j)| &= \sum_{j=1}^k \left| \frac{1}{n} \sum_{i=1}^n I(t_{j-1} \leq X_i < t_j) e^{\beta_0^T \mathbf{Z}_i} \mathbf{Z}_i \right| \\ &\leq \sum_{j=1}^k \frac{1}{n} \sum_{i=1}^n I(t_{j-1} \leq X_i < t_j) e^{\beta_0^T \mathbf{Z}_i} |\mathbf{Z}_i| \\ &\leq M^* \sum_{j=1}^k \frac{1}{n} \sum_{i=1}^n I(t_{j-1} \leq X_i < t_j), \end{aligned}$$

where M^* is a suitable upper bound. Interchanging the order of summation we have

$$\begin{aligned} \sum_{j=1}^k |\Delta \mathbf{S}^{(1)}(\beta_0, t_j)| &\leq M^* \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k I(t_{j-1} \leq X_i < t_j) \\ &= M^* \frac{1}{n} \sum_{i=1}^n I(X_i < \tau) \\ &= M^* (P(X < \tau) + o_p(1)) \\ &\leq M^* + o_p(1). \end{aligned}$$

Likewise for $\mathbf{S}^{(0)}(\beta_0, s)$. To complete the proof we will need to show that the property of TVBP is closed under reciprocal and product operations. Indeed, from $|\Delta l_n(t_j)| \leq \Delta g_n(t_j)/b^2$, where $l_n = 1/g_n$ and $g_n(x) \geq b > 0$ for all x , one can conclude that if g_n has TVBP then l_n also has TVBP. By conditions A.2 and A.3, for adequately large n , $\mathbf{S}^{(0)}(\beta_0, s)$ is bounded away from 0 for all s , hence its reciprocal has TVBP. The proof of Theorem 6.9 of [2] shows that $h = fg$ satisfies $|\Delta h(t_j)| \leq A|\Delta g(t_j)| + B|\Delta f(t_j)|$, where $A = \sup\{f(t) : t \in [0, \tau]\}$ and $B = \sup\{g(t) : t \in [0, \tau]\}$, from which we can conclude that $\bar{\mathbf{Z}}(\beta_0, s)$ has TVBP.

For the first term in Eq. (A.27), again apply Lemma 2 of [11] to obtain

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \int_0^\tau [(Z_i - \bar{Z}(\beta_0, t))] dN_i^u(t) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau [(Z_i - \bar{Z}(\beta_0, t))] dM_i^u(t) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^\tau [(Z_i - \bar{Z}(\beta_0, t))] dM_i^u(t) + o_p(1) \\ &= \mathbf{U}_{n,2}(\beta_0) + o_p(1). \end{aligned}$$

Then, applying condition D.1, it follows that

$$\mathbf{B}_n(\beta_0, \theta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau [Z_i - \bar{Z}(\beta_0, t)] [\text{Grad}(m(\mathbf{W}_i, \theta_0))]^T dN_i(t) + o_p(1) = \mathbf{B}_0 + o_p(1).$$

Eq. (A.21) now implies that the second term on the right side of Eq. (A.26) is $\mathbf{U}_{n,1}(\beta_0, \theta_0) + o_p(1)$. Finally, the consistency of $\hat{\theta}$ implies $\mathbf{A}_n(\beta_0, \hat{\theta}) \xrightarrow{P} -\Sigma_C$, establishing Eq. (A.25).

We now compute the second and cross product moments of $\mathbf{U}_{n,2}(\beta_0)$, $\mathbf{U}_{n,j}(\beta_0, \theta_0)$, $j = 1, 3$, given in Eq. (A.25). By applying iterated expectation with conditioning on \mathbf{W} we can show that

$$E \left[\left(\frac{\delta - m(\mathbf{W}, \theta_0)}{m(\mathbf{W}, \theta_0)\bar{m}(\mathbf{W}, \theta_0)} \right)^2 \text{Grad}_\theta(m(\mathbf{W}, \theta_0)) \text{Grad}^T(m(\mathbf{W}, \theta_0)) \right] = \mathbf{I}_0. \quad (\text{A.28})$$

It follows from Eq. (A.28) that

$$E [\mathbf{U}_{n,1}(\beta_0, \theta_0) \mathbf{U}_{n,1}^T(\beta_0, \theta_0)] = \mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{I}_0 (\mathbf{B}_0 \mathbf{I}_0^{-1})^T = \mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T. \quad (\text{A.29})$$

From [1] and Eq. (A.5), we have that

$$E [\mathbf{U}_{n,2}(\beta_0) \mathbf{U}_{n,2}^T(\beta_0)] = \int_0^\tau \mathbf{v}(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt = \Sigma_C. \quad (\text{A.30})$$

Again, applying iterated expectation with conditioning on X and Z , it follows that

$$\begin{aligned} E [\mathbf{U}_{n,3}(\beta_0, \theta_0) \mathbf{U}_{n,3}^T(\beta_0, \theta_0)] &= E [(\delta - m(X, Z, \theta_0))^2 (Z - \bar{Z}(\beta_0, X))^{\otimes 2} I(X \leq \tau)] \\ &= E [m(X, Z, \theta_0) \bar{m}(X, Z, \theta_0) (Z - \bar{Z}(\beta_0, X))^{\otimes 2} I(X \leq \tau)] \\ &= \Sigma_C - \mathbf{V}_0, \end{aligned} \quad (\text{A.31})$$

where Σ_C and \mathbf{V}_0 are defined by Eqs. (A.7) and (A.8) respectively. However, note that

$$\begin{aligned} E [\mathbf{U}_{n,2}(\beta_0) \mathbf{U}_{n,3}^T(\beta_0, \theta_0)] &= -E [\delta(\delta - m(X, Z, \theta_0))(Z - \bar{Z}(\beta_0, X))^{\otimes 2} I(X \leq \tau)] \\ &= -E [\mathbf{U}_{n,3}(\beta_0, \theta_0) \mathbf{U}_{n,2}^T(\beta_0, \theta_0)]. \end{aligned} \quad (\text{A.32})$$

