Summer 8-31-2014

Confidence bands for survival curves using model assisted cox regression

Shoubhik Mondal
New Jersey Institute of Technology

ABSTRACT

CONFIDENCE BANDS FOR SURVIVAL CURVES USING MODEL ASSISTED COX REGRESSION

by
Shoubhik Mondal

The goal of this dissertation is to develop informative subject-specific simultaneous confidence bands (SCBs) for survival functions from right censored data. The approach is based on an extension of semiparametric random censorship models (SRCMs) to Cox regression, which produces reliable and more informative SCBs. SRCMs derive their rationale from their ability to utilize parametric ideas within the random censorship environment. Incorporating SRCMs into the existing framework produces more powerful procedures to analyze right censored data. The first part of the project focuses on proposing new estimators of Cox regression parameters and the cumulative baseline hazard function, and deriving their large sample properties. Under correct parametric specification, the proposed estimators 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. 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.

The second and final part of the project involves the deployment of the newly proposed estimators to obtain more informative SCBs for subject-specific survival curves. Simulation results are presented to compare the performance of the proposed SCBs with the SCBs that are based only on standard Cox. The new SCBs provide correct empirical coverage and are more informative. The proposed SCBs are illustrated with two real examples. An extension to handle missing censoring indicators is also outlined.

CONFIDENCE BANDS FOR SURVIVAL CURVES USING MODEL ASSISTED COX REGRESSION

by
Shoubhik Mondal

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences
Department of Mathematics and Computer Science, Rutgers-Newark

August 2014

APPROVAL PAGE

CONFIDENCE BANDS FOR SURVIVAL CURVES USING MODEL ASSISTED COX REGRESSION

Shoubhik Mondal

Dr. Sundarraman Subramanian, Dissertation Advisor — Date
Associate Professor of Statistics, New Jersey Institute of Technology

---

Dr. Sunil Dhar, Committee Member — Date
Professor of Statistics, New Jersey Institute of Technology

Dr. Wenge Guo, Committee Member — Date
Assistant Professor of Statistics, New Jersey Institute of Technology

---

Dr. Ji Meng Loh, Committee Member — Date
Associate Professor of Statistics, New Jersey Institute of Technology

---

Dr. Satrajit Roychowdhury, Committee Member — Date
Senior Associate Director, Novartis Pharmaceutical Company

BIOGRAPHICAL SKETCH

Author: Shoubhik Mondal

Degree: Doctor of Philosophy

Date: August 2014

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Master of Science in Statistics
  University of Calcutta, Kolkata, India, 2008
- Bachelor of Science in Statistics
  St. Xavier's College, Kolkata, India, 2006

Major: Probability and Applied Statistics

Presentations and Publications:

S. Mondal, S. Subramanian, "Model Assisted Cox Regression," *Journal of Multivariate Analysis*, 123, 281-303, 2014.

S. Mondal, S. Subramanian, "Simultaneous Confidence Bands for Cox Regression from Semiparametric Random Censorship," *S*ubmitted, 2014.

S. Mondal, "Model Assisted Cox Regression," *C*ontributed Talk, ENAR Spring Meeting, Baltimore, 2014.

S. Mondal, "Confidence Bands for Survival Functions in Cox Regression Framework," Statistics seminar series, New Jersey Institute of Technology, 2013.

S. Mondal, "Model Assisted Cox Regression," Poster presentation at annual graduate student research day, New Jersey Institute of Technology, 2013.

S. Mondal, "Model Assisted Cox Regression," Statistics seminar series, New Jersey Institute of Technology, 2012.

An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.

John Tukey

ACKNOWLEDGMENT

I am very grateful to my advisor, Dr. Sundarraman Subramanian, who guided me in the right direction not only as an academic advisor but also as a true well wisher and a friend. His constant encouragement and support helped me to strongly move past some of my most difficult times. Without his help it would not have been possible to complete this dissertation.

I would like to extend my gratitude to the other members of my dissertation committee, Dr. Sunil Dhar, Dr. Wenge Guo, Dr. Ji Meng Loh and Dr. Satrajit Roychowdhury, for their encouragement and support throughout my PhD.

I would also like to thank late Dr. Manish Bhattacharjee, Dr. Aridaman Jain, Dr. Daljit Singh Ahluwalia, Dr. Jonathan Luke, Dr. John K. Bechtold, and the other faculty and staff members of the Department of Mathematical Sciences for their support during the last 4 years.

I take this opportunity to mention the names of my parents, Mr. Joydev Mondal and late Mrs. Umarani Mondal, in an effort to convey my gratitude to them for their continuous guidance and sacrifice, which made me achieve this goal.

Last, but certainly not the least, I would like to thank my wife, Anjana, for her support, care, and guidance throughout my study.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

In biomedical applications it is common practice to report pointwise confidence intervals (PCIs) for survival curves to facilitate comparison of two treatments. PCIs over any desired region are easy to calculate and hence very attractive to practitioners. However, they can lead to incorrect judgment regarding treatment efficacy when treatments have time-varying effect, say, when a treatment may have high early survival but lower long-term survival. Consider the example of comparing allogenic bone marrow transplant (BMT) versus conventional chemotherapy (CC) for chronic myelogenous leukemia (CML), discussed in Zhang and Klein [1]. PCIs indicated that the threshold for BMT inefficacy was up to 4.5 years, after which period it appeared that the BMT group would have a lower mortality rate than the CC group. However, the simultaneous confidence bands (SCBs) that Zhang and Klein [1] presented indicated a more conservative threshold of about 6 years, implying that the BMT group had a higher mortality rate for up to 6 years. Although PCIs produce smaller widths than SCBs, which, when extended to a region, produce the perception of increased discriminative ability through smaller enclosed areas than SCBs, this may hardly be the case in reality. Specifically, PCIs guarantee correct coverage for each isolated point *separately* but not for a multitude of points *jointly*. In the case of the CML example referred above, the PCIs of the difference of the two survival curves computed at 4.5 years and 4.75 years would not cover the true difference at those years jointly with 95% confidence. The smaller enclosed area of a PCI over a region, therefore, may be an artifact of this limitation. SCBs would give a better picture of global variability than PCIs. In their analysis of the Mayo data base, Dickson et al. [2] make a strong case for computing subject-specific survival function estimates accompanied by SCBs. In this dissertation, we focus on constructing new and more informative SCBs for subject-specific survival functions in the Cox regression framework.

The Cox proportional hazards (PH) model, because of its simplicity, is widely used to investigate the effect of covariates on the survival time. For Cox PH regression, under the framework of the *random censorship model* (RCM), Burr and Doss [3], Lin, Fleming, and Wei [4], and Lai and Su [5] developed SCBs for subject-specific survival and quantile functions; Zhang and Klein [1] developed SCBs for the difference of two survival functions. Dabrowska and Ho [6] developed SCBs for the transition probabilities in a Markov chain model with intensities specified by the Cox PH model. Wei and Schaubel [7] constructed SCBs under nonproportional hazards to compare hemodialysis and peritoneal dialysis of end-stage renal disease patients, and debated the conclusions of previous studies. Gilbert et al. [8] analyzed data from a cholera vaccine study and developed SCBs for the log-hazard ratios of the cholera vaccine and placebo, using which they concluded that the vaccine lost its efficacy after 3 years and suggested that a more durable immune response needed to be developed in reformed cholera vaccines. Shen and Cheng [9], Yin and Hu [10], and Lee and Hyun [11] also developed SCBs, but under the additive risk model. In the case of homogeneous right censored data, SCBs for cumulative hazard, survival and quantile functions have been extensively investigated; see the recent paper of Subramanian and Zhang [12], who developed one sample model-based SCBs for survival curves, for a list of past work. They exploited the idea of *semiparametric random censorship model* (SRCM), introduced by Dikta [13], and developed a two stage bootstrap procedure to produce new SCBs which are more informative than the Hall Wellner(HW) and Nair's Equal Precision (EP) bands.

The SRCM framework to survival function estimation for a homogeneous population 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 [14] 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 SRCM 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 ([14]). Combining SRCMs with a standard Cox PH analysis is likely to produce improved parameter estimates and, in turn, more informative SCBs. This approach, not yet implemented in the literature, will be our focus.

This dissertation is organized as follows. In Chapter 2, we propose and implement our SRCM-based Cox PH analysis, involving a large sample study of our proposed estimators. The results obtained in this chapter are then utilized in Chapter 3 to produce a set of new SCBs for the subject specific survival function.

1.1 The Cox Proportional Hazard Regression

We begin with a brief review of Cox PH regression. When there are no MCIs, the standard set-up is the observation of n independent and identically distributed triplets $(X_i, \delta_i, \boldsymbol{Z}_i), i = 1, \ldots, n$, where $X = \min(T, C)$ is the minimum of the failure and censoring times, δ is the censoring indicator given by

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{if } T_i > C_i, \end{cases}$$

and $\boldsymbol{Z}$ denotes a $p \times 1$ vector of covariates. Under standard random censorship, T and C are conditionally independent given $\boldsymbol{Z}$. To analyze the influence of covariates on the survival time, Cox [15] proposed the PH model, where the conditional hazard function of the failure time given $\boldsymbol{Z}$ takes the form

$$\lambda(t|\boldsymbol{Z}=\boldsymbol{z}) \;=\; \lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{z}}. \tag{1.1}$$

Here, $\boldsymbol{\beta}$ is the $p \times 1$ regression parameter and $\lambda_0(t)$ is a baseline hazard function, which is independent of the covariates. Writing $N(t) = I(X \leq t)$, $Y_j(t) = I(X_j \geq t)$, $j = 1, \ldots, n$, it is well known ([16]) that the Cox partial likelihood estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, solves the equation $S_{n\mathrm{C}}(\boldsymbol{\beta}) = 0$, where

$$S_{n\mathrm{C}}(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} \int_0^{\infty} \delta_i \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j} \boldsymbol{Z}_j}{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right] dN_i(t). \tag{1.2}$$

Tsiatis [17] and Andersen and Gill [18] proved consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$ and the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{C}} - \boldsymbol{\beta}_0)$, with the covariance function Σ_0 given by Eq. (A.7) in the Appendix. Breslow's [19] estimator of the cumulative baseline hazard function is given by

$$\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\mathrm{C}}) \;=\; \sum_{i=1}^{n} \int_0^{t} \frac{\delta_i}{\sum_{j=1}^{n} Y_j(s) e^{\hat{\boldsymbol{\beta}}_{\mathrm{C}}^T \boldsymbol{Z}_j}} dN_i(s). \tag{1.3}$$

Andersen and Gill [18] also derived the weak convergence of $n^{1/2}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\mathrm{C}}) - \Lambda_0(t))$. The baseline survival function, denoted by $\hat{S}_0(t)$, is calculated from the relation

$$\hat{S}_0(t) = \prod_{i=1}^{n} \left[1 - \int_0^{t} \frac{\delta_i}{\sum_{j=1}^{n} Y_j(s) e^{\hat{\boldsymbol{\beta}}_{\mathrm{C}}^T \boldsymbol{Z}_j}} dN_i(s) \right]. \tag{1.4}$$

For a subject with given covariate value $\boldsymbol{Z} = \boldsymbol{z}_0$, the survival function is given by

$$\hat{S}(t; \boldsymbol{z}_0) = \{\hat{S}_0(t)\}^{\exp(\hat{\boldsymbol{\beta}}_{\mathrm{C}}^T \boldsymbol{z}_0)}. \tag{1.5}$$

The estimate on the left side of Eq. (1.5) provides the basic building block for constructing SCBs. In this dissertation, we show how to exploit SRCMs, to produce improved estimates that directly lead to more informative SCBs.

1.2 Semiparametric Random Censorship Models

The semiparametric random censorship framework for the homogeneous case, captures the dependency of censoring indicators on the observed time through a correctly specified model for $m(t)$. Incorporation of SRCMs into Cox PH regression involves replacing the censoring indicator with a model-based estimate of its conditional expectation given the observed time and covariates. A motivating example may highlight the need to explore the dependency of censoring indicators on the observed time and covariates as well. In the case of the well-known Stanford heart transplant data, most patients who were under observation for more than 500 days after heart transplant were censored. Furthermore, as would be perhaps expected, older patients (with age, say, forty plus) appeared more susceptible to the event (death). However, looking at the segment given in Table 1.1 below, this general trend appears to be violated by some stray data segments.

Table 1.1 Segment of Stanford Heart Transplant Data

Patient ID	*Observation time (days)*	*Status*	*Age*
171	231	0	52
173	188	0	52
178	107	0	46

The strength of SRCMs may perhaps lie in their ability to capture the trend exhibited by the bulk of the data. Specifically, by capturing the dependence of the censoring status on the observation time and age through a model, one may be able to provide a correct weight for the censoring status that is more in line with patient characteristics. For the proposed parametric assist, any binary regression model may be explored to arrive at a satisfactory specification. Popular choices such as logistic, Cauchy, probit, complementary log-log, and generalized PH models [13] may be explored and their adequacy can be checked using available formal goodness-of-fit tests and diagnostic checking procedures in statistical software (R, SAS, MINITAB etc.).

1.3 Simultaneous Confidence Bands

In this section, we provide a brief review of the various approaches that have been proposed in the literature to construct SCBs for survival functions in both the homogeneous as well as the subject-specific case.

For homogeneous right censored data, Hall-Wellner [20] and equal precision (EP) bands [21] are the most popular. Akritas [22] proposed the bootstrap to construct his SCBs. Subramanian and Zhang [12] provided an exhaustive list of references for research on SCBs for the homogeneous case. They proposed a two-stage bootstrap procedure based on the SRCMs to produce new SCBs which performed better than the HW and EP bands even in the presence of significant model misspecification.

Lin, Fleming, and Wei [4] constructed SCBs for subject-specific survival curves based on the Cox PH model. They used the Gaussian multiplier bootstrap (GMB) to obtain the thresholds required for SCB construction. To better explain this procedure, for given $\boldsymbol{Z} = \boldsymbol{z}_0$, let $\hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0)$ denote the standard Cox based subject-specific cumulative hazard estimate, and let $W(t;\boldsymbol{z}_0)$ denote the influence function of $n^{1/2}(\hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0) - \Lambda(t;\boldsymbol{z}_0))$. Then,

$$n^{1/2}(\hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0) - \Lambda(t;\boldsymbol{z}_0)) = n^{-1/2}\sum_{i=1}^{n} W_i(t;\boldsymbol{z}_0) + o_p(1) \ \equiv \mathbb{W}(t;\boldsymbol{z}_0) + o_p(1),$$

uniformly in t. For $i = 1, \ldots, n$, introducing the perturbed version $W_i^*(t;\boldsymbol{z}_0) = \hat{W}_i(t;\boldsymbol{z}_0)G_i$, where G_i are independent standard normal random variables and $\hat{W}_i$ is a consistent estimate of W_i, Lin et al. (1994) proved that the conditional distribution of $\mathbb{W}^*(t;\boldsymbol{z}_0) = n^{-1/2}\sum_{i=1}^{n} W_i^*(t;\boldsymbol{z}_0)$ given the observed data has the same weak limit as $W(t;\boldsymbol{z}_0)$. Applying the continuous mapping theorem, the upper α quantile of $n^{1/2}\|(\hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0) - \Lambda(t;\boldsymbol{z}_0))\|_{t_1}^{t_2}$ can be approximated by that of $\|W^*(t;\boldsymbol{z}_0)\|_{t_1}^{t_2}$, where $\|\cdot\|$ is the supnorm. A simple, linear, $100(1-\alpha)\%$ SCB for $\Lambda(t;\boldsymbol{z}_0)$ is given by $\hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0) - n^{-1/2}q_\alpha, \hat{\Lambda}_{\mathrm{C}}(t;\boldsymbol{z}_0) + n^{-1/2}q_\alpha$. Lin et al. [4] computed more elaborate SCBs by considering

$$B(t;\boldsymbol{z}_0) = n^{1/2} g(t;\boldsymbol{z}_0)[\phi(\hat{\Lambda}(t;\boldsymbol{z}_0)) - \phi(\Lambda(t;\boldsymbol{z}_0))],$$

where ϕ is known function whose derivative ϕ' is continuous and nonzero in the time interval $[(t_1, t_2) : 0 \le t_1 \le t_2 \le \tau]$ and the weight function $g(t;\boldsymbol{z}_0)$ converges to a nonnegative bounded function uniformly on $[t_1, t_2]$. Typically, g is taken as the reciprocal of the standard error of $\phi(\hat{\Lambda})$. The functional delta method gives an alternative expression for $B(t;\boldsymbol{z}_0)$,

$$\hat{B}(t;\boldsymbol{z}_0) = g(t;\boldsymbol{z}_0)\phi'(\hat{\Lambda}(t;\boldsymbol{z}_0))\mathbb{W}^*(t;\boldsymbol{z}_0).$$

Then, one can determine the desired threshold by simply extracting q_α, the upper-α quantile of the distribution of $\|\hat{B}(t;\boldsymbol{z}_0)\|_{t_1}^{t_2}$. For example, taking $\phi(x) = \log x$ and $g(t;\boldsymbol{z}_0) = \hat{\Lambda}(t;\boldsymbol{z}_0)/\hat{\sigma}(t;\boldsymbol{z}_0)$, a $100(1-\alpha)\%$ SCB for $\log \hat{\Lambda}$ over $[t_1, t_2]$ is given by

$$\log \hat{\Lambda}(t;\boldsymbol{z}_0) \mp n^{-1/2} q_\alpha \frac{\hat{\sigma}(t;\boldsymbol{z}_0)}{\hat{\Lambda}(t;\boldsymbol{z}_0)}.$$

Since $\log \Lambda = \log(-\log S)$, one obtains

$$\hat{S}(t;\boldsymbol{z}_0)^{\exp(\pm n^{-1/2} q_\alpha \hat{\sigma}(t;\boldsymbol{z}_0)/\hat{\Lambda}(t;\boldsymbol{z}_0))}. \tag{1.6}$$

As will be shown in Chapter 3, our proposed SCBs provide superior performance than the ones given by Eq. (1.6).