Next observe that $E [\mathbf{U}_{n,1}(\beta_0, \theta_0) \mathbf{U}_{n,2}^T(\beta_0, \theta_0)]$ and $E [\mathbf{U}_{n,1}(\beta_0, \theta_0) \mathbf{U}_{n,3}^T(\beta_0, \theta_0)]$ cancel out:

$$\begin{aligned} E [\mathbf{U}_{n,1}(\beta_0, \theta_0) \mathbf{U}_{n,2}^T(\beta_0, \theta_0)] &= \mathbf{B}_0 \mathbf{I}_0^{-1} E \left[\left\{ \frac{\delta - m(\mathbf{W}, \theta_0)}{m(\mathbf{W}, \theta_0)\bar{m}(\mathbf{W}, \theta_0)} \text{Grad}(m(\mathbf{W}, \theta_0)) \right\} \right. \\ &\quad \times \left. \int_0^\tau \{Z - \bar{Z}(\beta_0, t)\}^T dM^u(t) \right] \\ &= \mathbf{B}_0 \mathbf{I}_0^{-1} E \left[\frac{\delta(\delta - m(\mathbf{W}, \theta_0))}{m(\mathbf{W}, \theta_0)\bar{m}(\mathbf{W}, \theta_0)} \right. \\ &\quad \times \left. \text{Grad}(m(\mathbf{W}, \theta_0))(Z - \bar{Z}(\beta_0, X))^T I(X \leq \tau) \right] \\ &= \mathbf{B}_0 \mathbf{I}_0^{-1} E [\text{Grad}(m(\mathbf{W}, \theta_0))(Z - \bar{Z}(\beta_0, X))^T I(X \leq \tau)], \end{aligned}$$

with $E [\mathbf{U}_{n,1}(\beta_0, \theta_0) \mathbf{U}_{n,3}^T(\beta_0, \theta_0)]$ computing to the negative of the last expression above. From Eqs. (A.29)–(A.32), we obtain the asymptotic covariance matrix Σ given by Eq. (A.10).

Remark. To prove the efficiency of $\hat{\beta}$, see Proposition 1, we will need a convenient, if slightly lengthy, form of Σ . From Eqs. (A.29)–(A.32), this alternate form is readily given by

$$\Sigma = \Sigma_C^{-1} \left[\mathbf{B}_0 \mathbf{I}_0^{-1} \mathbf{B}_0^T + \int_0^\tau \mathbf{v}(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt - E \left[m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} I(X \leq \tau) \right] \right] \Sigma_C^{-1}. \quad (\text{A.33})$$

Extension over $[0, \tau_H]$ Taylor's expansion of $\mathbf{S}_n(\hat{\beta}, \hat{\theta})$ about (β_0, θ_0) yields

$$\mathbf{S}_n(\hat{\beta}, \hat{\theta}) = \mathbf{S}_n(\beta_0, \theta_0) + \left(\frac{\partial}{\partial \beta} \mathbf{S}_n(\beta^*, \theta_0) \right) (\hat{\beta} - \beta_0) + \left(\frac{\partial}{\partial \theta} \mathbf{S}_n(\hat{\beta}, \theta^*) \right) (\hat{\theta} - \theta_0). \quad (\text{A.34})$$

To extend the above proof over $[0, \tau_H]$, we need to show that the tail parts of the integrals in the definitions of (i) $n^{-1/2} \mathbf{S}_n(\beta_0, \theta_0)$, (ii) $n^{-1} \partial \mathbf{S}_n(\beta^*, \theta_0) / \partial \beta$ and (iii) $n^{-1} \partial \mathbf{S}_n(\hat{\beta}, \theta^*) / \partial \theta$ are each negligible. To prove (i), it remains to show that the tail part of the integral in the definition of $\mathbf{S}_n(\beta_0, \theta_0)$, namely [see also Eq. (2.4)]

$$R_{n,1}(\tau, \tau_H; \beta_0, \theta_0) \equiv n^{-1/2} \sum_{i=1}^n \int_\tau^{\tau_H} m(t, \mathbf{Z}_i, \theta_0) [\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t)] dN_i(t),$$

is negligible. By the method of proof given in page 308 of [9], we show that, for any $\epsilon > 0$ and $\delta > 0$, $P \left\{ \sup_{\tau \leq s \leq \tau_H} |R_{n,1}(\tau, s; \beta_0, \theta_0)| > \epsilon \right\}$ is bounded above by

$$\frac{\delta}{\epsilon^2} + \frac{1}{\delta} E \left(\frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} (\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t))^2 Y_i(t) \exp(\beta_0^T \mathbf{Z}_i) d\Lambda_0(t) \right). \quad (\text{A.35})$$

The rest of the proof follows exactly as in pages 308 and 309 of [9]. From Eq. (A.2), it follows that

$$R_{n,1}(\tau, \tau_H; \beta_0, \theta_0) = n^{-1/2} \sum_{i=1}^n \int_\tau^{\tau_H} m(t, \mathbf{Z}_i, \theta_0) [\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t)] dM_i(t).$$

We assume $p = 1$ for simplicity. By the extension of Lengart's inequality indicated in Corollary 3.4.1 of [9], and for any $\epsilon > 0$ and $\delta > 0$, we have

$$P \left\{ \sup_{\tau \leq s \leq \tau_H} |R_{n,1}(\tau, s; \beta_0, \theta_0)| > \epsilon \right\} \leq \frac{\delta}{\epsilon^2} + P \left(\tilde{R}_{n,1}(\tau, \tau_H, \beta_0, \theta_0) > \delta \right), \quad (\text{A.36})$$

where, using Eq. (A.4) and condition A.6, we obtain

$$\begin{aligned} \tilde{R}_{n,1}(\tau, \tau_H, \beta_0, \theta_0) &\equiv \frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} m^2(t, \mathbf{Z}_i, \theta_0) [\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t)]^2 d\langle M_i, M_i \rangle(t) \\ &= \frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} m(t, \mathbf{Z}_i, \theta_0) [\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t)]^2 Y_i(t) \exp(\beta_0^T \mathbf{Z}_i) \lambda_0(t) dt. \end{aligned}$$

Applying Markov's inequality, the second term on the right hand side of Eq. (A.36) is bounded above by the second term on the right hand side of Eq. (A.35).

To prove (ii), for example, we have

$$R_{n,2}(\tau, \tau_H, \beta, \theta_0) = \frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} m(t, \mathbf{Z}_i, \theta_0) \left\{ \left(\frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^{\otimes 2} - \frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} \right\} dN_i(t).$$

To prove that $\sup_{\beta \in \mathcal{B}} |R_{n,2}(\tau, \tau_H, \beta, \theta_0)|$ is negligible, we will need to extend condition A.2 over all of $[0, \tau_H]$. Then, uniformly for $\beta \in \mathcal{B}$, we have

$$R_{n,2}(\tau, \tau_H, \beta, \theta_0) = \frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} m(t, \mathbf{Z}_i, \theta_0) \left\{ \left(\frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^{\otimes 2} - \frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} \right\} dN_i(t) + o_p(1),$$

so that

$$\begin{aligned} \sup_{\beta \in \mathcal{B}} |R_{n,2}(\tau, \tau_H, \beta, \theta_0)| &\leq \sup_{\beta \in \mathcal{B}, t \in [0, \tau_H]} \left| \left(\frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^{\otimes 2} - \frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} \right| \frac{1}{n} \sum_{i=1}^n \int_\tau^{\tau_H} dN_i(t) + o_p(1) \\ &= \sup_{\beta \in \mathcal{B}, t \in [0, \tau_H]} \left| \left(\frac{\mathbf{s}^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^{\otimes 2} - \frac{\mathbf{s}^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} \right| (H(\tau_H) - H(\tau)) + o_p(1), \end{aligned}$$

which is negligible as $\tau \uparrow \tau_H$.