CHAPTER 2

MODEL ASSISTED COX REGRESSION

2.1 Introduction

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 afore-mentioned conditional expectation, which, in addition to its dependence on the observed event time, may now also depend on a set of covariates $\boldsymbol{Z}$. 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 $\boldsymbol{Z}$, $m(x, \boldsymbol{z}) = P(\delta = 1|X = x, \boldsymbol{Z} = \boldsymbol{z})$ is the ratio of the conditional event-time hazard to the conditional *total* hazard [13, 23]. 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 $\boldsymbol{Z}$. Standard Cox analysis ignores this relationship by leaving the conditional censoring 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 3.3, 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 [23] extended the Koziol–Green model [24] to the subject-specific setting implicit in Cox regression, which subsumes an earlier approach [25] as well. In terms of finite sample performance, both the proposed and Yuan [23] estimators perform equally well, see Section 3.3. Our proposed method, however, offers a more attractive alternative for the following reasons. Yuan [23] 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 $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, compromising to some extent the simplicity of standard Cox regression analysis. Indeed, for a logistic model Yuan's [23] 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 $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}(t)$, the proposed estimators of $\boldsymbol{\beta}$ and $\Lambda_0(t)$, respectively, are each asymptotically as or more efficient than the standard Cox regression estimators.

Liu and Wang (2010) proposed two estimators of $\boldsymbol{\beta}$ to account for the MCIs. They only focused on estimation of $\boldsymbol{\beta}$, which is a limitation. Their first estimator denoted $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$, 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 3.3 reveal that $\hat{\boldsymbol{\beta}}$ performs as well as or better than $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$.

This chapter is further organized as follows. In Section 2.2, we present our proposed estimators and provide theoretical comparisons with standard Cox PH regression estimators. We then present our proposed extension when there are MCIs. In Section 3.3, we present the results of simulation studies comparing the proposed and other approaches under discussion. We also provide two illustrations using data from a heart transplant study and a study on recidivism, and an additional illustration using pseudo-real data. Technical details are given in the appendix A.

2.2 Estimators and Large Sample Results

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

$$S_{\mathrm{C}}(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} \int_0^{\infty} \delta_i \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j} \boldsymbol{Z}_j}{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right] dN_i(t). \tag{2.1}$$

Breslow's [19] estimator of the baseline cumulative hazard function is given by

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

Andersen and Gill [18] proved that $\hat{\boldsymbol{\beta}}_{\mathrm{C}} \xrightarrow{P} \boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{C}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{C}}^{-1})$, where Eq. (A.7) defines $\boldsymbol{\Sigma}_{\mathrm{C}}$. They also derived the weak convergence of $n^{1/2}(\hat{\Lambda}_{0\mathrm{C}}(t) - \Lambda_0(t))$, with the limiting variance function given by the first two terms of Eq. (A.48); see also, Tsiatis [17].

2.2.1 Censoring Indicators Always Observed

To tie the SRC framework to Cox regression, we write $m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) = P(\delta = 1|X, \boldsymbol{Z})$. The unknown $\boldsymbol{\theta} \in \mathbb{R}^k$, whose true value is $\boldsymbol{\theta}_0$, can be estimated by maximizing the quantity

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

Let $\hat{\boldsymbol{\theta}}$ denote the maximizer of Eq. (3.1). Then $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ [26]. Our proposed estimator $\hat{\boldsymbol{\beta}}$ is obtained by solving the equation $S_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = 0$, where

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

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

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

Our estimator of the baseline cumulative hazard function is given by

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

Theorem 1 and proposition 1 below give the large sample results of our proposed estimators.

Theorem 1. *When the parametric model for $m(x, \boldsymbol{z})$ is correctly specified and when conditions **A.1**–**A.6** and **D.1** hold, (i) $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, as $n \to \infty$, where $\boldsymbol{\Sigma}$ is given by Eq. (A.12); and (ii) $n^{1/2}(\hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(\cdot)) \xrightarrow{\mathcal{D}} \mathbb{U}(\cdot)$ in $D[0, \tau]$, where $\mathbb{U}$ is a zero-mean Gaussian process with covariance function $\sigma(\cdot, \cdot)$ given by Eq. (A.19).*

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

Proof Write $\langle \cdot, \cdot \rangle$ for the inner product in Euclidean space. Note that $\boldsymbol{B_0}$ is given by Eq. (A.11) and $D_r(m(x, \boldsymbol{z}, \boldsymbol{\theta}_0)) = \partial m(x, \boldsymbol{z}, \boldsymbol{\theta}) / \partial \theta_r |_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}, r = 1, \ldots, k$. From Eq. (A.7) and Eq. (A.36), we get the expression for $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}$ as

$$\boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \left[\boldsymbol{B_0} \boldsymbol{I_0}^{-1} \boldsymbol{B_0}^T - E \left\{ m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) \bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) (\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau) \right\} \right] \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}.$$

Applying Loewner's ordering [27] it suffices to prove for any $\boldsymbol{a} = (a_1, ..., a_p)^T$ that

$$\boldsymbol{a}^T \left[\boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{B_0}^T - E\left\{m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right\}\right] \boldsymbol{a} \leq 0. \tag{2.7}$$

We follow Dikta [13], who uses the following result in Rao [28]: For any $\boldsymbol{b} \in I\!R^k$,

$$\sup_{\boldsymbol{h} \in I\!R^k - \{0\}} \frac{\langle \boldsymbol{h}, \boldsymbol{b} \rangle^2}{\langle \boldsymbol{h}, \boldsymbol{I_0}\boldsymbol{h} \rangle} = \langle \boldsymbol{b}, \boldsymbol{I_0}^{-1}\boldsymbol{b} \rangle. \tag{2.8}$$

Note that $\boldsymbol{b} = (b_1, \ldots, b_k)^T = \boldsymbol{B_0}^T\boldsymbol{a}$ links the first term of (2.7) with the right hand side of Eq. (2.8): $\langle \boldsymbol{b}, \boldsymbol{I_0}^{-1}\boldsymbol{b} \rangle = \boldsymbol{a}^T\boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{B_0}^T\boldsymbol{a}$. It then suffices to show that for all $\boldsymbol{h} \in I\!R^k - \{0\}$

$$\frac{\langle \boldsymbol{h}, \boldsymbol{b} \rangle^2}{\langle \boldsymbol{h}, \boldsymbol{I_0}\boldsymbol{h} \rangle} \leq \boldsymbol{a}^T E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right] \boldsymbol{a}. \tag{2.9}$$

Let $G(x, \boldsymbol{z})$ denote the joint cumulative distribution function of $(X, \boldsymbol{Z})$ and $\Gamma_t = [0, t] \times I\!R^p$. Let $G^1(t, \boldsymbol{z}) = \int_{\Gamma_t} m(u, \boldsymbol{z}, \boldsymbol{\theta}_0)\bar{m}(u, \boldsymbol{z}, \boldsymbol{\theta}_0))d\,G(u, \boldsymbol{z})$. For fixed $\boldsymbol{h} \in I\!R^k - \{0\}$,

$$\begin{aligned}
\langle \boldsymbol{h}, \boldsymbol{b} \rangle^2 &= \left[E\left(\sum_{i=1}^{k} h_i D_i(m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^T I(X \leq \tau)\right)\boldsymbol{a}\right]^2 \\
&= \left[\int_{\Gamma_\tau} \sum_{i=1}^{k} \frac{h_i D_i(m(t, \boldsymbol{z}, \boldsymbol{\theta}_0))(\boldsymbol{z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^T}{m(t, \boldsymbol{z}, \boldsymbol{\theta}_0)\bar{m}(t, \boldsymbol{z}, \boldsymbol{\theta}_0)} d\,G^1(t, \boldsymbol{z})\boldsymbol{a}\right]^2.
\end{aligned}$$

Applying the Cauchy–Schwarz inequality,

$$\begin{aligned}
\langle \boldsymbol{h}, \boldsymbol{b} \rangle^2 \leq &\left[\int_{\Gamma_\tau} \left(\boldsymbol{a}^T(\boldsymbol{z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2}\boldsymbol{a}\right) d\,G^1(t, \boldsymbol{z}) \right. \\
&\left. \times \int_{\Gamma_\tau} \left(\sum_{i=1}^{k} \frac{h_i D_i(m(t, \boldsymbol{z}, \boldsymbol{\theta}_0))}{m(t, \boldsymbol{z}, \boldsymbol{\theta}_0)\bar{m}(t, \boldsymbol{z}, \boldsymbol{\theta}_0)}\right)^2 d\,G^1(t, \boldsymbol{z})\right].
\end{aligned}$$

The second integral on the right hand side above equals $\langle \boldsymbol{h}, I(\boldsymbol{\theta}_0)\boldsymbol{h} \rangle$. The first integral is just the right hand side of inequality (2.9), proving the asymptotic efficiency of $\hat{\boldsymbol{\beta}}$.

The efficiency of $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ is proved likewise. Note that the sum of the first two terms on the right side of Eq. (A.48) gives the variance of $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\mathrm{C}})$. It remains to prove that

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

where $\boldsymbol{d}_0(t)$ and $\alpha(t, X, Z, \boldsymbol{\beta}_0)$ are given by Eq. (A.16) and Eq. (A.17), respectively. Proceeding as before, with $\boldsymbol{b} = \boldsymbol{d}_0(t)$, it suffices to show that for all $\boldsymbol{h} = (h_1, ..., h_k) \in \mathbb{R}^k - \{0\}$,

$$\frac{\langle \boldsymbol{h}, \boldsymbol{b}\rangle^2}{\langle \boldsymbol{h}, I(\boldsymbol{\theta}_0)(\boldsymbol{h})\rangle} \leq E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\left(\alpha(t, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)^2 I(X \leq \tau)\right]. \quad (2.10)$$

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

$$\begin{aligned}
\langle \boldsymbol{h}, \boldsymbol{b}\rangle^2 &= \left[E\left\{\sum_{i=1}^k h_i D_i(m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))\alpha(t, X, \boldsymbol{Z}, \boldsymbol{\beta}_0) I(X \leq \tau)\right\}\right]^2 \\
&= \left[\int_{\Gamma_\tau}\left(\sum_{i=1}^k \frac{h_i D_i(m(s, \boldsymbol{z}, \boldsymbol{\theta}_0))}{m(s, \boldsymbol{z}, \boldsymbol{\theta}_0)\bar{m}(s, \boldsymbol{z}, \boldsymbol{\theta}_0)}\right)\alpha(t, s, \boldsymbol{z}, \boldsymbol{\beta}_0) d\, G^1(s, \boldsymbol{z})\right]^2 \\
&\leq \int_{\Gamma_\tau}\left(\alpha(t, s, \boldsymbol{z}, \boldsymbol{\beta}_0)\right)^2 d\, G^1(s, \boldsymbol{z}) \\
&\quad \times \int_{\Gamma_\tau}\left(\sum_{i=1}^k \frac{h_i D_i(m(s, \boldsymbol{z}, \boldsymbol{\theta}_0))}{m(s, \boldsymbol{z}, \boldsymbol{\theta}_0)\bar{m}(s, \boldsymbol{z}, \boldsymbol{\theta}_0)}\right)^2 d\, G^1(s, \boldsymbol{z}).
\end{aligned}$$

The first integral above is readily seen to be the right side of inequality (2.10). The second integral above is just $\langle \boldsymbol{h}, I(\boldsymbol{\theta}_0)\boldsymbol{h}\rangle$. This completes the proof.

2.2.2 Censoring Indicators Missing at Random

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

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

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

Theorem 2. *When the model for $m(x,\boldsymbol{z})$ is correctly specified, when the MCIs are MAR, and conditions **A.1**–**A.6** and **D.1** hold, then (i) $\hat{\boldsymbol{\beta}}\xrightarrow{P}\boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{\mathcal{D}}\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}_{\mathrm{M}})$, as $n\to\infty$, where $\boldsymbol{\Sigma}_{\mathrm{M}}$ is given by Eq. (A.13); (ii) $n^{1/2}(\hat{\Lambda}_0(\cdot,\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})-\Lambda_0(\cdot))\xrightarrow{\mathcal{D}}\mathbb{V}(\cdot)$ in $D[0,\tau]$, where $\mathbb{V}$ is a zero-mean Gaussian process with covariance function $\sigma_{\mathrm{M}}(\cdot,\cdot)$ given by Eq. (A.21).*

Remark When $\pi\equiv 1$ (no MCIs), then $\tilde{\boldsymbol{I}}_{\mathbf{0}}=\boldsymbol{I}_{\mathbf{0}}$ and $\boldsymbol{\Sigma}_{\mathrm{M}}$ reduces to $\boldsymbol{\Sigma}$; furthermore, $\sigma_{\mathrm{M}}(t_1,t_2)$ reduces to $\sigma(t_1,t_2)$.

2.3 Simulation 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 $\hat{\boldsymbol{\beta}}$, 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 $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, 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 Yuan [23] and Liu and Wang [29] are denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{Y}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$, respectively. Note that $\hat{\boldsymbol{\beta}}_{\mathrm{Y}}$ applies only in the absence of MCIs and comparison with $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$ applies only when there are MCIs.

2.3.1 Censoring Indicators Always Observed

We provide comparisons between the proposed, standard Cox, and Yuan [23] estimators.

No Misspecification The covariate was one dimensional and uniformly distributed over $[1, 2]$. The conditional event-time and censoring hazards given Z were each with rate $\exp(Z)$ 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 $m(X, Z, \theta) = 1/(1 + exp(-\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1 Z))$. 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 Figure 3.1.

Our proposed estimator $\hat{\boldsymbol{\beta}}$ outperforms $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$ significantly, while being comparable to $\hat{\boldsymbol{\beta}}_Y$. The relative improvement $[100 \times (Cox - proposed)/Cox]$ over standard Cox varied from 4% to 15% in terms of MSE. For baseline cumulative hazard, the relative improvement of $\hat{\Lambda}$ over the Breslow estimator in terms of MISE varied from 3% to 9%. The proposed estimator performed better than Yuan's [23] 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 [23] estimators over standard Cox varied between 10% and 20%.

Asymptotic confidence intervals based on $\hat{\boldsymbol{\beta}}_{\mathrm{Y}}$ require the bootstrap, see Yuan [23]. In Figure 2.2 the ECP and EML of the 95% asymptotic confidence intervals for $\boldsymbol{\beta}$, based on the large sample distributions of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, 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.

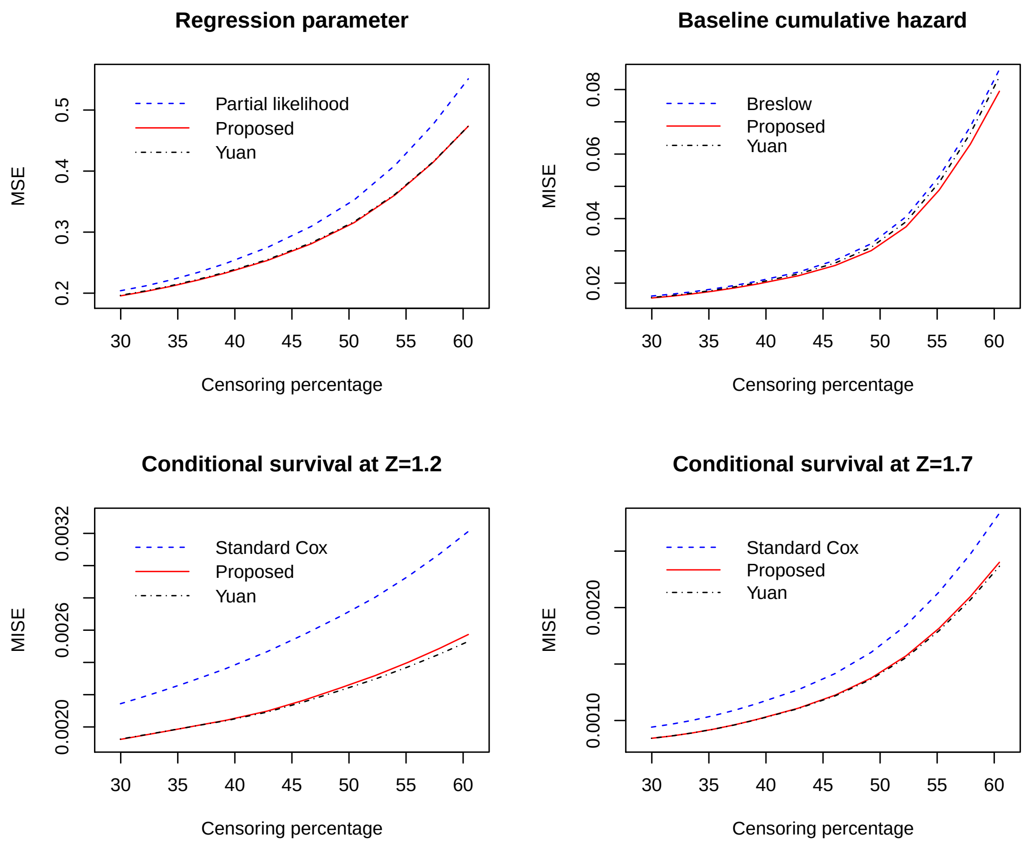

Figure 2.1 Comparison of estimators when m is correctly specified. The mean squared error (MSE) and the mean integrated squared error (MISE) are plotted against censoring rate.

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 Figure 3.2.

The proposed estimator $\hat{\boldsymbol{\beta}}$ outperformed $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, 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

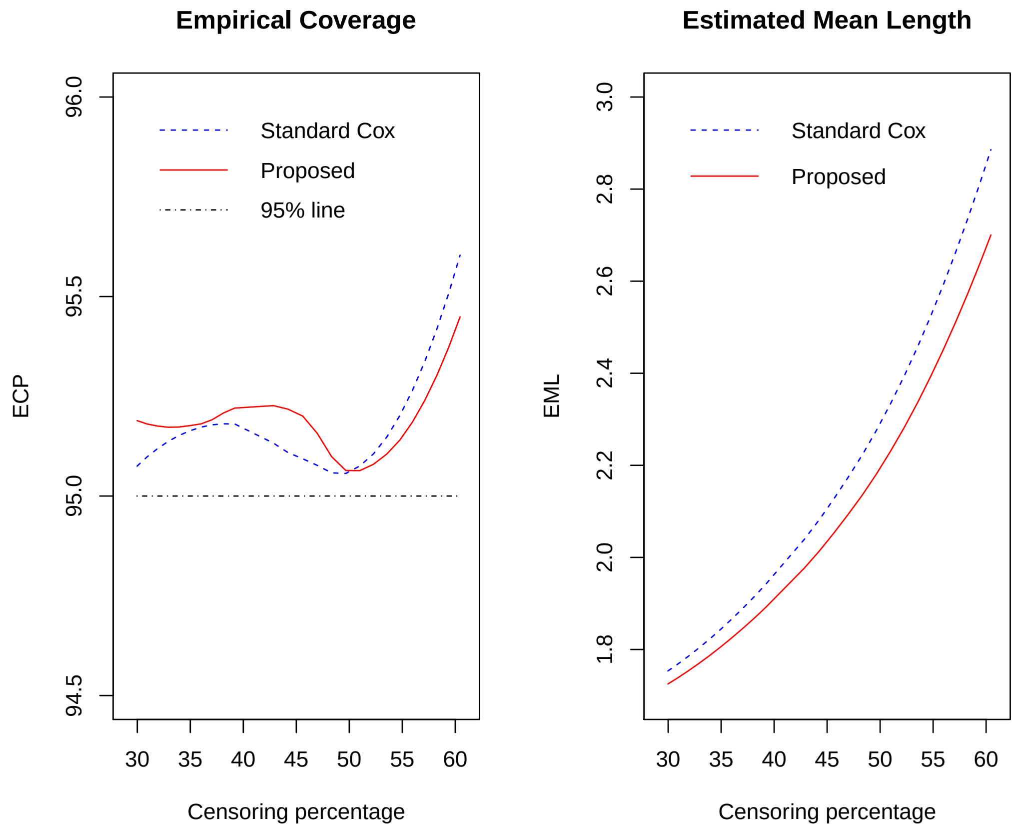

Figure 2.2 Empirical coverage probability (ECP) and estimated average length (EML) of 95% confidence intervals are plotted for various censoring rates when m is correctly specified.

survival with two covariate levels $Z = 1.2$ and $Z = 1.7$, the relative improvement of the proposed and Yuan [23] estimators over standard Cox varied between 2% and 45%. Thus, even under significant model misspecification, our proposed estimators outperformed the standard Cox estimators.

In Figure 2.4, the ECP and EML of the 95% asymptotic confidence intervals for $\boldsymbol{\beta}$, based on $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, 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.

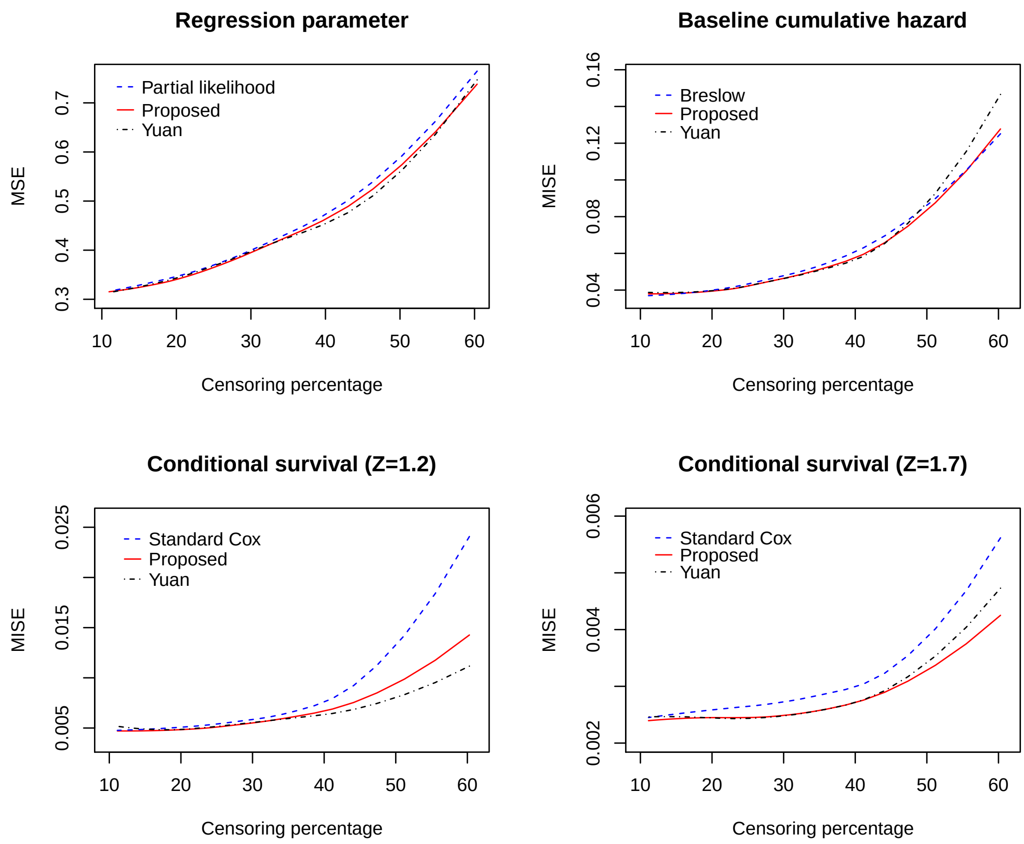

Figure 2.3 Comparison of estimators when m is misspecified. The mean squared error (MSE) and the mean integrated squared error (MISE) are plotted against censoring rate.

2.3.2 Censoring Indicators Missing at Random

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

Simulation Study in the Absence and Presence of Misspecification Since Liu and Wang [29] did not investigate estimation of Λ_0, here we only provide comparisons between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$. Data were simulated according to the designs in Subsection 2.3.1 (no misspecification) and Subsection 2.3.1 (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

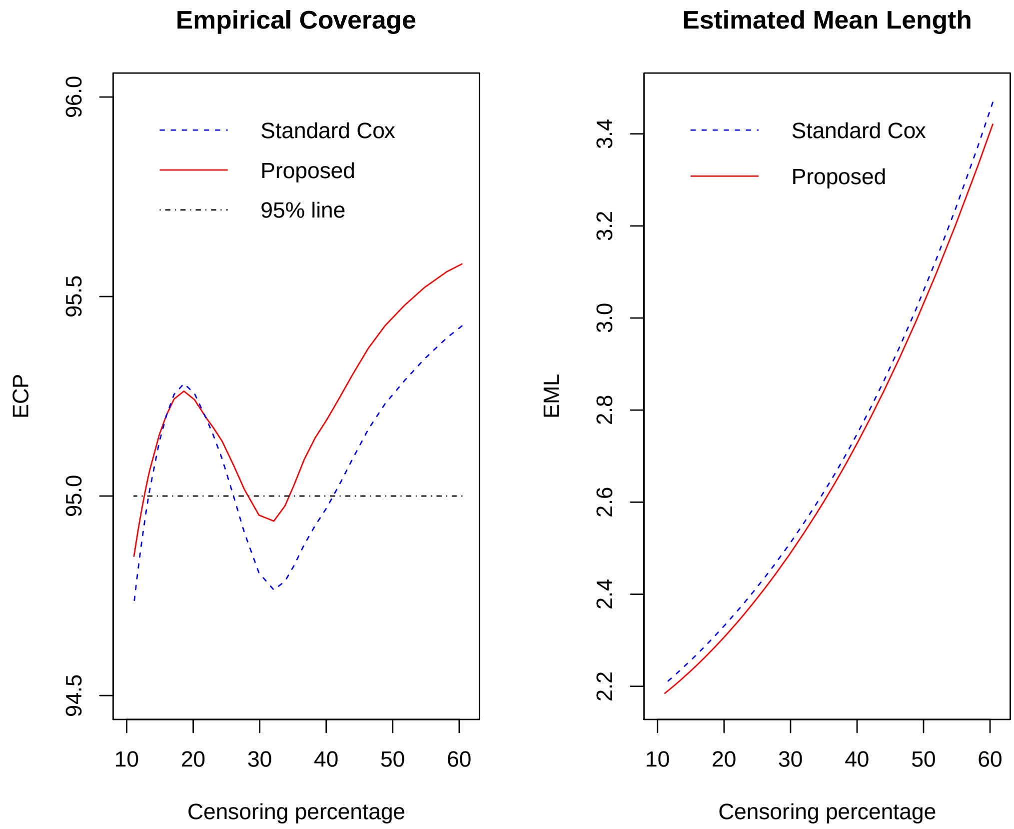

Figure 2.4 Empirical coverage probability (ECP) and estimated average length (EML) of 95% confidence intervals plotted for various censoring rates when the model for m is misspecified.

10,000 replications, and the sample size was 100. Results are shown in Figure 2.5. When m was correctly specified, $\hat{\boldsymbol{\beta}}$ offered a reduction of up to 8% in MSE over $\hat{\boldsymbol{\beta}}_W$. When m was misspecified, $\hat{\boldsymbol{\beta}}$ offered a reduction of up to 7% over $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$ when the MR was 20% and up to 1.14% for 44% MR.

2.4 Real Examples

2.4.1 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 $\boldsymbol{Z}$ =

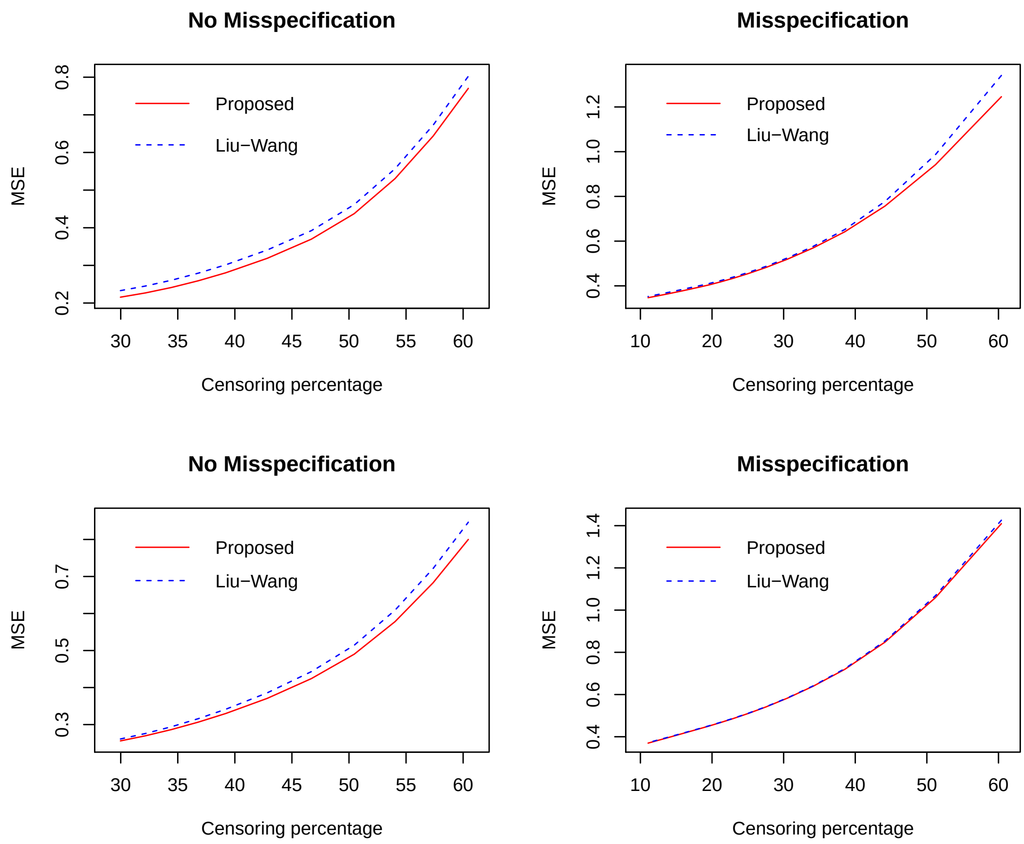

Figure 2.5 Comparison of estimators when censoring indicators are missing at random.

$(Z_1, Z_2)^T$, where $Z_1 = \text{age} - 41.7$ and $Z_2 = Z_1^2$ has been reported to fit the SHT data well (e.g. Escobar and Meeker [30]). We employed a logistic model for m with covariates X (time) and Z_1 that were selected by a stepwise regression procedure: $\text{logit}(m(X, Z, \boldsymbol{\theta})) = \theta_0 + \theta_1 X + \theta_2 Z_1$. The results are presented in Table 2.1. 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 Figure 2.6 (along with two other curves, see Section 2.4.3). 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

Table 2.1 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

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.

2.4.2 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 (Rossi, Berk and Lenihan [31]). 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 [32] 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

$$\text{logit}(m(t, \boldsymbol{Z}, \boldsymbol{\theta})) \quad = \quad \theta_1 \, \text{age} + \theta_2 \, \text{prior}. \tag{2.12}$$

Among several choices investigated, Eq. (2.12) was optimal in terms of goodness-of-fit.

In Table 2.2, 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. (2.12). Table 2.3 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.*

Table 2.2 Standard Cox and Proposed Estimates after First-Stage Analysis of Recidivism Data

	Standard Cox			*Proposed*		
Variable	*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

Table 2.3 Standard Cox and Proposed Estimates after Second-Stage Analysis of Recidivism Data

	Standard Cox			*Proposed*		
Variable	*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

2.4.3 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 2.4, are quite close to the complete data estimates (cf. Table 2.1), 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 Figure 2.6. 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 2.4 Estimates of Standford 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

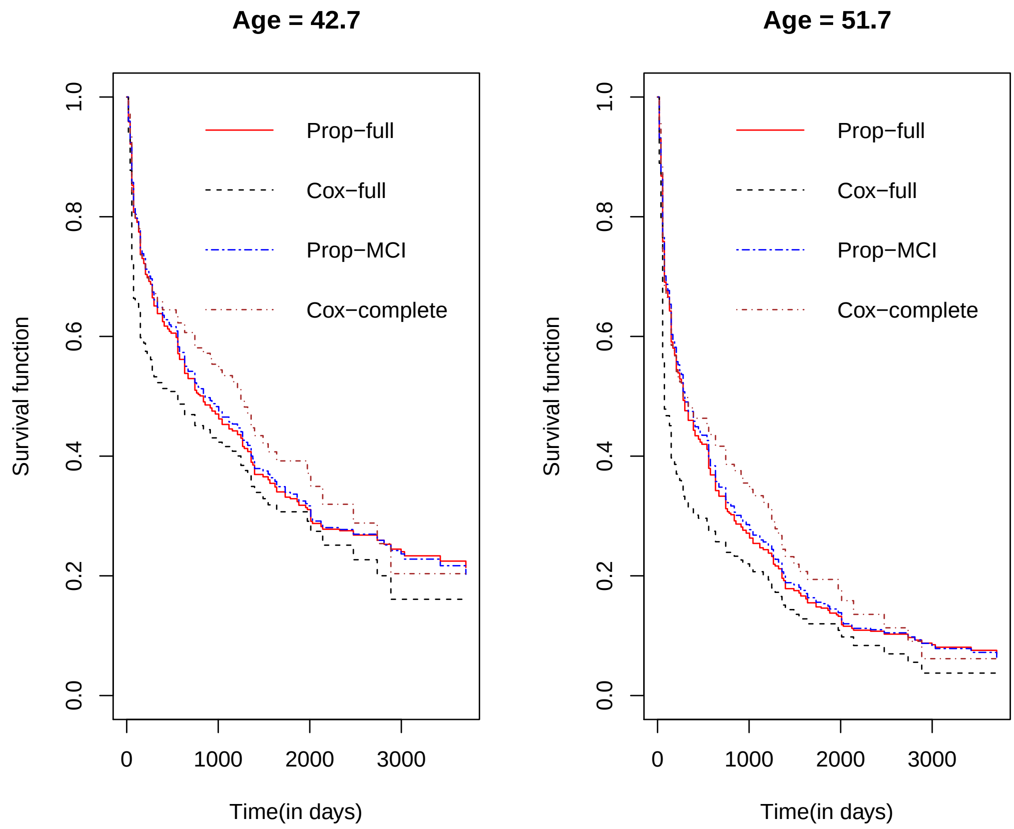

Figure 2.6 Subject specific survival function when censoring indicators are missing at random.

2.5 Conclusion

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 Yuan [23], whose method, however, can not be applied when the censoring indicators are missing at random. Under MCIs, our method performed better than Liu and Wang's [29] 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. 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 Collett [33]. In Section 3.3, 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.

CHAPTER 3

NEW CONFIDENCE BANDS

3.1 Introduction

In this chapter, we develop new subject-specific SCBs for survival curves from standard Cox PH regression assisted by *semiparametric random censorship models* proposed in the first chapter. Two-sample SCBs when the group-specific hazards are proportional are a special case, where the single covariate is the group indicator. The SRCM-based survival function estimator for the homogeneous case is semiparametric efficient [34]. SRCMs-assisted Cox PH regression produces more efficient estimators of the regression parameter vector and baseline cumulative hazard function, can improve upon marginal decisions produced by a standard Cox analysis, and is able to handle missing censoring indicators (MCIs) without undue extra computational effort. Subject-specific survival functions, which form the basic building block for subject-specific SCBs, have not been studied for the model-based scenario, however.

This approach leads to new estimators of $\boldsymbol{\beta}$, the regression parameter, and $\Lambda_0(t)$, the baseline cumulative hazard, which are asymptotically as or more efficient than their standard Cox counterparts. Plugging in the new estimators into a standard representation [35] leads to $\hat{\Lambda}(t, \boldsymbol{z})$, our proposed estimator of the subject-specific cumulative hazard function. For subject with covariate $\boldsymbol{z}_0$, we derive an asymptotic representation of the process $\hat{\mathbb{H}}(\cdot) = n^{1/2}(\hat{\Lambda}(\cdot, \boldsymbol{z}_0) - \Lambda(\cdot, \boldsymbol{z}_0))$, from which the weak convergence of $\hat{\mathbb{H}}$ to $\mathbb{H}$, a zero-mean Gaussian process, can be deduced. The method of SCB construction relies on the capability to obtain the upper-α quantile of the distribution of $||\hat{\mathbb{H}}||_{t_1}^{t_2}$, the supremum of $\hat{\mathbb{H}}$ over $[t_1, t_2] \subset [0, \tau_0)$, where τ_0 is the right endpoint of the support of the distribution of X, see Section 3.2. Unlike for the homogeneous case, a standard Brownian bridge approximation to its distribution, however, is not a viable option, since $\mathbb{H}$ does not have independent increments. Lin, Fleming, and Wei [4] faced a similar problem and employed the Gaussian multiplier bootstrap [36], henceforth indicated as GMB.