A.1.3. Weak convergence of the baseline cumulative hazard estimator

Recalling $\alpha(t, X, \mathbf{Z}, \boldsymbol{\beta}_0)$ defined by Eq. (A.15), we introduce the quantities

$$\begin{aligned} L_{n,1}(t, \boldsymbol{\beta}_0) &= n^{-1/2} \sum_{i=1}^n \int_0^t \alpha(t, s, \mathbf{Z}_i, \boldsymbol{\beta}_0) dM_i^u(s), \\ L_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= -n^{-1/2} \sum_{i=1}^n \int_0^t (\delta_i - m(s, \mathbf{Z}_i, \boldsymbol{\theta}_0)) \alpha(t, s, \mathbf{Z}_i, \boldsymbol{\beta}_0) dN_i(s), \\ L_{n,3}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^n \frac{\delta_i - m(\mathbf{W}_i, \boldsymbol{\theta}_0)}{m(\mathbf{W}_i, \boldsymbol{\theta}_0) \bar{m}(\mathbf{W}_i, \boldsymbol{\theta}_0)} [\mathbf{d}_0(t)]^T \mathbf{I}_0^{-1} \text{Grad}(m(\mathbf{W}_i, \boldsymbol{\theta}_0)), \end{aligned}$$

and develop an asymptotic representation, uniformly for $t \in [0, \tau]$, given by

$$n^{1/2}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t)) = L_{n,1}(t, \boldsymbol{\beta}_0) + \sum_{j=2}^3 L_{n,j}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1). \quad (\text{A.37})$$

Let $\text{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}, \boldsymbol{\theta}))$ and $\text{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}, \boldsymbol{\theta}))$ be the vectors of partial derivatives of $\hat{\Lambda}_0(t, \boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ respectively. Let $\boldsymbol{\theta}^*$ denote a value on the line joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Define $\boldsymbol{\beta}^*$ likewise between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Taylor's expansions of $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$ yields,

$$\begin{aligned} \hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t) &= \left[\hat{\Lambda}_0(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \Lambda_0(t) \right] + \left\langle \text{Grad}_{\boldsymbol{\beta}} \left(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0) \right), (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right\rangle \\ &\quad + \left\langle \text{Grad}_{\boldsymbol{\theta}} \left(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*) \right), (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right\rangle. \end{aligned} \quad (\text{A.38})$$

First note that $\text{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \hat{\boldsymbol{\theta}})) = -\mathbf{C}_0(t) + o_p(1)$, uniformly over $[0, \tau]$. Indeed,

$$\text{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \hat{\boldsymbol{\theta}})) = -\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}^*, s)}{(\mathbf{S}^{(0)}(\boldsymbol{\beta}^*, s))^2} m(s, \mathbf{Z}_i, \hat{\boldsymbol{\theta}}) dN_i(s).$$

Since $\hat{\boldsymbol{\beta}}$ is consistent, we can apply conditions A.2 and A.3 to replace $\mathbf{S}^{(1)}(\boldsymbol{\beta}^*, s)/(\mathbf{S}^{(0)}(\boldsymbol{\beta}^*, s))^2$ in the integrand above with its limit $\mathbf{s}^{(1)}(\boldsymbol{\beta}_0, s)/(\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s))^2$ plus a remainder term, which is $o_p(1)$ uniformly for $s \in [0, \tau]$. It follows from the consistency of $\hat{\boldsymbol{\theta}}$, strong law of large numbers, and Eq. (A.2) that, uniformly for $t \in [0, \tau]$,

$$\text{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \hat{\boldsymbol{\theta}})) = -\int_0^t \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0, s)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s)} \lambda_0(s) ds + o_p(1) = -\mathbf{C}_0(t) + o_p(1).$$

Likewise, we can show that $\text{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)) = \mathbf{D}_0(t) + o_p(1)$ uniformly for $t \in [0, \tau]$. For the first term on the right side of Eq. (A.38), write $\Lambda_0^*(t) = \int_0^t I\{ \sum_{i=1}^n Y_i(x) > 0 \} \lambda_0(x) dx$, so that $\Lambda_0^*(t) - \Lambda_0(t) = o_p(n^{-1/2})$, see page 300 of [9]. Then

$$\begin{aligned} \hat{\Lambda}_0(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \Lambda_0(t) &= \left[\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dN_i^u(s)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}_0, s)} - \Lambda_0^*(t) \right] \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\delta_i - m(s, \mathbf{Z}_i, \boldsymbol{\theta}_0)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}_0, s)} dN_i(s) + o_p(n^{-1/2}). \end{aligned} \quad (\text{A.39})$$

The first term on the right side of Eq. (A.39) is the sum of $E_{n,1}(t, \boldsymbol{\beta}_0)$ given below and a remainder term, which, by Lengart's inequality (see, for example, page 308 of [9], where an extension of their corollary 3.4.1 is employed), is $o_p(n^{-1/2})$, uniformly for $t \in [0, \tau]$:

$$E_{n,1}(t, \boldsymbol{\beta}_0) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dM_i^u(s)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s)}.$$

The second term on the right side of Eq. (A.39) is the sum of $E_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ below and a remainder term, which, by Lemma 2 of [11], is $o_p(n^{-1/2})$, uniformly over $[0, \tau]$:

$$E_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = -\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\delta_i - m(s, \mathbf{Z}_i, \boldsymbol{\theta}_0)}{\mathbf{s}^{(0)}(\boldsymbol{\beta}_0, s)} dN_i(s).$$

From Eq. (A.25), the second term on the right side of Eq. (A.38) contributes *three* expressions. The *second* expression combined with $n^{1/2}E_{n,1}(t, \boldsymbol{\beta}_0)$ gives $L_{n,1}(t, \boldsymbol{\beta}_0)$. The *third* expression combined with $n^{1/2}E_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ gives $L_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$.

The first expression combined with the third term on the right side of Eq. (A.38) gives $L_{n,3}(t, \beta_0, \theta_0)$. Thus Eq. (A.37) holds. The multivariate central limit theorem now implies finite dimensional convergence.