We apply the GMB to the derived asymptotic representation of $\mathbb{H}$ to obtain thresholds needed for constructing our proposed equal-precision (EP) and Hall–Wellner (HW) type SCBs. Simulation and sensitivity studies presented in Section 3.3 show that our proposed SCBs provide approximately correct coverage. For censoring rates (CRs) between 20% and 50%, the proposed SCBs gave a percent relative reduction in estimated average enclosed areas (EAEAs) and estimated average widths (EAWs) amounting to between 5% and 15% over their only competitor, namely, the standard Cox based SCBs developed by Lin, Fleming and Wei [36]. We also provide an extension that handles MCIs.

This chapter is organized as follows. In Section 3.2, we present our proposed bands and the asymptotic results. In Section 3.3, we present a set of simulation studies to showcase the performance of our proposed bands. Next we provide two illustrations using real datasets. Technical details, such as the asymptotic validity of the GMB, are given in the appendix B.

3.2 Simultaneous Confidence Bands

Recall that for the Cox PH model, the conditional hazard function of the failure time given the vector $\boldsymbol{Z}$ takes the form $\lambda(t|\boldsymbol{Z}=\boldsymbol{z})=\lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{z}}$, where $\boldsymbol{\beta}$ is a $p\times 1$ regression parameter and $\lambda_0(t)$ is an unspecified baseline hazard function. Under the RCM setup, the observed data constitute n independent and identically distributed triplets $(X_i,\delta_i,\boldsymbol{Z}_i), i=1,\ldots,n$, 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 $\boldsymbol{Z}$ is the $p\times 1$ vector of covariates. We assume that T and C are conditionally independent given $\boldsymbol{Z}$. When there are MCIs, we introduce a missingness indicator ξ (1 when δ is observed and 0 when missing), with the observed data then being the n independent and identically distributed quintuplets $(X_i,\xi_i,\sigma_i,\boldsymbol{Z}_i), i=1,\ldots,n,$

where $\sigma = \xi\delta$. SRCMs attach a model for $m(x, \boldsymbol{z})$. Specifically, $m(x, \boldsymbol{z}, \boldsymbol{\theta}_0) = P(\delta = 1|X = x, \boldsymbol{Z} = \boldsymbol{z})$, where $m(x, \boldsymbol{z}, \boldsymbol{\theta})$ has a known form with unknown $\boldsymbol{\theta} \in I\!R^k$, the true value of which is $\boldsymbol{\theta}_0$. The likelihood for $\boldsymbol{\theta}$, given by,

$$\prod_{i=1}^{n} \{m(X_i, \boldsymbol{Z}_i, \boldsymbol{\theta})\}^{\sigma_i} \{1 - m(X_i, \boldsymbol{Z}_i, \boldsymbol{\theta})\}^{\xi_i - \sigma_i}, \tag{3.1}$$

reduces to the standard one when there are no MCIs, see Sections 2.2.1 and 2.2.2. Let $\hat{\boldsymbol{\theta}}$ denote the maximizer of (3.1) and write $N_i(t) = I(X_i \leq t)$, $Y_i(t) = I(X_i \geq t)$. Then $\hat{\boldsymbol{\beta}}$, the SRCM-adjusted estimator of $\boldsymbol{\beta}$, solves the equation $S_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = 0$, where

$$S_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \int_0^{\tau_0} m(t, \boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}) \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j} \boldsymbol{Z}_j}{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right] dN_i(t). \tag{3.2}$$

Under appropriate regularity conditions, in Theorem 1 we proved the existence of $\hat{\boldsymbol{\beta}}$ that converges to $\boldsymbol{\beta}_0$ in probability as $n \to \infty$, and that the asymptotic variance of $\boldsymbol{a}^T \hat{\boldsymbol{\beta}}$ is no greater than its RCM counterpart. We also proposed the alternate Breslow-type estimator of the baseline cumulative hazard function, given by

$$\hat{\Lambda}_0(t) \;\; = \;\; \sum_{i=1}^{n} \int_0^t \frac{m(s, \boldsymbol{Z}_i, \hat{\boldsymbol{\theta}})}{\sum_{j=1}^{n} Y_j(s) e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_j}} dN_i(s), \tag{3.3}$$

and proved the weak convergence and asymptotic efficiency of its normalized version.

Let $\Lambda(t, \boldsymbol{z}_0)$ denote the conditional cumulative hazard of T given $\boldsymbol{Z} = \boldsymbol{z}_0$. The new SCBs for subject-specific survival functions will be based on $\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{z}_0)$, the estimate of $\Lambda(t, \boldsymbol{z}_0)$:

$$\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{z}_0) \;\equiv\; \hat{\Lambda}(t, \boldsymbol{z}_0) \;\; = \;\; \sum_{i=1}^{n} \int_0^t \frac{m(u, \boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}) \exp(\hat{\boldsymbol{\beta}}^T \boldsymbol{z}_0)}{\sum_{j=1}^{n} Y_j(u) e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_j}} dN_i(u). \tag{3.4}$$

Let M denote the martingale associated with the counting process N, see Eq. (A.3). In Appendix B.1, we show that $\hat{\mathbb{H}}(t) = n^{1/2}(\hat{\Lambda}(t, \boldsymbol{z}_0) - \Lambda(t, \boldsymbol{z}_0))$ admits the asymptotic representation $\hat{\mathbb{H}}(t) = L_{n,1}(t) + L_{n,2}(t) + o_p(1)$, uniformly for $t \in [0, \tau_0)$,

$$L_{n,1}(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau_0} m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0) a_i(t, u) dM_i(u) \tag{3.5}$$

$$L_{n,2}(t) = n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i - m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)}{m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i, \boldsymbol{\theta}_0)} \boldsymbol{b}^T(t) \boldsymbol{I_0}^{-1} \mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)), \tag{3.6}$$

$\boldsymbol{W} = (X, \boldsymbol{Z}^T)^T$, $\bar{m}(\cdot, \boldsymbol{\theta}) = 1 - m(\cdot, \boldsymbol{\theta})$, and $\boldsymbol{I}_0$, $a(t, u)$, and $\boldsymbol{b}(t)$ are given by Eqs. (A.1), (B.7), and (B.8), respectively. The vector $\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{w}, \boldsymbol{\theta}))$ is defined in Appendix B.1. From the limiting covariance function, see Eq. (B.21), we arrive at the variance function given by

$$V(t) = C(t, t) = E\left[m^2(\boldsymbol{W}, \boldsymbol{\theta}_0) a^2(t, X) I(X < \tau_0)\right] + \boldsymbol{b}^T(t) \boldsymbol{I}_0^{-1} \boldsymbol{b}(t). \tag{3.7}$$

Using Eq. (B.22) we now show that the new estimator $\hat{\Lambda}(t, \boldsymbol{z}_0)$ is asymptotically as or more efficient than $\hat{\Lambda}_{\mathrm{C}}(t, \boldsymbol{z}_0)$, the standard Cox based conditional cumulative hazard estimator.

Proposition 2. *When the parametric model for $m(t, \boldsymbol{z})$ is correctly specified, the estimator $\hat{\Lambda}(t, \boldsymbol{z}_0)$ is asymptotically as or more efficient than $\hat{\Lambda}_{\mathrm{C}}(t, \boldsymbol{z}_0)$.*

Proof To prove efficiency, an alternate form of $V(t)$, that also incorporates the asymptotic variance of the standard Cox based conditional cumulative hazard estimator given $\boldsymbol{Z} = \boldsymbol{z}_0$, will be convenient. From Eq. (B.22), we can write

$$\begin{aligned} V(t) &= E\left[(m(\boldsymbol{W}, \boldsymbol{\theta}_0) - m(\boldsymbol{W}, \boldsymbol{\theta}_0)(1 - m(\boldsymbol{W}, \boldsymbol{\theta}_0))) a^2(t, X) I(X < \tau_0)\right] \\ &\quad + \boldsymbol{b}^T(t) \boldsymbol{I}_0^{-1} \boldsymbol{b}(t). \end{aligned}$$

The first term of $V(t)$, namely $E(m(\boldsymbol{W}, \boldsymbol{\theta}_0) a^2(t, X) I(X < \tau_0))$, can be shown to be the standard Cox counterpart of $V(t)$, that is, the asymptotic variance of the normalized conditional cumulative hazard estimator. It is given by

$$V_{\mathrm{C}}(t) = \int_0^t \frac{\exp(2\boldsymbol{\beta}_0^T \boldsymbol{z}_0) \lambda_0(s)}{s^{(0)}(\boldsymbol{\beta}_0, s)} ds + \boldsymbol{E}_0^T(t, \boldsymbol{\beta}_0) \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \boldsymbol{E}_0(t, \boldsymbol{\beta}_0),$$

where a calculation involving a cross product term is 0, as was shown in a similar computation involving the quantity 'T_2' in the Section A.1.3 of the Appendix A. Therefore,

$$V(t) \quad = \quad V_{\mathrm{C}}(t) - E\left[m(\boldsymbol{W},\boldsymbol{\theta}_0)(1-m(\boldsymbol{W},\boldsymbol{\theta}_0))a^2(t,X)I(X<\tau_0)\right] + \boldsymbol{b}^T(t)\boldsymbol{I}_0^{-1}\boldsymbol{b}(t).$$

It remains to prove that

$$\boldsymbol{b}^T(t)\boldsymbol{I}_0^{-1}\boldsymbol{b}(t) \le E\left[m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)\bar{m}(X,\boldsymbol{Z},\boldsymbol{\theta}_0)a^2(t,X)I(X\le\tau_0)\right],$$

which, however, can be shown by following the steps given in the second proof of the Proposition 1, completing the proof. □

To apply the GMB, suppose that $G_1,\ldots,G_n$ denote independent standard normal random variables, generated independent of the data. Using $\hat{\boldsymbol{\theta}}$ and consistent estimates $\hat{a}(t,u)$ and $\hat{\boldsymbol{b}}(t)$, form $\hat{\mathbb{H}}^*(t) = L^*_{n,1}(t) + L^*_{n,2}(t)$, where

$$L^*_{n,1}(t) \quad = \quad n^{-1/2}\sum_{i=1}^{n} m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})\hat{a}_i(t,X_i)I(X_i<\tau_0)G_i, \tag{3.8}$$

$$L^*_{n,2}(t) \quad = \quad n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i - m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}{m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})\bar{m}(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}\hat{\boldsymbol{b}}^T(t)\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}}))G_i. \tag{3.9}$$

Let $[t_1,t_2] \subset [0,\tau_0)$. In Appendix B.2 we show that, for almost all sample sequences, $\hat{\mathbb{H}}^*(\cdot)$ has the same distribution as $\hat{\mathbb{H}}(\cdot)$. By the continuous mapping theorem, for almost all sample sequences, $\|\hat{\mathbb{H}}^*\|_{t_1}^{t_2}$ has the same distribution as $\|\hat{\mathbb{H}}\|_{t_1}^{t_2}$. For $0<\alpha<1$, let q_α denote the upper-α quantile of the distribution of $\|\hat{\mathbb{H}}^*\|_{t_1}^{t_2}$. A linear $100(1-\alpha)\%$ SCB for $\Lambda(t,\boldsymbol{z}_0)$ is given by

$$\left(\hat{\Lambda}(t,\boldsymbol{z}_0) - n^{-1/2}q_\alpha \quad , \quad \hat{\Lambda}(t,\boldsymbol{z}_0) + n^{-1/2}q_\alpha\right).$$

To construct more elaborate SCBs let φ be a known function whose derivative φ' is continuous and nonzero in $[t_1,t_2]\subset[0,\tau_0)$. Let $\hat{g}(t,\boldsymbol{z}_0)$ denote a weight function that converges uniformly in $[t_1,t_2]$ to the nonnegative bounded function $g(t,\boldsymbol{z}_0)$. Defining

$$\hat{\mathbb{B}}(t) = n^{1/2}\hat{g}(t,\boldsymbol{z}_0)[\varphi(\hat{\Lambda}(t,\boldsymbol{z}_0)) - \varphi(\Lambda(t,\boldsymbol{z}_0))],$$

by the functional delta method, the distribution of $\hat{\mathbb{B}}(t)$ can be approximated by

$$\hat{\mathbb{B}}^*(t) = \hat{g}(t, \boldsymbol{z}_0)\varphi'(\hat{\Lambda}(t, \boldsymbol{z}_0))\hat{\mathbb{H}}^*(t).$$

Let $\varphi(x) = \log x$. Let $\hat{\sigma}^2(t, \boldsymbol{z}_0)$ denote a consistent estimate of the asymptotic variance function of $\hat{\mathbb{H}}(t)$, see Eq. (B.9). Taking $\hat{g}(t, \boldsymbol{z}_0) = \hat{\Lambda}(t, \boldsymbol{z}_0)/\hat{\sigma}(t, \boldsymbol{z}_0)$, let $q_{1,\alpha}$ denote the upper-α quantile of the distribution of $\|\hat{\mathbb{B}}^*\|_{t_1}^{t_2}$. A $100(1-\alpha)\%$ SCB for $\log \hat{\Lambda}$ over $[t_1, t_2]$ is given by

$$\log \hat{\Lambda}(t, \boldsymbol{z}_0) \mp n^{-1/2} q_{1,\alpha} \hat{\sigma}(t, \boldsymbol{z}_0)/\hat{\Lambda}(t, \boldsymbol{z}_0).$$

Since $\log \Lambda = \log(-\log S)$, we obtain an EP-type band (Nair, 1984) as

$$\hat{S}(t, \boldsymbol{z}_0)^{\exp\left(\pm n^{-1/2} q_{1,\alpha} \hat{\sigma}(t,\boldsymbol{z}_0)/\hat{\Lambda}(t,\boldsymbol{z}_0)\right)}. \tag{3.10}$$

Taking $\hat{g}(t, \boldsymbol{z}_0) = \hat{\Lambda}(t, \boldsymbol{z}_0)/(1 + \hat{\sigma}^2(t, \boldsymbol{z}_0))$, let $q_{2,\alpha}$ be the upper-α quantile of the distribution of $\|\hat{\mathbb{B}}^*\|_{t_1}^{t_2}$. Then, a $100(1-\alpha)\%$ HW type SCBs for $S(t, \boldsymbol{z}_0)$, are given by

$$\begin{array}{ll} \text{(Transformed)} & \hat{S}(t, \boldsymbol{z}_0)^{\exp\left(\pm n^{-1/2} q_{2,\alpha}(1+\hat{\sigma}^2(t,\boldsymbol{z}_0))/\hat{\Lambda}(t,\boldsymbol{z}_0)\right)} \\ \text{(Untransformed)} & \hat{S}(t, \boldsymbol{z}_0) \mp n^{-1/2} q_{2,\alpha} \dfrac{1 + \hat{\sigma}^2(t, \boldsymbol{z}_0)}{\hat{\Lambda}(t, \boldsymbol{z}_0)}. \end{array} \tag{3.11}$$

Remark When there are MCIs ([14]), replace $L_{n,2}(t)$ given by Eq. (3.6) with

$$L_{n,2}(t) = n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i(\delta_i - m(\boldsymbol{W}_i, \boldsymbol{\theta}_0))}{m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i, \boldsymbol{\theta}_0)} \boldsymbol{b}^T(t) \tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1} \mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)), \tag{3.12}$$

where $\tilde{\boldsymbol{I}}_{\boldsymbol{0}}$ is given in the Eq. A.2. The Similar adjustment will be needed for $L_{n,2}^*(t)$, given by Eq. (3.9).

3.3 Simulation Results

In this section we report the results of simulation studies, based on sample size 100. Each method is first examined for performance efficacy, in terms of its empirical coverage probability (ECP), which is the proportion of 10,000 SCBs that include

$S(t, \boldsymbol{z}_0)$ for all $t \in [t_1, t_2]$. When the ECPs were found to be close to the nominal value of 95%, further comparisons between the methods were then based on the two measures *estimated average enclosed area* (EAEA) and *estimated average width* (EAW). Suppose that $u_j,\ j = 1, \ldots, n$ are the ordered observed minimums. Let $1 \le m_1 < n$ denote the integer such that $u_{m_1} \le t_1 < u_{m_1+1}$, and let m_2 denote the integer such that $m_1 < m_2 < n$ and $u_{m_1+1} < u_{m_2} < t_2 \le u_{m_2+1}$. Then, as in Subramanian and Zhang [12], the EAW and EAEA are defined over the interval $[u_{m_1}, u_{m_2}]$ by

$$\text{EAEA} = \frac{1}{k}\sum_{i=1}^{k}\left\{\sum_{j=m_1}^{m_2} l_j \Delta_{u_j}\right\}_i, \qquad \text{EAW} = \frac{1}{k}\sum_{i=1}^{k}\left\{\sum_{j=m_1}^{m_2} l_j \Delta_{S_j}\right\}_i,$$

where l_j denotes the width of the band computed at u_j, k denotes the number of replications, $\Delta_{u_j} = u_{j+1} - u_j$ and $\Delta_{S_j} = S_j - S_{j-1}$. The endpoints t_1 and t_2 were chosen as the 0.25 and 0.75 quantiles of the ordered values of $X_1, \ldots, X_n$. The critical values $q_{1,\alpha}$ and $q_{2,\alpha}$ in Eqs. (3.10)–(3.11) were based on 1,000 bootstrap replications.

To showcase specificity (turning in correct ECPs when m is correctly specified), the case of SRCMs with no misspecification was considered first; here, the exact models that were used to generate the censoring indicators were fitted. Since, in practice, misspecification is often the norm than the exception, the sensitivity of the proposed SCBs to misspecified models was also considered; here the fitted model for m, was different from that which generated the indicators. The same approach was also employed when there were MCIs. However, in the absence of a competitor for the MCIs scenario, only ECPs are presented. We also performed a robustness study that used an ill-fitting nonparametric model for the missingness probability.

3.3.1 Censoring Indicators Always Observed

Absence of MCIs permits comparison of the proposed SCBs with the SCBs developed by Lin et. al [4]. We shall denote their EP-type SCBs as Cox EP, and their HW-type transformed and untransformed SCBs as Cox HW1 and Cox HW2, respectively.

Specificity Study The one-dimensional covariate Z was taken to be uniformly distributed over $(1, 2)$. The conditional event-time and censoring hazards given $Z = z$ were $\exp(z)$ and $\exp(\gamma z)$, respectively, where γ was selected to give censoring rates (CRs) between 20% and 50%. The baseline hazard was taken to be unity. With these specifications, the true model for the conditional probability, $m(x, z) = P(\delta = 1|X = x, Z = z)$, turns out to be the logit model given by $m(z, \boldsymbol{\theta}) = 1/(1 + \exp(-\boldsymbol{\theta}^T \boldsymbol{w}))$, where $\boldsymbol{\theta} = (\theta_0, \theta_1)^T = (0, 1 - \gamma)^T$ and $\boldsymbol{w} = (1, z)^T$. This true model was fitted and SCBs for $S(t, z_0), t \in [t_1, t_2]$, were computed at two covariate levels $z_0 = 1.2$ and $z_0 = 1.8$. The ECP and the percentage relative reduction (PRR) in proposed EAEA and EAW values over the standard Cox counterparts are plotted as a function of the CR in Figure 3.1. The PRR is given by the formula 100(Cox − Proposed)/Cox. The upper panel of graphs are for $z_0 = 1.2$ and the lower panel are for $z_0 = 1.8$.

Both methods gave SCBs providing ECPs close to the nominal 95%. The PRR in EAEA and EAW values provided by the proposed SCBs over their standard Cox counterparts, for both Z values, varied from 3.5% to 12%. Thus, the proposed SCBs are more informative than the standard Cox ones, especially for higher censoring rates.

Sensitivity Study Here Z was taken as before, one dimensional and having the uniform distribution over $(1, 2)$. The conditional event-time hazard given $Z = z$ was taken as a Cox PH model with covariate $\log z$, the scalar regression parameter $\beta = 1$, and the baseline hazard equal to 1. The conditional censoring time was taken as uniform over $(0, \gamma Z)$, where γ was chosen to give CRs between 10% and 40%. 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

$$m(x, z, \boldsymbol{\theta}) = 0.5 + \frac{1}{\pi} \arctan(\theta_0 + \theta_1 x + \theta_2 z) \tag{3.13}$$

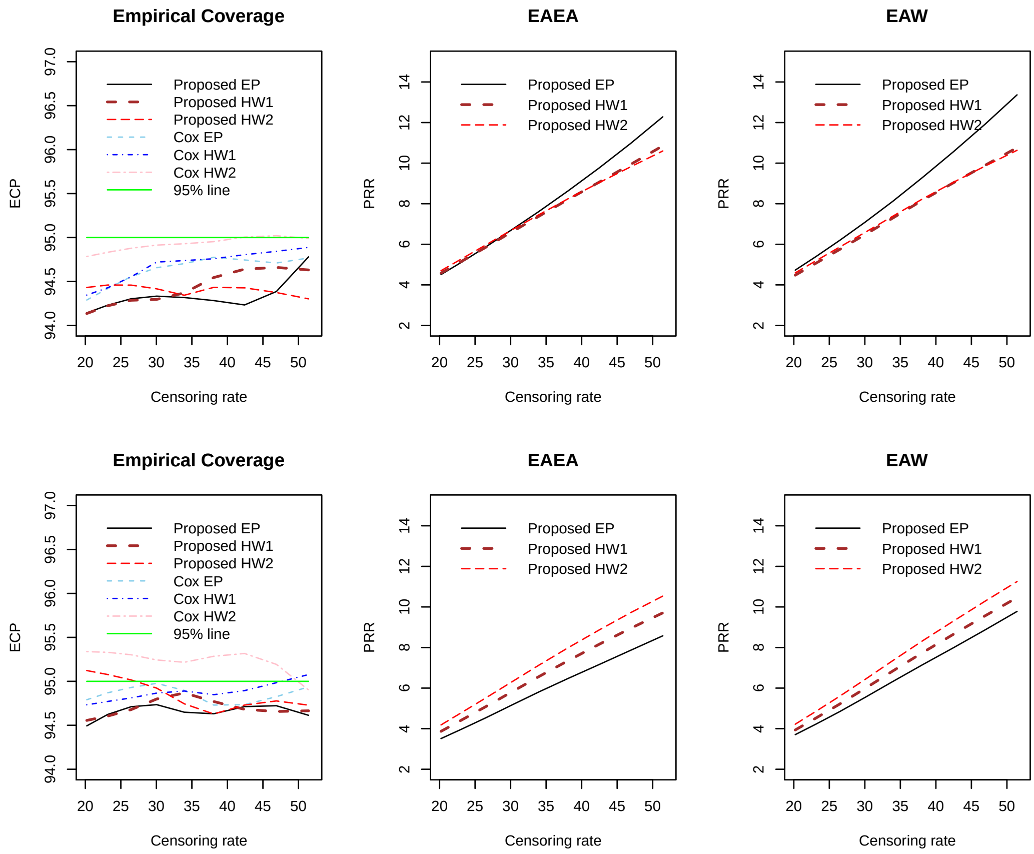

Figure 3.1 MCIs absent and the model for m is correctly specified. The ECP and relative reduction in EAEA and EAW values of proposed over standard Cox SCBs are plotted against censoring rate.

to the generated censoring indicators. The results are shown in Figure 3.2.

All the proposed SCBs still provided ECPs close to the nominal 95%. The PRR in EAEA and EAW values provided by the proposed SCBs over their standard Cox counterparts varied between 2% and 7%, with increased reduction seen for higher CRs. Thus, even when the model was misspecified, the proposed SCBs are more informative than the standard Cox ones.

3.3.2 Censoring Indicators Missing at Random

Let $\pi(x, \boldsymbol{z})$ denote the conditional expectation of the missingness indicator ξ given $X = x$ and $\boldsymbol{Z} = \boldsymbol{z}$. We investigate four scenarios resulting from different combinations

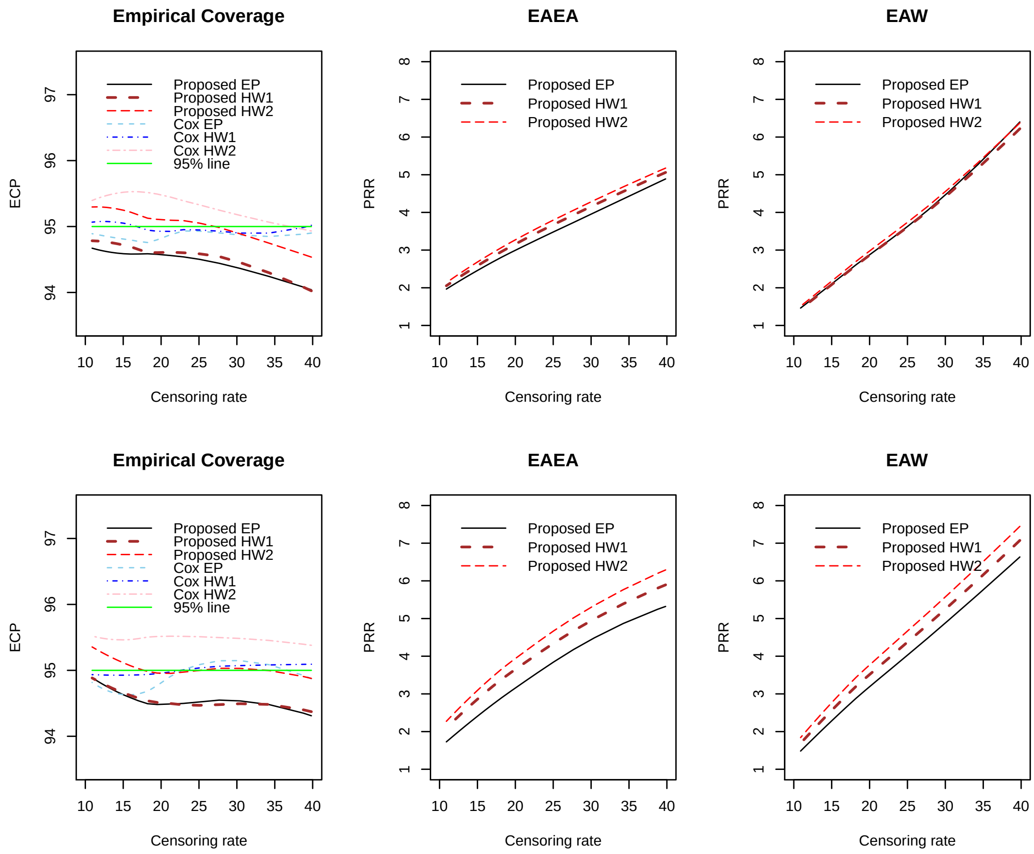

Figure 3.2 MCIs absent and the model for m is misspecified. The ECP and relative reduction in EAEA and EAW values of proposed over standard Cox SCBs are plotted against censoring rate.

of specificity and sensitivity study for m and π. *For example, we indicate* ***specificity*** $\times$ ***specificity*** *to denote the case that there is no misspecification of m as well as π.* In the absence of other methods to compare with the proposed, only ECPs are reported for each scenario.

Specificity $\times$ Specificity Study Data were generated according to the scheme described in Section 3.3.1 and the true logit model was fitted for m as in Section 3.3.1. To impute missingness, the logit model $\pi(x, z, \alpha) = 1/(1+\exp(-\alpha(x+z)))$ was used, where α was chosen to give the two missingness rates (MRs) 10% and 40%. We fitted the true logit model $\pi(x, z, \boldsymbol{\theta}) = 1/(1+\exp(-\boldsymbol{\theta}^T \boldsymbol{w}))$, where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^T = (0, \alpha, \alpha)^T$

and $\boldsymbol{w} = (1, x, z)^T$. The graphs in the upper half of Figure 3.3 give ECPs as a function of the CR, for the two MRs 10% and 40% and for $z_0 = 1.2$. The graphs in the lower half are for $z_0 = 1.8$. The ECPs are close to the nominal 95%.

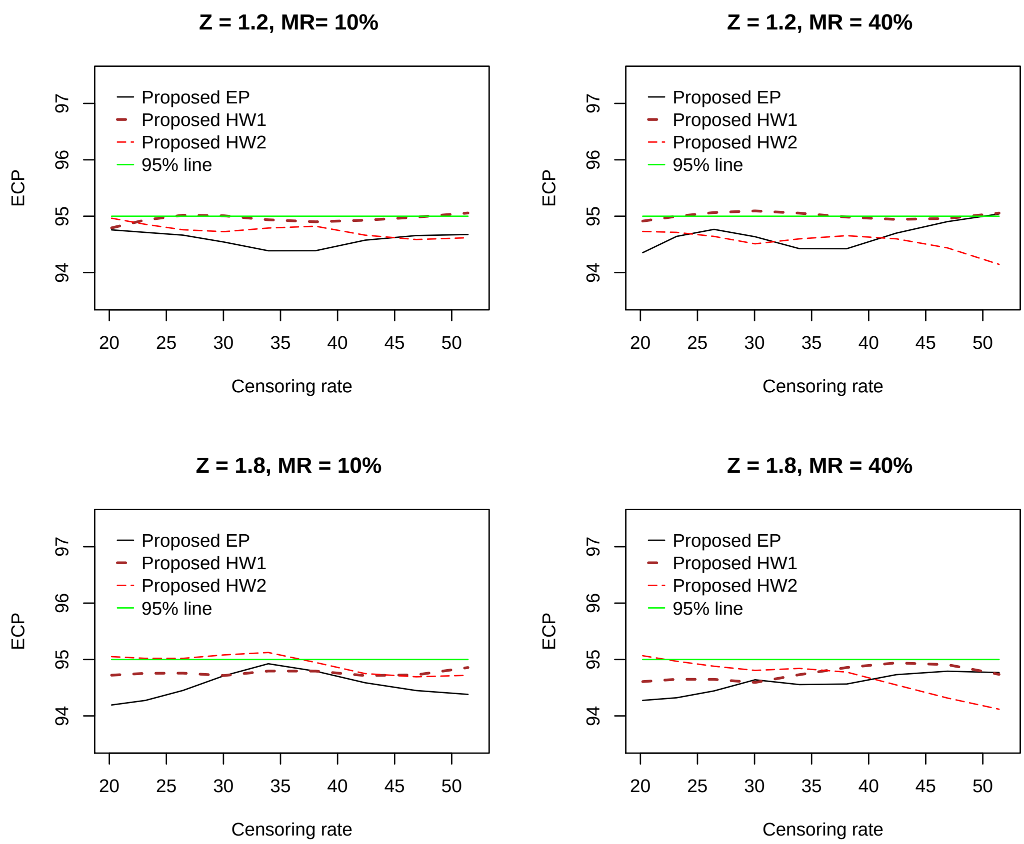

Figure 3.3 Specificity × Specificity study. Correctly specified models for both m and missingness probability. The ECPs of the proposed SCBs are plotted against censoring rate.

Sensitivity × Specificity study Data were generated according to the scheme described in Section 3.3.1 and the Cauchy link, Eq. (3.13), was fitted for m. Thus, the model for m was misspecified. The model for π was not misspecified and fitting was as in Section 3.3.2. The plots of ECP versus CR, shown in Figure 3.4, indicate that the proposed SCBs provide ECPs close to the nominal 95%.

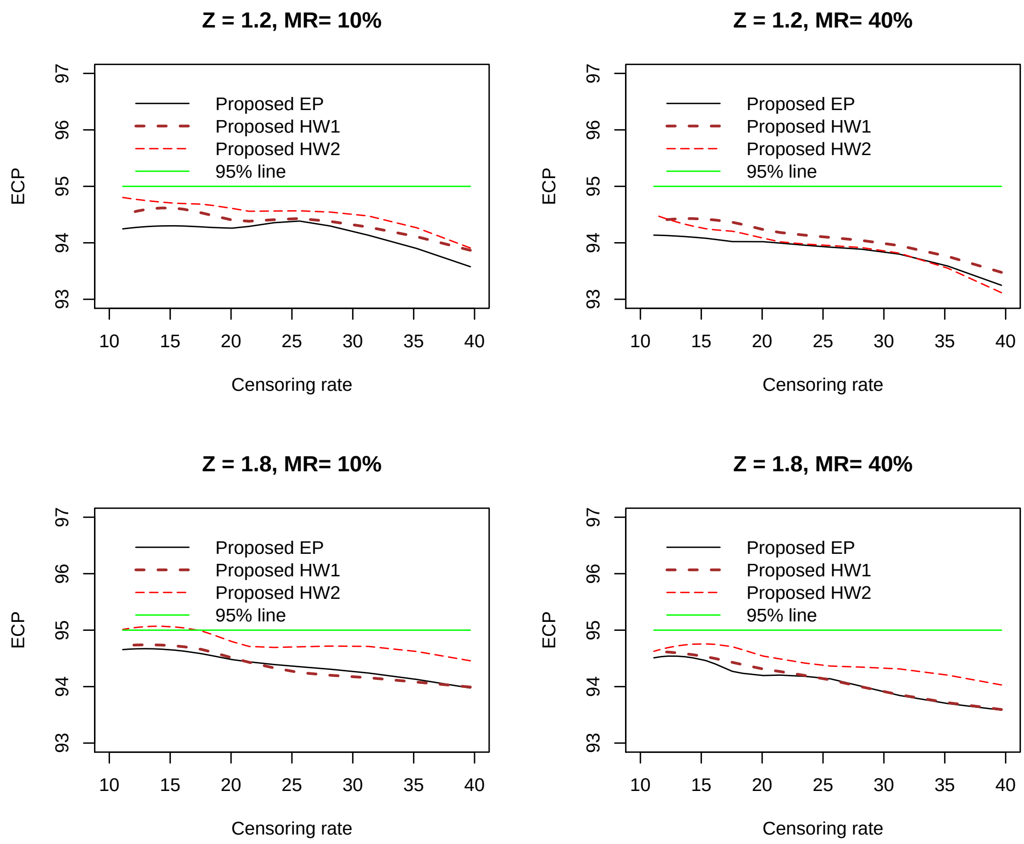

Figure 3.4 Sensitivity $\times$ Specificity study. Correctly specified model for the missingness probability but model for m is misspecified. The ECPs of the proposed SCBs are plotted against censoring rate.

Specificity $\times$ Sensitivity study Data were generated according to the scheme described in Section 3.3.2. Since m was correctly specified, its fitting was as in Section 3.3.2. We estimated π using the average of the ξ_i's. This global nonparametric estimator was a misspecified model for π, which depends on both X and Z, see Section 3.3.2. Figure 3.5 shows the plot of ECP versus CR for the two values $z_0 = 1.2$ and $z_0 = 1.8$, and the two MRs as described in Section 3.3.2. The plot indicates that even with the crude estimator of π the proposed SCBs provided ECPs close to 95%.

Sensitivity $\times$ Sensitivity study Data were again generated according to the scenario described in Section 3.3.1 and the model for m is misspecified by fitting

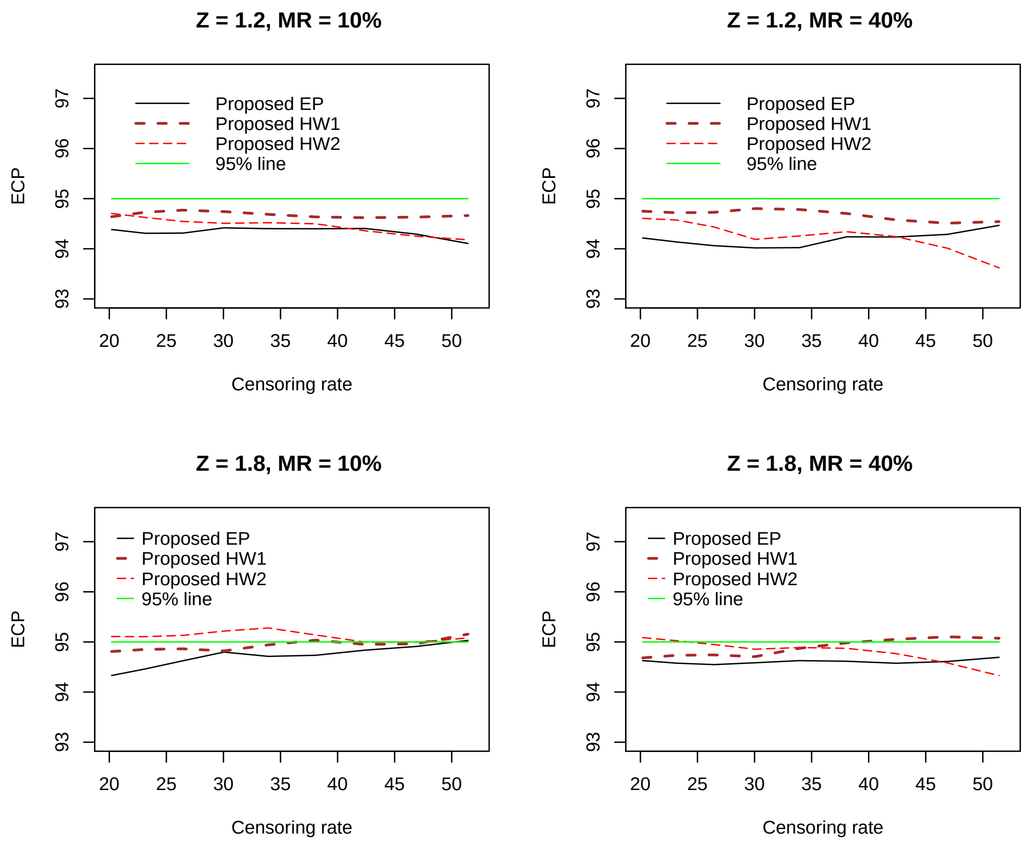

Figure 3.5 Specificity × Sensitivity study. Correctly specified model for m but the model for the missingness probability is misspecified. The ECPs of the proposed SCBs are plotted against censoring rate.