It remains to prove tightness. First, $L_{n,1}(t, \beta_0)$ converges weakly in $D[0, \tau]$ to a zero-mean Gaussian martingale and is tight. Next, $L_{n,2}(t, \beta_0, \theta_0)$ can be decomposed into two sums and we show tightness of the first sum as follows: for $0 \leq t_1 \leq t_2 \leq \tau$ and $i = 1, \dots, n$, we write

$$\chi_i(\beta_0, \theta_0, t_1, t_2) = (\delta_i - m(X_i, \mathbf{Z}_i, \theta_0))I(t_1 < X_i \leq t_2)/s^{(0)}(\beta_0, X_i)$$

and note that, conditioning by X and \mathbf{Z} , $E(\chi_i(\beta_0, \theta_0, t_1, t_2)) = 0$. Then we have that

$$\begin{aligned} & \overline{\lim}_{n \rightarrow \infty} E \left[n^{-1/2} \sum_{i=1}^n \chi_i(\beta_0, \theta_0, t_1, t_2) \right]^4 \\ &= \overline{\lim}_{n \rightarrow \infty} E \left[\frac{1}{n^2} \left(\sum_{i=1}^n \chi_i^4(\beta_0, \theta_0, t_1, t_2) + 3 \sum_{i \neq j=1}^n \chi_i^2(\beta_0, \theta_0, t_1, t_2) \chi_j^2(\beta_0, \theta_0, t_1, t_2) \right) \right] \\ &\leq \overline{\lim}_{n \rightarrow \infty} E \left[\frac{1}{n^2} \left(3 \sum_{i=1}^n \chi_i^4(\beta_0, \theta_0, t_1, t_2) + 3 \sum_{i \neq j=1}^n \chi_i^2(\beta_0, \theta_0, t_1, t_2) \chi_j^2(\beta_0, \theta_0, t_1, t_2) \right) \right] \\ &= 3 \overline{\lim}_{n \rightarrow \infty} E \left[\frac{1}{n} \sum_{i=1}^n \chi_i^2(\beta_0, \theta_0, t_1, t_2) \right]^2 \\ &\leq 3E \left[\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \chi_i^2(\beta_0, \theta_0, t_1, t_2) \right]^2 \quad (\text{see problem 3.2.6, page 47 of [5]}) \\ &= 3[E(\chi^2(\beta_0, \theta_0, t_1, t_2))]^2 = 3(\mu(t_2) - \mu(t_1))^2, \end{aligned}$$

where $\mu(t) = \int_0^t dG^1(s, \mathbf{z})/s^{(0)}(\beta_0, s)^2$; see the proof of Proposition 1 where $G^1(t, \mathbf{z})$ was defined. Note that $\mu(\cdot)$ is a finite and continuous measure allowing us to appeal to formula (30) on page 52 of [19]. Tightness of the second sum of $L_{n,2}(t, \beta_0, \theta_0)$ as well as that of $L_{n,3}(t, \beta_0, \theta_0)$ follows likewise. Finite dimensional convergence and tightness of $n^{1/2}(\hat{\Lambda}_0(\cdot, \hat{\beta}, \hat{\theta}) - \Lambda_0(\cdot))$ imply its weak convergence in $D[0, \tau]$.

We shall suppress β_0 and θ_0 from the $L_{n,j}$, $j = 1, 2, 3$. Using Eq. (A.3), for $0 \leq t_1, t_2 \leq \tau$,

$$\begin{aligned} E(L_{n,1}(t_1)L_{n,1}(t_2)) &= E \int_0^\tau (\alpha(t_1, s, \mathbf{Z}, \beta_0)) (\alpha(t_2, s, \mathbf{Z}, \beta_0)) Y(s) e^{\beta_0^T \mathbf{Z}} \lambda_0(s) ds \\ &= E \int_0^\tau (\alpha(t_1, s, \mathbf{Z}, \beta_0)) (\alpha(t_2, s, \mathbf{Z}, \beta_0)) dN^u(s) \\ &= E[m(X, \mathbf{Z}, \theta_0) (\alpha(t_1, X, \mathbf{Z}, \beta_0)) (\alpha(t_2, X, \mathbf{Z}, \beta_0)) I(X \leq \tau)]. \end{aligned} \quad (\text{A.40})$$

In the intermediate calculations here, we write $V(\delta|X, \mathbf{Z}) = m(X, \mathbf{Z}, \theta_0)\bar{m}(X, \mathbf{Z}, \theta_0)$ for the conditional variance of δ given X and \mathbf{Z} . Since $E(\delta|X, \mathbf{Z}) = m(X, \mathbf{Z}, \theta_0)$ we have that

$$\begin{aligned} E(L_{n,2}(t_1)L_{n,2}(t_2)) &= E[(\delta - m(X, \mathbf{Z}, \theta_0))^2 (\alpha(t_1, X, \mathbf{Z}, \beta_0)) (\alpha(t_2, X, \mathbf{Z}, \beta_0)) I(X \leq \tau)] \\ &= E[V(\delta|X, \mathbf{Z}) (\alpha(t_1, X, \mathbf{Z}, \beta_0)) (\alpha(t_2, X, \mathbf{Z}, \beta_0)) I(X \leq \tau)]. \end{aligned} \quad (\text{A.41})$$

Furthermore, applying Eq. (A.28), it follows that

$$E(L_{n,3}(t_1)L_{n,3}(t_2)) = [\mathbf{d}_0(t_1)]^T \mathbf{I}_0^{-1} \mathbf{I}_0 \mathbf{I}_0^{-1} [\mathbf{d}_0(t_2)] = [\mathbf{d}_0(t_1)]^T \mathbf{I}_0^{-1} [\mathbf{d}_0(t_2)]. \quad (\text{A.42})$$

The cross-product moment computations are similar. Using Eq. (A.3) we again obtain

$$\begin{aligned} E(L_{n,1}(t_1)L_{n,2}(t_2)) &= -E[(\delta - m(X, \mathbf{Z}, \theta_0)) (\alpha(t_1, X, \mathbf{Z}, \beta_0)) (\alpha(t_2, X, \mathbf{Z}, \beta_0)) I(X \leq \tau)] \\ &= -E(L_{n,2}(t_1)L_{n,2}(t_2)) \\ &= -E(L_{n,1}(t_2)L_{n,2}(t_1)). \end{aligned} \quad (\text{A.43})$$

Finally, analogous calculations show that

$$\begin{aligned} E(L_{n,1}(t_1)L_{n,3}(t_2)) &= -E(L_{n,2}(t_1)L_{n,3}(t_2)) \\ E(L_{n,1}(t_2)L_{n,3}(t_1)) &= -E(L_{n,2}(t_2)L_{n,3}(t_1)). \end{aligned} \quad (\text{A.44})$$

From Eqs. (A.40)–(A.43) and (A.43), we obtain the asymptotic covariance function $\sigma(t_1, t_2)$ given by Eq. (A.17). When $t_1 = t_2 = t$, we have the variance function $\sigma(t, t)$ given by Eq. (A.18).

Remark. To prove the efficiency of $\hat{\Lambda}(t, \hat{\beta}, \hat{\theta})$, see Proposition 1, we will need the form

$$\begin{aligned} \sigma(t, t) &= \int_0^t \frac{\lambda_0(s)}{s^{(0)}(\beta_0, s)} ds + [\mathbf{C}_0(t)]^T \Sigma_C^{-1} \mathbf{C}_0(t) + [\mathbf{d}_0(t)]^T \mathbf{I}_0^{-1} \mathbf{d}_0(t) \\ &\quad - E \left[m(X, \mathbf{Z}, \theta_0) \bar{m}(X, \mathbf{Z}, \theta_0) (\alpha(t, X, \mathbf{Z}, \beta_0))^2 I(X \leq \tau) \right]. \end{aligned} \quad (\text{A.45})$$

Note that, when $t_1 = t_2 = t$, the third term of Eq. (A.45) is just the right side of Eq. (A.42). Similarly, the fourth term of Eq. (A.45) is just the sum of the right side of Eq. (A.41) and two times the right side of Eq. (A.43). It remains to show that the first two terms of Eq. (A.45) are contributed by the right side of Eq. (A.40). Accordingly for $t_1 = t_2 = t$, write the latter quantity as $T_1 + T_2 + T_3$ and note that

$$\begin{aligned} T_1 &= E \left(\frac{m(X, \mathbf{Z}, \theta_0)}{(s^{(0)}(\beta_0, X))^2} I(X \leq t) \right) = E \int_0^t \frac{1}{(s^{(0)}(\beta_0, s))^2} dN^u(s) \\ &= E \int_0^t \frac{1}{(s^{(0)}(\beta_0, s))^2} Y(s) e^{\beta_0^T \mathbf{Z}} \lambda_0(s) ds = \int_0^t \frac{\lambda_0(s)}{s^{(0)}(\beta_0, s)} ds, \end{aligned}$$

which is the first term on the right side of Eq. (A.45). Furthermore, using Eq. (A.7) we have

$$\begin{aligned} T_3 &= E \left[m(X, \mathbf{Z}, \theta_0) [\mathbf{C}_0(t)]^T \Sigma_C^{-1} (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} \Sigma_C^{-1} \mathbf{C}_0(t) I(X \leq \tau) \right] \\ &= [\mathbf{C}_0(t)]^T \Sigma_C^{-1} E \left[m(X, \mathbf{Z}, \theta_0) (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X))^{\otimes 2} I(X \leq \tau) \right] \Sigma_C^{-1} \mathbf{C}_0(t) \\ &= [\mathbf{C}_0(t)]^T \Sigma_C^{-1} \mathbf{C}_0(t), \end{aligned}$$

which is the second term on the right side of Eq. (A.45). Finally, the term T_2 is zero as follows:

$$\begin{aligned} T_2 &= -2E \left(m(X, \mathbf{Z}, \theta_0) \frac{I(X \leq t)}{s^{(0)}(\beta_0, X)} [\mathbf{C}_0(t)]^T \Sigma_C^{-1} (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X)) I(X \leq \tau) \right) \\ &= -2E \left(\frac{m(X, \mathbf{Z}, \theta_0)}{s^{(0)}(\beta_0, X)} [\mathbf{C}_0(t)]^T \Sigma_C^{-1} (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, X)) N(t) \right) \\ &= -2E \left(\int_0^t \frac{1}{s^{(0)}(\beta_0, s)} [\mathbf{C}_0(t)]^T \Sigma_C^{-1} (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, s)) m(s, \mathbf{Z}, \theta_0) dN(s) \right) \\ &= -2E \int_0^t \frac{1}{s^{(0)}(\beta_0, s)} [\mathbf{C}_0(t)]^T \Sigma_C^{-1} (\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, s)) Y(s) e^{\beta_0^T \mathbf{Z}} \lambda_0(s) ds. \end{aligned}$$

But $E \left[(\mathbf{Z} - \bar{\mathbf{z}}(\beta_0, s)) Y(s) e^{\beta_0^T \mathbf{Z}} \right] = \mathbf{s}^{(1)}(\beta_0, s) - \bar{\mathbf{z}}(\beta_0, s) s^{(0)}(\beta_0, s) = 0$ and hence $T_2 = 0$.

A.2. Proof of Theorem 2

A.2.1. Distribution of $\hat{\beta}$

Defining the normalized sums

$$\begin{aligned} \tilde{\mathbf{U}}_{n,1}(\beta_0, \theta_0) &= n^{-1/2} \sum_{i=1}^n \mathbf{B}_0 \mathbf{I}_0^{-1} \frac{\xi_i(\delta_i - m(\mathbf{W}_i, \theta_0))}{m(\mathbf{W}_i, \theta_0) \bar{m}(\mathbf{W}_i, \theta_0)} \text{Grad}(m(\mathbf{W}_i, \theta_0)) \\ \tilde{\mathbf{U}}_{n,2}(\beta_0) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t)] dM_i^u(t) \\ \tilde{\mathbf{U}}_{n,3}(\beta_0, \theta_0) &= -n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [(\delta_i - m(t, \mathbf{Z}_i, \theta_0)) (\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t))] dN_i(t) \\ \tilde{\mathbf{U}}_{n,4}(\beta_0, \theta_0) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau (1 - \xi_i) [m(t, \mathbf{Z}_i, \theta_0) (\mathbf{Z}_i - \bar{\mathbf{z}}(\beta_0, t))] dM_i(t), \end{aligned}$$

we develop an asymptotic representation given by

$$n^{1/2}(\hat{\beta} - \beta_0) = \Sigma_C^{-1} \left[\tilde{\mathbf{U}}_{n,1}(\beta_0, \theta_0) + \tilde{\mathbf{U}}_{n,2}(\beta_0) + \sum_{j=3}^4 \tilde{\mathbf{U}}_{n,j}(\beta_0, \theta_0) \right] + o_p(1). \quad (\text{A.46})$$

Basically, we will require approximations for $n^{-1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, see Eq. (A.26). Under MAR, we have the following representation for the MLE of $\boldsymbol{\theta}$ (cf. [22] or [25]):

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{-1/2} \sum_{i=1}^n \tilde{\mathbf{I}}_0^{-1} \frac{\xi_i(\delta_i - m(\mathbf{W}_i, \boldsymbol{\theta}_0))}{m(\mathbf{W}_i, \boldsymbol{\theta}_0)\bar{m}(\mathbf{W}_i, \boldsymbol{\theta}_0)} \text{Grad}(m(\mathbf{W}_i, \boldsymbol{\theta}_0)) + o_p(1). \quad (\text{A.47})$$

From Eq. (A.47), it follows that the second term on the right side of Eq. (A.26) produces $\tilde{\mathbf{U}}_{n,1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. Next write $n^{-1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \mathbf{I}_1 + \mathbf{I}_2$, where

$$\begin{aligned} \mathbf{I}_1 &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0)(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] dN_i(t), \\ \mathbf{I}_2 &= n^{-1/2} \sum_{i=1}^n \int_0^\tau (1 - \xi_i) [m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0)(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] dN_i(t). \end{aligned}$$

However, \mathbf{I}_1 can be further decomposed as a sum of \mathbf{I}_{11} , \mathbf{I}_{12} and $\tilde{\mathbf{U}}_{n,3}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, where

$$\begin{aligned} \mathbf{I}_{11} &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t)] dN_i^u(t), \\ \mathbf{I}_{12} &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [(\delta_i - m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0))(\bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t) - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))] dN_i(t). \end{aligned}$$