Cauchy link, Eq. (3.13), as before. Also, π was misspecified with nonparametric estimator given in Section 3.3.2. Figure 3.6 shows that ECPs for the proposed SCBs are close to 95% except for untransformed bands, with $z_0 = 1.2$, 40% MR and high CRs between 35% and 40%, when the ECPs fell slightly below 93%. It must be noted, however, that such misspecification is unrealistic in practice. Parametric specifications of π (e.g., Cauchy or logit) will frequently root out such severe misspecification.

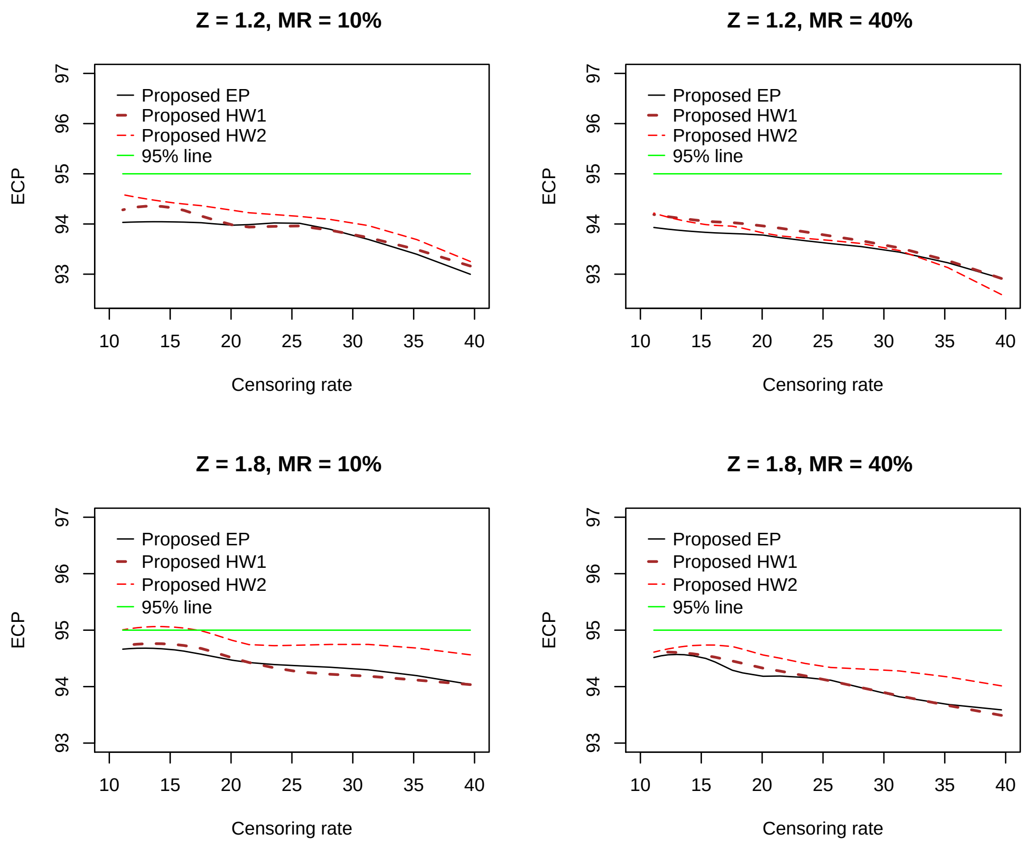

Figure 3.6 Sensitivity × Sensitivity study. Misspecified models for both m and the missingness probability. The ECPs of the proposed SCBs are plotted against censoring rate.

3.4 Real Examples

3.4.1 Illustration Using Primary Biliary Cirrhosis Data

We illustrate the proposed SCBs using the primary biliary cirrhosis (PBC) data from the Mayo clinic database. Lin et al. [4] reported that the Cox PH model with five covariates, namely age, log(albumin), log(bilirubin), Oedema, and log(prothrombin time), denoted by $Z_1, \ldots, Z_5$ respectively, provided a good fit for the PBC data. Our analysis was based on the 416 patients with complete data on the covariates, $Z_1, \ldots, Z_5$. We fitted several available binary response regression models for m, with link functions like Cauchy, logit, probit, and complementary log-log, and used the Akaike Information Criterion and the Bayesian Information Criterion to determine the most adequate one, which turned out to be the Cauchy model with covariates X, Z_1, Z_3, Z_4, Z_5. Note that Z_2 was found to be insignificant for the purposes of fitting m but was otherwise used in the analysis. We therefore, fitted the model

$$m(\boldsymbol{W}, \boldsymbol{\theta}) = 0.5 + \arctan(\theta_0 + \theta_1 X + \theta_2 Z_1 + \theta_3 Z_3 + \theta_4 Z_4 + \theta_5 Z_5). \tag{3.14}$$

We have plotted in Figure 3.7 the estimated subject-specific survival curves accompanied by 95% transformed HW- and EP-type SCBs for subjects with the covariate measurement $\boldsymbol{z}_0 = (51, 3.4, 1.8, 0, 10.74)$ — note that age was 51, log(albumin) was 3.4 and so on. We have plotted the standard Cox-based transformed HW- and EP-type bands for comparison. As noted by Lin et al. [4], the estimated subject-specific cumulative hazard is very small for $t < 2$, due to which the lower bounds of the HW-type SCBs were very low over that region. We therefore, set the lower bound for all $t < 2$, equal to the lower bound calculated at $t = 2$. The proposed HW-type SCBs are more informative (in the sense of being more tight) over the entire time span. Comparison of the EP-type bands show that the new SCBs are narrower for most of the time span except for the years between 6 and 8, where they are similar to their standard Cox counterparts. The PRR values, given in Table 3.1, also indicate that the proposed HW- and EP-type SCBs performed better than their standard Cox counterparts.

3.4.2 A Second Illustration Using Kidney Transplant Data

For our second illustration, we chose the data on the death time of 863 patients who underwent kidney transplant at the Ohio State University Transplant center between 1982 and 1992. Patients were censored if they were lost to follow-up or alive till June 30, 1992. The data are described in Section 1.7 of Klein and Moeschberger [37]. Three covariates, namely, age, race and gender of patients were listed. A model checking in terms of Cox-Snell residuals ensured that Cox proportional hazard model fitted well for this data set and only age came out as the significant covariate whereas gender

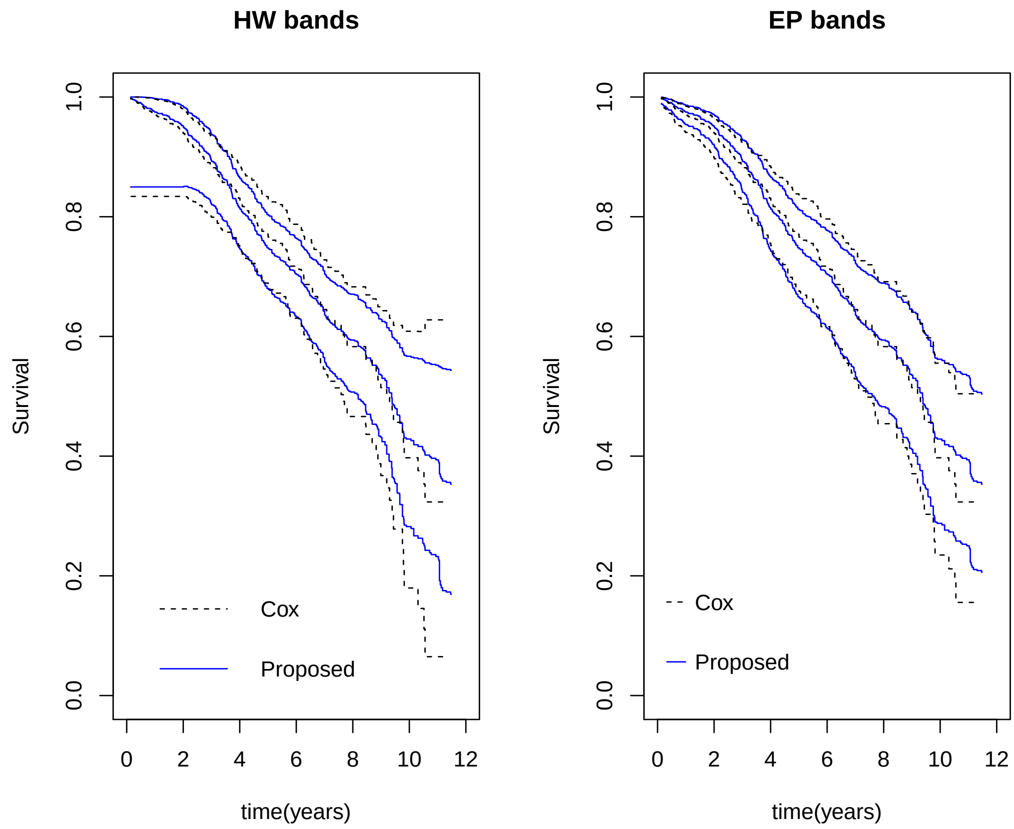

Figure 3.7 Primary Biliary Cirrhosis data analysis: comparison of proposed and standard Cox-based SCBs for subject-specific survival curves.

and race were not important. The final model we fit for m is the logit model,

$$m(X, \boldsymbol{Z}, \boldsymbol{\theta}) = 1/(1 + \exp(-\theta_0 - \theta_1 \text{time} - \theta_2 \text{age})). \tag{3.15}$$

We have plotted the estimated subject specific survival curve of the patient with age 43 (mean age of all patients) with two kinds of SCBs, transformed EP and HW type, in Figure 3.8. The estimated subject specific cumulative hazards were very small for $t < 162$ days and this made lower bounds of HW-type SCBs were very low over that region. We set the lower bound for $t < 162$ is equal to the lower bound calculated at $t = 162$. Similar adjustment was done at $t = 104$ for Cox counterpart. Figure 3.8 shows that both EP and HW bands are narrower than the Cox counterparts. Also,

Table 3.1 gives the PRR values for two types of bands which indicates that proposed bands performed significantly better than their Cox counterparts.

Table 3.1 Percentage Relative Reduction (PRR) in Estimated Average Enclosed Area (EAEA) and Estimated Average Width (EAW) of Proposed SCBs over Standard Cox.

	EP type		HW type	
Data	EAEA	EAW	EAEA	EAW
Primary biliary cirrhosis	6.43	18.06	15.69	29.45
Kidney transplant	10.67	23.33	16.52	23.79

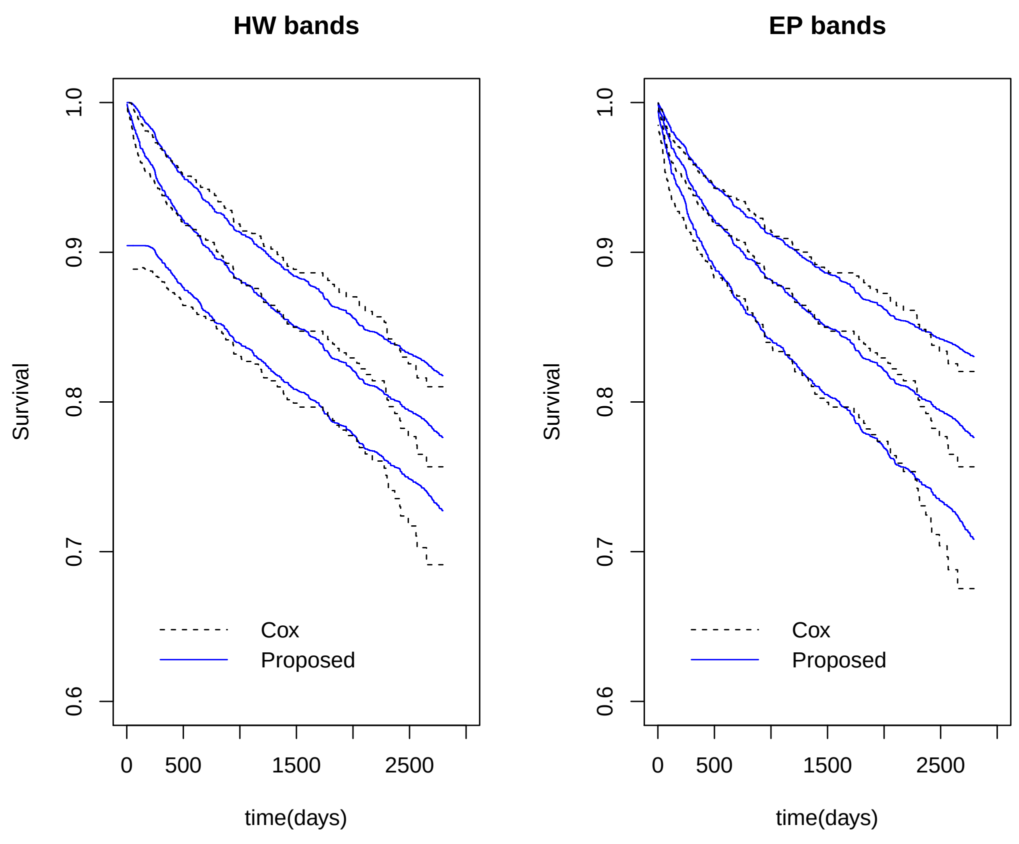

Figure 3.8 Kidney Transplant data analysis: comparison of proposed and standard Cox-Based SCBs for subject-specific survival curves.

3.5 Conclusion

Model assisted Cox PH regression offers significant strengthening of standard Cox PH analysis. We have developed several SCBs for subject-specific survival, the construction of which requires the application of the GMB. The proposed SCBs are more informative than existing ones, even when there may be some parametric misspecification. Indeed, they have the potential to strengthen marginal conclusions obtained by standard Cox based SCBs. Furthermore, they are easy to compute in the presence of MCIs, a facility not shared by the existing SCBs; and can be extended to produce SCBs for the difference of two survival functions ([1]).

CHAPTER 4

CONCLUDING REMARKS AND FUTURE WORK

This research is concerned with constructing simultaneous confidence bands for a subject-specific survival curve under model-based Cox regression. The asymptotic properties of the proposed model-based Cox regression estimators have been derived and it has been shown that, when the model for conditional probability is correctly specified, the model-based estimator of the regression parameter is asymptotically more efficient than its standard Cox counterpart. On the basis of numerical studies it is clear 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, the proposed approach gave better estimates of Cox regression parameters. The proposed method is comparable to that of Yuan (2005), whose method, however, can not be applied when the censoring indicators are missing at random. Under MCIs, the proposed method performed better than Liu and Wang's (2010) procedure, who, however, did not provide any estimates for baseline hazard. The proposed method provides a unified approach that can handle both cases of absence and presence of MCIs without any extra effort.

The results of chapter 2 have been published in *Journal of Multivariate Analysis*, a mainstream statistics journal. During the review process, 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 $k(\boldsymbol{z})$ as the density function of $\boldsymbol{Z}$, the adjusted likelihood, after factoring in the model for m, takes the form [see also Yuan

[23])

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

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 likelihood, whose maximizer can be found numerically, see Lu (2008), who 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 $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$ and the Breslow estimator $\Lambda_{0\mathrm{C}}$, see Lin and Zeng (2007). It would be a worthwhile direction for further research to investigate whether such direct maximization of the likelihood would yield improved estimators of $\boldsymbol{\beta}$ and $\Lambda_0(t)$, without compromising the simplicity of analysis as well. Yuan (2005) obtained a likelihood by *profiling* out λ_0 in the above likelihood, which he then maximized to obtain his estimators of $\boldsymbol{\beta}$ and $\Lambda_0(t)$. However, our proposed approach offers an attractive alternative and performs as well as his estimators.

APPENDIX A

ASYMPTOTIC PROPERTIES

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 Fleming and Harrington [16].