It is straightforward to invoke Lemma 2 of [11] and show that $\mathbf{I}_{12} = o_p(1)$. Let $n\hat{\rho} = \sum_{i=1}^n \xi_i$. Using Eq. (A.3), we obtain

$$\begin{aligned} \mathbf{I}_{11} &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i [\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t)] dM_i^u(t) \\ &\quad + n^{-1/2} \sum_{i=1}^n \int_0^\tau (\xi_i - \hat{\rho}) [(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] Y_i(t) e^{\beta_0^T \mathbf{Z}} \lambda_0(t) dt \\ &= \tilde{\mathbf{U}}_{n,2}(\boldsymbol{\beta}_0) + n^{-1/2} \sum_{i=1}^n \int_0^\tau (\xi_i - \hat{\rho}) [(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] Y_i(t) e^{\beta_0^T \mathbf{Z}} \lambda_0(t) dt + o_p(1), \end{aligned}$$

It is straightforward to show that

$$\begin{aligned} \mathbf{I}_2 &= n^{-1/2} \sum_{i=1}^n \int_0^\tau (1 - \xi_i) [m(t, \mathbf{Z}_i, \boldsymbol{\theta}_0)(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] dM_i(t) \\ &\quad + n^{-1/2} \sum_{i=1}^n \int_0^\tau ((1 - \xi_i) - (1 - \hat{\rho})) [(\mathbf{Z}_i - \bar{\mathbf{Z}}(\boldsymbol{\beta}_0, t))] Y_i(t) e^{\beta_0^T \mathbf{Z}} \lambda_0(t) dt + o_p(1). \end{aligned}$$

The first term of \mathbf{I}_2 is $\tilde{\mathbf{U}}_{n,4}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ and the second term cancels out the second term of \mathbf{I}_{11} . Eq. (A.46) holds. Asymptotic normality follows by the multivariate central limit theorem.

By MAR and applying iterated expectation with conditioning on W we can show that

$$E \left[\left(\frac{\xi(\delta - m(\mathbf{W}, \boldsymbol{\theta}_0))}{m(\mathbf{W}, \boldsymbol{\theta}_0)\bar{m}(\mathbf{W}, \boldsymbol{\theta}_0)} \right)^2 \text{Grad}(m(\mathbf{W}, \boldsymbol{\theta}_0)) \text{Grad}^T(m(\mathbf{W}, \boldsymbol{\theta}_0)) \right] = \tilde{\mathbf{I}}_0. \quad (\text{A.48})$$

In the moment calculations below, we shall suppress $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$ appearing in the $\tilde{\mathbf{U}}_{n,j}$. Martingale integrals in our expectation calculations contribute 0. We will often utilize Eq. (A.3).

$$\begin{aligned} E[\tilde{\mathbf{U}}_{n,1}^{\otimes 2}] &= \mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} \tilde{\mathbf{I}}_0 (\mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1})^T = \mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} \mathbf{B}_0^T, \\ E[\tilde{\mathbf{U}}_{n,2}^{\otimes 2}] &= E \left[\int_0^\tau \pi(t, \mathbf{Z})(\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} Y(t) \lambda_0(t) e^{\beta_0^T \mathbf{Z}} dt \right] \\ &= E \left[\int_0^\tau \pi(t, \mathbf{Z})(\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} d\{N^u(t) - M^u(t)\} \right] \\ &= E[\pi(X, \mathbf{Z})m(X, \mathbf{Z}, \boldsymbol{\theta}_0)(\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)], \end{aligned}$$

$$\begin{aligned}
E \left[\tilde{\mathbf{U}}_{n,3}^{\otimes 2} \right] &= E \left[\xi (\delta - m(X, \mathbf{Z}, \boldsymbol{\theta}_0))^2 (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right] \\
&= E \left[\pi(X, \mathbf{Z}) m(X, \mathbf{Z}, \boldsymbol{\theta}_0) \bar{m}(X, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right], \\
E \left[\tilde{\mathbf{U}}_{n,4}^{\otimes 2} \right] &= E \left[(1 - \xi) (m(X, \mathbf{Z}, \boldsymbol{\theta}_0))^2 (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} d\langle M, M \rangle(t) \right] \\
&= E \left[\int_0^\tau (1 - \pi(t, \mathbf{Z})) m(t, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} Y(t) \lambda_0(t) e^{\boldsymbol{\beta}_0^T \mathbf{Z}} dt \right].
\end{aligned}$$

Using Eq. (A.3), it follows that

$$E \left[\tilde{\mathbf{U}}_{n,4}^{\otimes 2} \right] = E \left[(1 - \pi(X, \mathbf{Z})) m(X, \mathbf{Z}, \boldsymbol{\theta}_0)^2 (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right].$$

Next, noting that δ are MAR and applying Eq. (A.3), we obtain

$$\begin{aligned}
E \left[\tilde{\mathbf{U}}_{n,1} \tilde{\mathbf{U}}_{n,2}^T \right] &= \mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} E \left[\frac{\xi \delta (\delta - m(\mathbf{W}, \boldsymbol{\theta}_0))}{m(\mathbf{W}, \boldsymbol{\theta}_0) \bar{m}(\mathbf{W}, \boldsymbol{\theta}_0)} \text{Grad}(m(\mathbf{W}, \boldsymbol{\theta}_0)) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^T I(X \leq \tau) \right] \\
&= \mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} E \left[\pi(X, \mathbf{Z}) \text{Grad}(m(\mathbf{W}, \boldsymbol{\theta}_0)) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^T I(X \leq \tau) \right]
\end{aligned}$$

Another calculation as above leads to $E \left[\tilde{\mathbf{U}}_{n,1} \tilde{\mathbf{U}}_{n,3}^T \right] = -E \left[\tilde{\mathbf{U}}_{n,1} \mathbf{U}_{n,2}^T \right]$. Finally, it is seen that

$$\begin{aligned}
E \left[\tilde{\mathbf{U}}_{n,2} \tilde{\mathbf{U}}_{n,3}^T \right] &= -E \left[\xi \delta (\delta - m(X, \mathbf{Z}, \boldsymbol{\theta}_0)) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right] \\
&= -E \left[\pi(X, \mathbf{Z}) m(X, \mathbf{Z}, \boldsymbol{\theta}_0) \bar{m}(X, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right] \\
&= -E \left[\tilde{\mathbf{U}}_{n,3} \tilde{\mathbf{U}}_{n,3}^T \right].
\end{aligned}$$

Also, $E(\tilde{\mathbf{U}}_{n,4} \tilde{\mathbf{U}}_{n,j}^T) = 0$, $j = 1, 2, 3$. Combining all the direct and cross-product moment expressions, the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is given by

$$\begin{aligned}
\boldsymbol{\Sigma}_M &= \boldsymbol{\Sigma}_C^{-1} \left[\mathbf{B}_0 \tilde{\mathbf{I}}_0^{-1} \mathbf{B}_0^T + E \left\{ \pi(X, \mathbf{Z}) m(X, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right\} \right. \\
&\quad + E \left\{ (1 - \pi(X, \mathbf{Z})) m(X, \mathbf{Z}, \boldsymbol{\theta}_0)^2 (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right\} \\
&\quad \left. - \left\{ \pi(X, \mathbf{Z}) m(X, \mathbf{Z}, \boldsymbol{\theta}_0) \bar{m}(X, \mathbf{Z}, \boldsymbol{\theta}_0) (\mathbf{Z} - \bar{\mathbf{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right\} \right] \boldsymbol{\Sigma}_C^{-1}.
\end{aligned} \tag{A.49}$$

Combining the last three terms inside the square brackets on the right hand side of Eq. (A.49), the asymptotic covariance matrix simplifies yielding Eq. (A.11).

A.3. Weak convergence of baseline cumulative hazard

We introduce the quantities

$$\begin{aligned}
\tilde{L}_{n,1}(t, \boldsymbol{\beta}_0) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i \alpha(t, s, \mathbf{Z}_i, \boldsymbol{\beta}_0) dM_i^u(s), \\
\tilde{L}_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= -n^{-1/2} \sum_{i=1}^n \int_0^\tau \xi_i (\delta_i - m(s, \mathbf{Z}_i, \boldsymbol{\theta}_0)) \alpha(t, s, \mathbf{Z}_i, \boldsymbol{\beta}_0) dN_i(s), \\
\tilde{L}_{n,3}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^n \frac{\xi_i (\delta_i - m(\mathbf{W}_i, \boldsymbol{\theta}_0))}{m(\mathbf{W}_i, \boldsymbol{\theta}_0) \bar{m}(\mathbf{W}_i, \boldsymbol{\theta}_0)} \left[[\mathbf{d}_0(t)]^T \tilde{\mathbf{I}}_0^{-1} \text{Grad}(m(\mathbf{W}_i, \boldsymbol{\theta}_0)) \right], \\
\tilde{L}_{n,4}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^n \int_0^\tau (1 - \xi_i) m(s, \mathbf{Z}_i, \boldsymbol{\theta}_0) \alpha(t, s, \mathbf{Z}_i, \boldsymbol{\beta}_0) dM_i(s),
\end{aligned}$$

and develop an asymptotic representation, uniformly for $t \in [0, \tau]$, given by

$$n^{1/2}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t)) = \tilde{L}_{n,1}(t, \boldsymbol{\beta}_0) + \sum_{j=2}^4 \tilde{L}_{n,j}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1). \tag{A.50}$$

Each $\tilde{L}_{n,k}(t, \beta_0, \theta_0)$, $k = 2, 3$ and $\tilde{L}_{n,1}(t, \beta_0)$ are complete case normalized sum, same as $L_{n,k}(t, \beta_0, \theta_0)$ and $L_{n,1}(t, \beta_0)$ defined in Appendix A.1.3, but with ξ_i attached to each summand and \mathbf{I}_0 replaced with $\tilde{\mathbf{I}}_0$. The proof follows the methods described in Appendix A.1.3. Specifically, Eq. (A.38) applies and it suffices to derive asymptotic representations for the three quantities on its right hand side. Because of Eq. (A.47) and consistency of $\hat{\beta}$, both $\text{Grad}_{\beta}(\hat{\Lambda}_0(t, \beta^*, \theta_0))$ and $\text{Grad}_{\theta}(\hat{\Lambda}_0(t, \hat{\beta}, \theta^*))$ still converge in probability to $-\mathbf{C}_0(t)$ and $\mathbf{D}_0(t)$ respectively, uniformly for $t \in [0, \tau]$. The first term on the right side of Eq. (A.38), namely

$$\hat{\Lambda}_0(t, \beta_0, \theta_0) - \Lambda_0(t) \equiv \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{m(s, \mathbf{Z}_i, \theta_0)}{S^{(0)}(\beta_0, s)} dN_i(s) - \Lambda_0(t),$$

can be expressed as $\tilde{E}_1(t) + \tilde{E}_2(t) + \tilde{E}_3(t) - \Lambda_0(t)$, where

$$\begin{aligned} \tilde{E}_1(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\xi_i \delta_i}{S^{(0)}(\beta_0, s)} dN_i(s), \\ \tilde{E}_2(t) &= -\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\xi_i (\delta_i - m(s, \mathbf{Z}_i, \theta_0))}{S^{(0)}(\beta_0, s)} dN_i(s) \\ \tilde{E}_3(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{(1 - \xi_i) m(s, \mathbf{Z}_i, \theta_0)}{S^{(0)}(\beta_0, s)} dN_i(s). \end{aligned}$$

Using Eq. (A.3) followed by an application of Lengart's inequality, $\tilde{E}_1(t)$ can be further decomposed as the following sum plus $o_p(n^{-1/2})$:

$$\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\xi_i}{s^{(0)}(\beta_0, s)} dM_i^u(s) + \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{(\xi_i - \hat{\rho})}{S^{(0)}(\beta_0, s)} Y_i(s) \lambda_0(s) e^{\beta_0^T \mathbf{Z}_i} ds + \hat{\rho} \Lambda_0(t). \quad (\text{A.51})$$

Write $\tilde{E}_{11}(t)$ for the first term of Eq. (A.51). Apply Lemma 2 of [11] to obtain

$$\tilde{E}_2(t) = -\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\xi_i (\delta_i - m(s, \mathbf{Z}_i, \theta_0))}{s^{(0)}(\beta_0, s)} dN_i(s) + o_p(n^{-1/2}). \quad (\text{A.52})$$

Likewise, we can also show that

$$\begin{aligned} \tilde{E}_3(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{(1 - \xi_i) m(s, \mathbf{Z}_i, \theta_0)}{s^{(0)}(\beta_0, s)} dM_i(s) - \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\xi_i - \hat{\rho}}{S^{(0)}(\beta_0, s)} Y_i(s) \lambda_0(s) e^{\beta_0^T \mathbf{Z}_i} ds \\ &\quad + (1 - \hat{\rho}) \Lambda_0(t) + o_p(n^{-1/2}). \end{aligned} \quad (\text{A.53})$$

Write $\tilde{E}_{31}(t)$ for the first term of Eq. (A.53). Note that the second terms of Eqs. (A.51) and (A.53) cancel out. Because of Eq. (A.46), the second term of Eq. (A.38) contributes four expressions. The *second* expression combined with $n^{1/2} \tilde{E}_{11}(t)$ gives $\tilde{L}_{n,1}(t, \beta_0)$. The *third* expression combined with $n^{1/2} \tilde{E}_2(t)$ gives $\tilde{L}_{n,2}(t, \beta_0, \theta_0) + o_p(1)$. The *first* expression combined with the third term on the right side of Eq. (A.38) gives $\tilde{L}_{n,3}(t, \beta_0, \theta_0)$. The *fourth* expression combined with $n^{1/2} \tilde{E}_{31}(t)$ gives $\tilde{L}_{n,4}(t, \beta_0, \theta_0)$, and Eq. (A.50) holds. Finite dimensional convergence follows by the multivariate central limit theorem and tightness can be verified as described before.