Let $\boldsymbol{W} = (X, \boldsymbol{Z})$ and $\bar{m}(\boldsymbol{w}, \boldsymbol{\theta}) = 1 - m(\boldsymbol{w}, \boldsymbol{\theta})$. Let $D_r(m(\boldsymbol{w}, \boldsymbol{\theta}))$ denote the partial derivative of $m(\boldsymbol{w}, \boldsymbol{\theta})$ with respect to θ_r. Write $\text{Grad}(m(\boldsymbol{w}, \boldsymbol{\theta})) = [D_1(m(\boldsymbol{w}, \boldsymbol{\theta})), \ldots, D_k(m(\boldsymbol{w}, \boldsymbol{\theta}))]^T$ and let $\boldsymbol{J}_{\boldsymbol{\theta}}(t, \boldsymbol{z}) = [(\text{Grad}(m(t, \boldsymbol{z}, \boldsymbol{\theta})))^{\otimes 2}]$, where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. Vectors and matrices will be in **bold**. The information in the absence of MCIs [see Eq. (4) on page 257 of [13]] and the presence of MCIs [see Eq. (3.11) on page 134 of [14]] are given by

$$\boldsymbol{I}(\boldsymbol{\theta}_0) \equiv \boldsymbol{I_0} = E\left(\frac{\boldsymbol{J}_{\boldsymbol{\theta}_0}(\boldsymbol{W})}{m(\boldsymbol{W}, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}, \boldsymbol{\theta}_0)}\right), \tag{A.1}$$

$$\tilde{\boldsymbol{I}}(\boldsymbol{\theta}_0) \equiv \tilde{\boldsymbol{I}}_{\boldsymbol{0}} = E\left(\frac{\pi(\boldsymbol{W})\boldsymbol{J}_{\boldsymbol{\theta}_0}(\boldsymbol{W})}{m(\boldsymbol{W}, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}, \boldsymbol{\theta}_0)}\right). \tag{A.2}$$

Writing $Y(t) = I(X \geq t)$, we define the following quantities; see Andersen and Gill [18]:

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

$$\boldsymbol{s}^{(m)}(\boldsymbol{\beta}, t) = E[Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}}\boldsymbol{Z}^{\otimes m}], \qquad m = 0, 1, 2;$$

$$\bar{\boldsymbol{Z}}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)}; \qquad \bar{\boldsymbol{z}}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)};$$

$$\boldsymbol{v}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}\right)^{\otimes 2}.$$

Note that $\Lambda_{X|\boldsymbol{Z}}(t)$, the conditional cumulative hazard function of X given $\boldsymbol{Z}$, is the sum of the conditional cumulative hazards of T and C, and that $m(t, \boldsymbol{z})$ is the ratio of the event-time hazard to the total hazard. It follows that $\Lambda_{X|\boldsymbol{Z}}(t) = \int_0^t e^{\boldsymbol{\beta}_0' \boldsymbol{Z}} \lambda_0(u) du / m(u, \boldsymbol{z}, \boldsymbol{\theta}_0)$. See also, page 497 of Yuan [23], where his $\gamma(x, z, \boldsymbol{\theta})$ satisfies the relation $1 + \gamma(x, z, \boldsymbol{\theta}) = 1/m(x, z, \boldsymbol{\theta})$. For each $i = 1, \ldots, n$, the counting process $N_i(t) = I(X_i \leq t)$, conditional on $\boldsymbol{Z}_i$, has a compensator $\int_0^t Y_i(s) d\Lambda_{X_i|\boldsymbol{Z}_i}(s)$ so that

$$M_i(t) = N_i(t) - \int_0^t \frac{Y_i(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} \lambda_0(u)}{m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)} du \tag{A.3}$$

is a martingale with respect to the sigma-field $\mathcal{F}_t = \sigma\{\boldsymbol{Z}_i, I(X_i \leq s), i = 1, \ldots, n : s \leq t\}$, see page 81 of Bickel et al. [38]. Furthermore, the counting process $N_i^u(t) = I(X_i \leq t, \delta_i = 1)$, conditional on $\boldsymbol{Z}_i$, has a compensator $\int_0^t Y_i(s) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} \lambda_0(s) ds$ so that

$$M_i^u(t) = N_i^u(t) - \int_0^t Y_i(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} \lambda_0(u) du \tag{A.4}$$

is a martingale with respect to the sigma-field $\breve{\mathcal{F}}_t = \sigma\{\boldsymbol{Z}_i, N_i^u(s), Y_i(s+), i = 1, \ldots, n : s \leq t\}$; see page 128 of Fleming and Harrington [16]. The corresponding predictable covariation processes are given by

$$\langle M^u, M^u \rangle(t) = \int_0^t Y(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}} \lambda_0(u) du; \tag{A.5}$$

$$\langle M, M \rangle(t) = \int_0^t \frac{Y(u) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}} \lambda_0(u)}{m(u, \boldsymbol{Z}, \boldsymbol{\theta}_0)} du. \tag{A.6}$$

See Eqs. (2.1)–(2.3) of Andersen and Gill [18] for $\langle M^u, M^u \rangle(t)$. Theorem 2.5.2 of Fleming and Harrington [16] yields the expression for $\langle M, M \rangle(t)$.

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

$$\begin{aligned}
\mathbf{\Sigma}_{\mathrm{C}} &= \int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, t)\lambda_0(t)dt, && \text{(A.7)}\\
\mathbf{\Sigma}_{\mathrm{C}} &= E\left[\int_0^\tau m(t, \boldsymbol{Z}, \boldsymbol{\theta}_0)\boldsymbol{v}(\boldsymbol{\beta}_0, t)dN(t)\right], && \text{(A.8)}\\
\mathbf{\Sigma}_{\mathrm{C}} &= E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta_0})(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right]. && \text{(A.9)}
\end{aligned}$$

Andersen and Gill [18] derived Eq. (A.7). Furthermore, following the proof of theorem 3.2 of Andersen and Gill [18], it can be shown that the right hand side of Eq. (A.7) also equals $E\left[\int_0^\tau (\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, s))^{\otimes 2} dN^u(s)\right]$. Apart from the fact that this term is the right hand side of Eq. (A.9), a simple conditional argument applied to this term also establishes Eq. (A.8). We will also need the following quantities:

$$\begin{aligned}
\boldsymbol{V_0} &= E\left[(m(X, \boldsymbol{Z}, \boldsymbol{\theta_0}))^2 (\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right], && \text{(A.10)}\\
\boldsymbol{B_0} &= E\left[\int_0^\tau [\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t)][\mathrm{Grad}(m(\boldsymbol{W}, \boldsymbol{\theta_0}))]^T dN(t)\right], && \text{(A.11)}\\
\mathbf{\Sigma} &= \mathbf{\Sigma}_{\mathrm{C}}^{-1}\left[\boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{B_0}^T + \boldsymbol{V_0}\right]\mathbf{\Sigma}_{\mathrm{C}}^{-1}, && \text{(A.12)}\\
\mathbf{\Sigma}_{\mathrm{M}} &= \mathbf{\Sigma}_{\mathrm{C}}^{-1}\left[\boldsymbol{B_0}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\boldsymbol{B_0}^T + \boldsymbol{V_0}\right]\mathbf{\Sigma}_{\mathrm{C}}^{-1}, && \text{(A.13)}\\
\boldsymbol{C_0}(t) &= \int_0^t \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, s)}{s^{(0)}(\boldsymbol{\beta}_0, s)}\lambda_0(s)ds, && \text{(A.14)}\\
\boldsymbol{D_0}(t) &= E\left[\int_0^t \frac{\mathrm{Grad}(m(s, \boldsymbol{Z}, \boldsymbol{\theta}_0))}{s^{(0)}(\boldsymbol{\beta}_0, s)}dN(s)\right], && \text{(A.15)}\\
\boldsymbol{d}_0(t) &= \boldsymbol{D_0}(t) - \boldsymbol{B_0}^T\mathbf{\Sigma}_{\mathrm{C}}^{-1}\boldsymbol{C_0}(t), && \text{(A.16)}\\
\alpha(t, X, \boldsymbol{Z}, \boldsymbol{\beta}_0) &= \frac{I(X \le t)}{s^{(0)}(\boldsymbol{\beta}_0, X)} - [\boldsymbol{C_0}(t)]^T\mathbf{\Sigma}_{\mathrm{C}}^{-1}[\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X)], && \text{(A.17)}\\
\gamma(t_1, t_2) &= E\left[m^2(X, \boldsymbol{Z}, \boldsymbol{\theta_0})\alpha(t_1, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\alpha(t_2, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)I(X \le \tau)\right], && \text{(A.18)}\\
\sigma(t_1, t_2) &= [\boldsymbol{d}_0(t_1)]^T\boldsymbol{I_0}^{-1}[\boldsymbol{d}_0(t_2)] + \gamma(t_1, t_2), && \text{(A.19)}\\
\sigma(t, t) &= [\boldsymbol{d}_0(t)]^T\boldsymbol{I_0}^{-1}[\boldsymbol{d}_0(t)] + \gamma(t, t), && \text{(A.20)}\\
\sigma_{\mathrm{M}}(t_1, t_2) &= [\boldsymbol{d}_0(t_1)]^T\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}[\boldsymbol{d}_0(t_2)] + \gamma(t_1, t_2), && \text{(A.21)}\\
\sigma_{\mathrm{M}}(t, t) &= [\boldsymbol{d}_0(t)]^T\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}[\boldsymbol{d}_0(t)] + \gamma(t, t). && \text{(A.22)}
\end{aligned}$$

Throughout the paper we shall assume that T and C are conditionally independent given $\boldsymbol{Z}$. We shall also need the following conditions to prove theorems 1 and 2.

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

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

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

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

A. 4. *The matrix* $\boldsymbol{\Sigma}_{\mathrm{C}}$ *[cf. Eqs. (A.7) - (A.9)] is positive definite.*

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

A. 6. *The function* $m(x, \boldsymbol{z})$ *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 Andersen and Gill [18] and in theorem 8.4.1. of Fleming and Harrington [16], both for the iid case, and considered in this paper as well. Although alternative set of weaker conditions involving the finiteness of the second moments of $\boldsymbol{Z}$ and $M(X, \boldsymbol{Z})$ (defined in condition **D.1** below) 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 Fleming and Harrington [16]. The set of conditions **A.2**–**A.4** were given by Andersen and Gill [18] for standard Cox and discussed well there; see also pages 289–290 of Fleming and Harrington [16]. 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* $\boldsymbol{z}$ *such that, for each* $r = 1,\ldots,k$, $|\mathrm{D}_r(m(x,\boldsymbol{z},\boldsymbol{\theta}))| \le M(x,\boldsymbol{z})$ *and* $E(M(X,\boldsymbol{Z})) < \infty$.

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 Dikta [13] , of which **D.1** is an extension.

Note that, since $\hat{\boldsymbol{\theta}}$ is derived via maximum likelihood [cf. Eq. (3.1)], it is an M-estimator and therefore, has the representation given by

$$\begin{aligned} n^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) &= n^{-1/2}\sum_{i=1}^{n} \boldsymbol{I_0}^{-1} \frac{\delta_i - m(\boldsymbol{W}_i,\boldsymbol{\theta}_0))}{m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i,\boldsymbol{\theta}_0)} \mathrm{Grad}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)) \\ &\quad + o_p(1). \end{aligned} \tag{A.23}$$

See page 241 of Tsiatis, Davidian, and McNeney [39]. For binary regression, see Example 5.40 of van der Vaart [40]. For more general m, adaptations of conditions in Perlman [41] as carried out by Dikta [13] , 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 $\breve{\boldsymbol{l}}_n(\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}})$, defined by Eq. (2.5), is free of $\boldsymbol{\beta}$, it follows that $\hat{\boldsymbol{\beta}}$ maximizes

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

In order to apply the methods of Andersen and Gill [18], we introduce the function

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

and show that it is concave. Indeed, applying Eq.(A.3) followed by condition **A.6**,

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

coincides with Andersen and Gill's [18] concave limit function of $A(\boldsymbol{\beta},\tau)$ (as well as of $X(\boldsymbol{\beta},\tau)$), see p. 1106 of their paper, and, hence, $\tilde{\boldsymbol{l}}(\boldsymbol{\beta},\boldsymbol{\theta}_0)$ is a concave function of $\boldsymbol{\beta}$. Since $\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})$ 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 Andersen and Gill [18], provided it can be shown that

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

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

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

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

$$\|\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})-\tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)\| = o_p(1) \quad (A.24)$$

$$\|\tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)-\tilde{\boldsymbol{l}}(\boldsymbol{\beta},\boldsymbol{\theta}_0)\| = o_p(1). \quad (A.25)$$

The strong law of large numbers implies Eq. (B.24). To prove Eq. (B.23), Taylor's expansion of $\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ yields

$$\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})-\tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0) = [\boldsymbol{l}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)-\tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)]+\left(\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{l}_n(\boldsymbol{\beta},\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0), \quad (A.26)$$

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.26) 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 $|\langle \text{Grad}_{\boldsymbol{\theta}}(m(x,\boldsymbol{z},\boldsymbol{\theta}^*)), (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\rangle| \leq kM(x,\boldsymbol{z})||\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0||$. Then,

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

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

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

A.1.2 Asymptotic Normality of $\hat{\beta}$

Defining three normalized sums of iid random variables by

$$
\begin{aligned}
\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= n^{-1/2}\sum_{i=1}^{n} \boldsymbol{B_0}\boldsymbol{I_0}^{-1}\frac{\delta_i - m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}{m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}\text{Grad}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)), \\
\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau} [(\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t))]\, dM_i^u(t), \\
\boldsymbol{U}_{n,3}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= -n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau} [(\delta_i - m(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0))(\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t))]\, dN_i(t),
\end{aligned}
$$

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

$$
n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) + \boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0) + \boldsymbol{U}_{n,3}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)] + o_p(1). \quad \text{(A.27)}
$$

Defining $n\boldsymbol{A}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}) = \partial \boldsymbol{S}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and $n\boldsymbol{B}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \partial \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, and using Taylor's expansion and the consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, we can show that

$$\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + n\boldsymbol{B}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + n\boldsymbol{A}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(n^{1/2}).$$

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

$$\begin{aligned} -\boldsymbol{A}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}}) \left(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right) &= n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\beta}_0, \theta_0) + \boldsymbol{B}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \left(n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right) \\ &\quad + o_p(1). \qquad \text{(A.28)} \end{aligned}$$

After some basic algebra, the first term on the right side of Eq. (A.28) can be expressed as

$$\begin{aligned} & n^{-1/2} \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \qquad \text{(A.29)} \\ = \; & n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left(\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t) \right) dN_i^u(t) + \boldsymbol{U}_{n,3}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \\ & + n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} (\delta_i - m(t, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)) \left(\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t) - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t) \right) dN_i(t). \qquad \text{(A.30)} \end{aligned}$$

We carry out in detail, once, an application of lemma 2 of Gilbert, McKeague, and Sun [42] 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.29) can be expressed as

$$n^{1/2} \int_0^{\tau} (\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t) - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t)) dQ_n(t),$$

where $nQ_n(t) = \sum_{i=1}^{n} \int_0^{\tau} (\delta_i - m(t, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)) dN_i(t)$. Note that $n^{1/2}Q_n(\cdot)$ converges weakly to a zero-mean Gaussian process on $D[0, \tau]$ and that $\sup_{t \in [0,\tau]} |\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t) - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t)| \xrightarrow{P} 0$. We can apply lemma 2 of Gilbert et al. [42] to conclude that the third term is $o_p(1)$, provided that $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, s)$ has total variation bounded in probability (TVBP) and $\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, s)$ 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 Apostol [43]. Since $\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, s)$ is a ratio of two monotonic functions the denominator being uniformly bounded away from zero (conditon **A**.3), it is of BV. Since $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, s)$ is a ratio of $\boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, s)$ and $S^{(0)}(\boldsymbol{\beta}_0, s)$, we first prove that $\boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, s)$ and $S^{(0)}(\boldsymbol{\beta}_0, s)$ each have TVBP. For simplicity, consider a one-dimensional covariate ($p = 1$) and let $0 = t_0 < t_1 ...... < t_{k-1} < t_k = \tau$ denote an arbitrary partition of the interval $[0, \tau]$. Defining $\Delta\boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, t_j) = \boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, t_j) - \boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, t_{j-1})$, we have

$$\begin{aligned}
\sum_{j=1}^{k} \left|\Delta\boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, t_j)\right| &= \sum_{j=1}^{k} \left|\frac{1}{n}\sum_{i=1}^{n} I(t_{j-1} \le X_i < t_j) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} \boldsymbol{Z}_i\right| \\
&\le \sum_{j=1}^{k} \frac{1}{n}\sum_{i=1}^{n} I(t_{j-1} \le X_i < t_j) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i} |\boldsymbol{Z}_i| \\
&\le M^* \sum_{j=1}^{k} \frac{1}{n}\sum_{i=1}^{n} I(t_{j-1} \le X_i < t_j),
\end{aligned}$$

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

$$\begin{aligned}
\sum_{j=1}^{k} \left|\Delta\boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0, t_j)\right| &\le M^* \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} I(t_{j-1} \le X_i < t_j) \\
&= M^* \frac{1}{n}\sum_{i=1}^{n} I(X_i < \tau) \\
&= M^* \left(P(X < \tau) + o_p(1)\right) \\
&\le M^* + o_p(1).
\end{aligned}$$

Likewise for $S^{(0)}(\boldsymbol{\beta}_0, s)$. To complete the proof we will need to show that the property of TVBP is closed under reciprocal and product operations. Indeed, from $|\Delta l_n(t_j)| \le \Delta g_n(t_j)/b^2$, where $l_n = 1/g_n$ and $g_n(x) \ge b > 0$ 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, $S^{(0)}(\boldsymbol{\beta}_0, s)$ is bounded away from 0 for all s, hence, its reciprocal has TVBP. The proof of theorem 6.9 of Apostol [43] shows that $h = fg$ satisfies $|\Delta h(t_j)| \le A|\Delta g(t_j)| + B|\Delta f(t_j)|$, where $A = \sup\{f(t) : t \in [0, \tau]\}$ and $B = \sup\{g(t) : t \in [0, \tau]\}$, from which we can conclude that $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, s)$ has TVBP.

For the first term in Eq. (A.29), again apply lemma 2 of Gilbert et al. [42] to obtain

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left[(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0,t))\right]dN_i^u(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left[(\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0,t))\right]dM_i^u(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left[(\boldsymbol{Z}_i-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t))\right]dM_i^u(t)+o_p(1) \\
&= \boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0)+o_p(1).
\end{aligned}
$$

Then, applying condition **D.1**, it follows that

$$
\begin{aligned}
\boldsymbol{B}_n(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{Z}_i-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t)\right]\left[\mathrm{Grad}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0))\right]^T dN_i(t)+o_p(1) \\
&= \boldsymbol{B_0}+o_p(1).
\end{aligned}
$$

Eq. (A.23) now implies that the second term on the right side of Eq. (A.28) is $\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)+o_p(1)$. Finally, the consistency of $\hat{\boldsymbol{\theta}}$ implies $\boldsymbol{A}_n(\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}})\xrightarrow{P}-\boldsymbol{\Sigma}_{\mathrm{C}}$, establishing Eq. (A.27).