Write $\alpha_{t_1, t_2}(X, \mathbf{Z}, \beta_0) = \alpha(t_1, X, \mathbf{Z}, \beta_0) \alpha(t_2, X, \mathbf{Z}, \beta_0)$. For the covariance function, analogous to the calculations given in Appendix A.1.3, the following expressions can be verified.

$$E(\tilde{L}_{n,1}(t_1) \tilde{L}_{n,1}(t_2)) = E[\pi(X, \mathbf{Z}) m(X, \mathbf{Z}, \theta_0) \alpha_{t_1, t_2}(X, \mathbf{Z}, \beta_0) I(X \leq \tau)], \quad (\text{A.54})$$

$$E(\tilde{L}_{n,2}(t_1) \tilde{L}_{n,2}(t_2)) = E[\pi(X, \mathbf{Z}) V(\delta | X, \mathbf{Z}) \alpha_{t_1, t_2}(X, \mathbf{Z}, \beta_0) I(X \leq \tau)], \quad (\text{A.55})$$

$$E(\tilde{L}_{n,3}(t_1) \tilde{L}_{n,3}(t_2)) = [\mathbf{d}_0(t_1)]^T \tilde{\mathbf{I}}_0^{-1} \tilde{\mathbf{I}}_0 \tilde{\mathbf{I}}_0^{-1} [\mathbf{d}_0(t_2)] = [\mathbf{d}_0(t_1)]^T \tilde{\mathbf{I}}_0^{-1} [\mathbf{d}_0(t_2)], \quad (\text{A.56})$$

$$E(\tilde{L}_{n,4}(t_1) \tilde{L}_{n,4}(t_2)) = E[(1 - \pi(X, \mathbf{Z})) m^2(X, \mathbf{Z}, \theta_0) \alpha_{t_1, t_2}(X, \mathbf{Z}, \beta_0) I(X \leq \tau)], \quad (\text{A.57})$$

$$E(\tilde{L}_{n,1}(t_1) \tilde{L}_{n,2}(t_2)) = E(\tilde{L}_{n,1}(t_2) \tilde{L}_{n,2}(t_1)) = -E(\tilde{L}_{n,2}(t_1) \tilde{L}_{n,2}(t_2)), \quad (\text{A.58})$$

$$E(\tilde{L}_{n,1}(t_1) \tilde{L}_{n,3}(t_2)) = -E(\tilde{L}_{n,2}(t_1) \tilde{L}_{n,3}(t_2)), \quad (\text{A.59})$$

$$E(\tilde{L}_{n,1}(t_2) \tilde{L}_{n,3}(t_1)) = -E(\tilde{L}_{n,2}(t_2) \tilde{L}_{n,3}(t_1)). \quad (\text{A.60})$$

Since $\tilde{L}_{n,4}(\cdot, \beta_0, \theta_0)$ is orthogonal to $\tilde{L}_{n,j}(\cdot, \beta_0, \theta_0)$, $j = 1, 2, 3$, we obtain from Eqs. (A.54)–(A.60) the final expression of the limiting covariance function $\sigma_M(t_1, t_2)$ given by Eq. (A.19).

References

- [1] P.K. Andersen, R.D. Gill, Cox's regression model for counting processes: a large sample study, *Ann. Statist.* 10 (1982) 1100–1120.
- [2] T.M. Apostol, *Mathematical Analysis*, second ed., Addison Wesley, Reading, MA, 1974.
- [3] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York, 1993.
- [4] N.E. Breslow, Contribution to the discussion on the paper by DR Cox, regression and life tables, *J. R. Stat. Soc. Ser. B* 34 (1972) 216–217.
- [5] K.L. Chung, *A Course in Probability Theory*, third ed., Academic Press, San Diego, CA, 2001.
- [6] D. Collett, *Modelling Binary Data*, second ed., CRS Press, Boca Raton, FL, 2002.
- [7] G. Dikta, On semiparametric random censorship models, *J. Statist. Plann. Inference* 66 (1998) 253–279.
- [8] L.A. Escobar, W.Q. Meeker, Assessing influence in regression analysis with censored data, *Biometrics* 48 (1992) 507–528.
- [9] T.R. Fleming, D.P. Harrington, *Counting Processes and Survival Analysis*, Wiley, New York, 1991.
- [10] J. Fox, *An R and S Plus Companion to Applied Regression*, 1st edition, Sage Publications Ltd, London, 2002.
- [11] P.B. Gilbert, I.W. McKeague, Y. Sun, The two-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy, *Biostatistics* 9 (2008) 263–276.
- [12] J.A. Koziol, S.B. Green, A Cramer-von mises statistic for randomly censored data, *Biometrika* 63 (1976) 465–474.
- [13] E.L. Lehmann, G. Casella, *Theory of Point Estimation*, Springer, New York, 1998.
- [14] C. Liu, Q. Wang, Semiparametric estimation for regression coefficients in the Cox model with failure indicators missing at random, *Statist. Sinica* 20 (2010) 1125–1142.
- [15] W. Lu, Maximum likelihood estimation in the proportional hazards cure model, *Ann. Inst. Statist. Math.* 60 (2008) 545–574.
- [16] M.D. Perlman, On the strong consistency of approximate maximum likelihood estimates, in: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1972, pp. 263–281.
- [17] C.R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.
- [18] P.H. Rossi, R.A. Berk, K.J. Lenihan, *Money, Work and Crime: Some Experimental Results*, Academic Press, New York, 1980.
- [19] G.R. Shorack, J.A. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.
- [20] M. Siotani, Some applications of Loewner's ordering on symmetric matrices, *Ann. Inst. Statist. Math.* 19 (1967) 245–259.
- [21] S. Subramanian, Model-based likelihood confidence intervals for survival functions, *Statist. Probab. Lett.* 82 (2012) 626–635.
- [22] S. Subramanian, The missing censoring-indicator model of random censorship, in: N. Balakrishnan, C.R. Rao. (Eds.), *Advances in Survival Analysis*, in: *Handbook of Statistics*, 23, 2004, pp. 123–141.
- [23] S. Subramanian, Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazard, *J. Statist. Plann. Inference* 84 (2000) 81–94.
- [24] A.A. Tsiatis, A large sample study of Cox's regression model, *Ann. Stat.* 9 (1981) 93–108.
- [25] A.A. Tsiatis, M. Davidian, B. Mcnney, Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure, *Biometrika* 89 (2002) 238–244.
- [26] A. Van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
- [27] M. Yuan, Semiparametric censorship model with covariates, *Test* 14 (2005) 1–26.
- [28] D. Zeng, D.Y. Lin, Maximum likelihood estimation in semiparametric regression models with censored data, *J. R. Stat. Soc. Ser. B* 69 (2007) 507–564.