We now compute the second and cross product moments of $\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0)$, $\boldsymbol{U}_{n,j}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, $j=1,3$, given in Eq. (A.27). By applying iterated expectation with conditioning on $\boldsymbol{W}$ we can show that

$$
E\left[\left(\frac{\delta-m(\boldsymbol{W},\boldsymbol{\theta}_0)}{m(\boldsymbol{W},\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right)^2\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W},\boldsymbol{\theta}_0))\mathrm{Grad}^T(m(\boldsymbol{W},\boldsymbol{\theta}_0))\right]=\boldsymbol{I_0}. \quad \text{(A.31)}
$$

It follows from Eq. (A.31) that

$$
E\left[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,1}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right] = \boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{I_0}\left(\boldsymbol{B_0}\boldsymbol{I_0}^{-1}\right)^T = \boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{B_0}^T \quad \text{(A.32)}
$$

From Andersen and Gill [18] and Eq. (A.7), we have that

$$
E\left[\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0)\boldsymbol{U}_{n,2}^T(\boldsymbol{\beta}_0)\right] = \int_0^{\tau}\boldsymbol{v}(\boldsymbol{\beta}_0,t)s^{(0)}(\boldsymbol{\beta}_0,t)\lambda_0(t)dt = \boldsymbol{\Sigma}_{\mathrm{C}}. \quad \text{(A.33)}
$$

Again, applying iterated expectation with conditioning on X and $\boldsymbol{Z}$, it follows that

$$\begin{aligned}
& E\left[\boldsymbol{U}_{n,3}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,3}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right] \\
= \;& E\left[(\delta - m(X,\boldsymbol{Z},\boldsymbol{\theta}_0))^2(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,X))^{\otimes 2} I(X\le\tau)\right] \\
= \;& E\left[m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)\bar{m}(X,\boldsymbol{Z},\boldsymbol{\theta}_0)(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,X))^{\otimes 2} I(X\le\tau)\right] \\
= \;& \boldsymbol{\Sigma}_{\mathrm{C}} - \boldsymbol{V_0}, \qquad\qquad (A.34)
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\mathrm{C}}$ and $\boldsymbol{V_0}$ are defined by Eq. (A.9) and Eq. (A.10), respectively. However, note that

$$\begin{aligned}
E\left[\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0)\boldsymbol{U}_{n,3}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right] &= -E\left[\delta(\delta - m(X,\boldsymbol{Z},\boldsymbol{\theta}_0))(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,X))^{\otimes 2} I(X\le\tau)\right] \\
&= -E\left[\boldsymbol{U}_{n,3}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,3}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right]. \qquad (A.35)
\end{aligned}$$

Next observe that $E\left[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,2}^T(\boldsymbol{\beta}_0)\right]$ and $E\left[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,3}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right]$ cancel out:

$$\begin{aligned}
& E\left[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,2}^T(\boldsymbol{\beta}_0)\right] \\
= \;& \boldsymbol{B_0}\boldsymbol{I_0}^{-1}E\left[\left\{\frac{\delta - m(\boldsymbol{W},\boldsymbol{\theta}_0)}{m(\boldsymbol{W},\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W},\boldsymbol{\theta}_0)}\mathrm{Grad}(m(\boldsymbol{W},\boldsymbol{\theta}_0))\right\}\right. \\
& \left.\times\int_0^\tau \{\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t)\}^T\, dM^u(t)\right] \\
= \;& \boldsymbol{B_0}\boldsymbol{I_0}^{-1}E\left[\frac{\delta(\delta - m(\boldsymbol{W},\boldsymbol{\theta}_0))}{m(\boldsymbol{W},\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right. \\
& \left.\times\,\mathrm{Grad}(m(\boldsymbol{W},\boldsymbol{\theta}_0))(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,X))^T I(X\le\tau)\right] \\
= \;& \boldsymbol{B_0}\boldsymbol{I_0}^{-1}E\left[\mathrm{Grad}(m(\boldsymbol{W},\boldsymbol{\theta}_0))(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,X))^T I(X\le\tau)\right],
\end{aligned}$$

with $E\left[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\boldsymbol{U}_{n,3}^T(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right]$ computing to the negative of the last expression above. From Eqs. (A.32)–(A.35), we obtain the asymptotic covariance matrix $\boldsymbol{\Sigma}$ given by Eq. (A.12).

Remark To prove the efficiency of $\hat{\boldsymbol{\beta}}$, see proposition 1, we will need a convenient, if slightly lengthy, form of $\boldsymbol{\Sigma}$. From Eqs. (A.32)–(A.35), this alternate form is readily given by

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}\Bigg[\boldsymbol{B_0}\boldsymbol{I_0}^{-1}\boldsymbol{B_0}^T + \int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}_0, t)s^{(0)}(\boldsymbol{\beta}_0, t)\lambda_0(t)dt \\ &\quad - E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right]\Bigg]\boldsymbol{\Sigma}_{\mathrm{C}}^{-1}. \quad \text{(A.36)}\end{aligned}$$

Extension over $[0, \tau_H)$ Taylor's expansion of $\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ about $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ yields

$$\begin{aligned}\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) &= \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \left(\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{S}_n(\boldsymbol{\beta}^*, \boldsymbol{\theta}_0)\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad + \left(\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \qquad \text{(A.37)}\end{aligned}$$

To extend the above proof over $[0, \tau_H)$, we need to show that the tail part of the integrals in the definitions of (i) $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, (ii) $n^{-1}\partial \boldsymbol{S}_n(\boldsymbol{\beta}^*, \boldsymbol{\theta}_0)/\partial \boldsymbol{\beta}$ and (iii) $n^{-1}\partial \boldsymbol{S}_n(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)/\partial \boldsymbol{\theta}$ are each negligible. To prove (i), it remains to show that the tail part of the integral in the definition of $S_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, namely [see also Eq. (2.4)]

$$R_{n,1}(\tau, \tau_H; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \equiv n^{-1/2}\sum_{i=1}^n \int_\tau^{\tau_H} m(t, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)\left[\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t)\right] dN_i(t),$$

is negligible. By the method of proof given in page 308 of Fleming and Harrington [16], we show that, for any $\epsilon > 0$ and $\delta > 0$, $P\left\{\sup_{\tau \leq s \leq \tau_H} |R_{n,1}(\tau, s; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)| > \epsilon\right\}$ is bounded above by

$$\frac{\delta}{\epsilon^2} + \frac{1}{\delta}E\left(\frac{1}{n}\sum_{i=1}^n \int_\tau^{\tau_H} \left(\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t)\right)^2 Y_i(t)\exp(\boldsymbol{\beta}_0^T \boldsymbol{Z}_i) d\Lambda_0(t)\right). \qquad \text{(A.38)}$$

The rest of the proof follows *exactly* as in pages 308 and 309 of Fleming and Harrington [16]. From Eq. (A.3), it follows that

$$R_{n,1}(\tau, \tau_H; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = n^{-1/2}\sum_{i=1}^n \int_\tau^{\tau_H} m(t, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)\left[\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t)\right] dM_i(t).$$

We assume $p = 1$ for simplicity. By the extension of Lenglart's inequality indicated in corollary 3.4.1 of Fleming and Harrington (1991), and for any $\epsilon > 0$ and $\delta > 0$, we have

$$P\left\{\sup_{\tau\leq s\leq\tau_H}|R_{n,1}(\tau,s;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)|>\epsilon\right\} \leq \frac{\delta}{\epsilon^2}+P\left(\tilde{R}_{n,1}(\tau,\tau_H,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)>\delta\right), \quad (A.39)$$

where, using Eq. (A.5) and condition **A.6**, we obtain

$$\begin{aligned}
&\tilde{R}_{n,1}(\tau,\tau_H,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\\
&\equiv \frac{1}{n}\sum_{i=1}^{n}\int_{\tau}^{\tau_H} m^2(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\left[\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0,t)\right]^2 d\langle M_i,M_i\rangle(t)\\
&=\frac{1}{n}\sum_{i=1}^{n}\int_{\tau}^{\tau_H} m(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\left[\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0,t)\right]^2 Y_i(t)\exp(\boldsymbol{\beta}_0^T\boldsymbol{Z}_i)\lambda_0(t)dt.
\end{aligned}$$

Applying Markov's inequality, the second term on the right hand side of Eq. (A.39) is bounded above by the second term on the right hand side of Eq. (A.38).

To prove (ii), for example, we have

$$\begin{aligned}
&R_{n,2}(\tau,\tau_H,\boldsymbol{\beta},\boldsymbol{\theta}_0)\\
&=\frac{1}{n}\sum_{i=1}^{n}\int_{\tau}^{\tau_H} m(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\left\{\left(\frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}-\frac{\boldsymbol{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right\}dN_i(t).
\end{aligned}$$

To prove that $\sup_{\boldsymbol{\beta}\in\mathcal{B}}|R_{n,2}(\tau,\tau_H,\boldsymbol{\beta},\boldsymbol{\theta}_0)|$ is negligible, we will need to extend condition **A**.2 over all of $[0,\tau_H)$. Then, uniformly for $\boldsymbol{\beta}\in\mathcal{B}$, we have

$$\begin{aligned}
&R_{n,2}(\tau,\tau_H,\boldsymbol{\beta},\boldsymbol{\theta}_0)\\
&=\frac{1}{n}\sum_{i=1}^{n}\int_{\tau}^{\tau_H} m(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\left\{\left(\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}-\frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right\}dN_i(t)+o_p(1),
\end{aligned}$$

so that

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathcal{B}}|R_{n,2}(\tau,\tau_H,\boldsymbol{\beta},\boldsymbol{\theta}_0)|\\
&\leq \sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau_H]}\left|\left(\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}-\frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right|\frac{1}{n}\sum_{i=1}^{n}\int_{\tau}^{\tau_H}dN_i(t)+o_p(1)\\
&= \sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau_H]}\left|\left(\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}-\frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right|(H(\tau_H)-H(\tau))+o_p(1),
\end{aligned}$$

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

A.1.3 Weak Convergence of Baseline Cumulative Hazard Estimator

Recalling $\alpha(t, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)$ defined by Eq. (A.17), we introduce the quantities

$$\begin{aligned} L_{n,1}(t, \boldsymbol{\beta}_0) &= n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \alpha(t, s, \boldsymbol{Z}_i, \boldsymbol{\beta}_0) dM_i^u(s), \\ L_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= -n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} (\delta_i - m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))\, \alpha(t, s, \boldsymbol{Z}_i, \boldsymbol{\beta}_0) dN_i(s), \\ L_{n,3}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i - m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)}{m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i, \boldsymbol{\theta}_0)} \left[[\boldsymbol{d}_0(t)]^T \boldsymbol{I_0}^{-1} \mathrm{Grad}(m(\boldsymbol{W}_i, \boldsymbol{\theta}_0) \right], \end{aligned}$$

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

$$n^{1/2}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t)) = L_{n,1}(t, \boldsymbol{\beta}_0) + \sum_{j=2}^{3} L_{n,j}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1). \tag{A.40}$$

Let $\mathrm{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}, \boldsymbol{\theta}))$ and $\mathrm{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}, \boldsymbol{\theta}))$ be the vector 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,

$$\begin{aligned} \hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t) &= \left[\hat{\Lambda}_0(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \Lambda_0(t)\right] + \left\langle \mathrm{Grad}_{\boldsymbol{\beta}}\left(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0)\right), \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right\rangle \\ &\quad + \left\langle \mathrm{Grad}_{\boldsymbol{\theta}}\left(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)\right), \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\rangle. \end{aligned} \tag{A.41}$$

First note that $\mathrm{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \hat{\boldsymbol{\theta}})) = -\boldsymbol{C_0}(t) + o_p(1)$, uniformly over $[0, \tau]$. Indeed,

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

Since $\hat{\boldsymbol{\beta}}$ is consistent, we can apply conditions **A.2** and **A.3** to replace the quantity $\boldsymbol{S}^{(1)}(\boldsymbol{\beta}^*, s)/(S^{(0)}(\boldsymbol{\beta}^*, s))^2$ in the integrand above with its limit $\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, s)/(s^{(0)}(\boldsymbol{\beta}_0, s))^2$ 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.3) that, uniformly for $t \in [0, \tau]$,

$$\mathrm{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \hat{\boldsymbol{\theta}})) = -\int_0^t \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, s)}{s^{(0)}(\boldsymbol{\beta}_0, s)} \lambda_0(s) ds + o_p(1) = -\boldsymbol{C_0}(t) + o_p(1).$$

Likewise, we can show that $\mathrm{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}_0(t,\hat{\boldsymbol{\beta}},\boldsymbol{\theta}^*)) = \boldsymbol{D_0}(t) + o_p(1)$ uniformly for $t \in [0,\tau]$. For the first term on the right side of Eq. (A.41), 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 Fleming and Harrington [16]. Then

$$\hat{\Lambda}_0(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - \Lambda_0(t) = \left[\frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dN_i^u(s)}{S^{(0)}(\boldsymbol{\beta}_0,s)} - \Lambda_0^*(t)\right] - \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{\delta_i - m(s,\boldsymbol{Z}_i,\boldsymbol{\theta}_0)}{S^{(0)}(\boldsymbol{\beta}_0,s)} dN_i(s) + o_p(n^{-1/2}) \quad \text{(A.42)}$$

The first term on the right side of Eq. (A.42) is the sum of $E_{n,1}(t,\boldsymbol{\beta}_0)$ given below and a remainder term, which, by Lenglart's inequality (see, for example, page 308 of Fleming and Harrington [16], where an extension of their corollary 3.4.1 is employed), is $o_p(n^{-1/2})$, uniformly for $t \in [0,\tau]$:

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

The second term on the right side of Eq. (A.42) is the sum of $E_{n,2}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ below and a remainder term, which, by lemma 2 of Gilbert et al. [42], is $o_p(n^{-1/2})$, uniformly over $[0,\tau]$:

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

From Eq. (A.27), the second term on the right side of Eq. (A.41) 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.41) gives $L_{n,3}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$. Thus Eq. (A.40) holds. The multivariate central limit theorem now implies finite dimensional convergence.

It remains to prove tightness. First, $L_{n,1}(t,\boldsymbol{\beta}_0)$ converges weakly in $D[0,\tau]$ to a zero-mean Gaussian martingale and is tight. Next, $L_{n,2}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ can be decomposed into two sums and we show tightness of the first sum as follows: For $0 \le t_1 \le t_2 \le \tau$

and $i = 1, \ldots, n$, we write

$$\chi_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2) = (\delta_i - m(X_i, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))I(t_1 < X_i \leq t_2)/s^{(0)}(\boldsymbol{\beta}_0, X_i)$$

and note that, conditioning by X and $\boldsymbol{Z}$, $E(\chi_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)) = 0$. Then we have that

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} E\left[n^{-1/2}\sum_{i=1}^{n}\chi_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right]^4 \\
&= \overline{\lim_{n\to\infty}} E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}\chi_i^4(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2) + 3\sum_{i\neq j=1}^{n}\chi_i^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\chi_j^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right)\right] \\
&\leq \overline{\lim_{n\to\infty}} E\left[\frac{1}{n^2}\left(3\sum_{i=1}^{n}\chi_i^4(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2) + 3\sum_{i\neq j=1}^{n}\chi_i^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\chi_j^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right)\right] \\
&= 3\,\overline{\lim_{n\to\infty}} E\left[\frac{1}{n}\sum_{i=1}^{n}\chi_i^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right]^2 \\
&\leq 3E\left[\overline{\lim_{n\to\infty}}\frac{1}{n}\sum_{i=1}^{n}\chi_i^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right]^2 \text{ (see problem 3.2.6, page 47 of Chung, 2001)} \\
&= 3\left[E\left(\chi^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, t_1, t_2)\right)\right]^2 \;=\; 3(\mu(t_2) - \mu(t_1))^2,
\end{aligned}$$

where $\mu(t) = \int_0^t dG^1(s, \boldsymbol{z})/(s^{(0)}(\boldsymbol{\beta}_0, s))^2$; see proof of proposition 1 where $G^1(t, \boldsymbol{z})$ was defined. Note that $\mu(\cdot)$ is a finite and continuous measure allowing us to appeal to formula (30) on page 52 of Shorack and Wellner [44]. Tightness of the second sum of $L_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ as well as that of $L_{n,3}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ follow likewise. Finite dimensional convergence and tightness of $n^{1/2}(\hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(\cdot))$ implies its weak convergence in $D[0, \tau]$.

Next we will compute all the cross terms to evaluate the covariance function for the $n^{1/2}(\hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(\cdot))$. We shall suppress $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ from the $L_{n,j}, j = 1, 2, 3$.

Using Eq. (A.4), for $0 \leq t_1, t_2 \leq \tau$,

$$\begin{aligned}
& E(L_{n,1}(t_1)L_{n,1}(t_2)) \\
= \;& E\int_0^\tau \left(\alpha(t_1, s, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, s, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) Y(s) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}} \lambda_0(s) ds \\
= \;& E\int_0^\tau \left(\alpha(t_1, s, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, s, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) dN^u(s) \\
= \;& E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\left(\alpha(t_1, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) I(X \leq \tau)\right]. \qquad \text{(A.43)}
\end{aligned}$$

In the intermediate calculations here, we write $V(\delta|X, \boldsymbol{Z}) = m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)$ for the conditional variance of δ given X and $\boldsymbol{Z}$. Since $E(\delta|X, \boldsymbol{Z}) = m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)$ we have that

$$\begin{aligned}
& E(L_{n,2}(t_1)L_{n,2}(t_2)) \\
= \;& E\left[(\delta - m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))^2 \left(\alpha(t_1, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) I(X \leq \tau)\right] \\
= \;& E\left[V(\delta|X, \boldsymbol{Z})\left(\alpha(t_1, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) I(X \leq \tau)\right]. \qquad \text{(A.44)}
\end{aligned}$$

Furthermore, applying Eq. (A.31), it follows that

$$\begin{aligned}
E(L_{n,3}(t_1)L_{n,3}(t_2)) &= [\boldsymbol{d}_0(t_1)]^T \boldsymbol{I_0}^{-1} \boldsymbol{I_0} \boldsymbol{I_0}^{-1} [\boldsymbol{d}_0(t_2)] \\
&= [\boldsymbol{d}_0(t_1)]^T \boldsymbol{I_0}^{-1} [\boldsymbol{d}_0(t_2)]. \qquad \text{(A.45)}
\end{aligned}$$

The cross-product moment computations are similar. Using Eq. (A.4) we again obtain

$$\begin{aligned}
& E(L_{n,1}(t_1)L_{n,2}(t_2)) \\
= \;& -E\left[\delta(\delta - m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))\left(\alpha(t_1, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)\left(\alpha(t_2, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right) I(X \leq \tau)\right] \\
= \;& -E(L_{n,2}(t_1)L_{n,2}(t_2)) \\
= \;& -E(L_{n,1}(t_2)L_{n,2}(t_1)). \qquad \text{(A.46)}
\end{aligned}$$

Finally, analogous calculations show that

$$\begin{aligned}
E(L_{n,1}(t_1)L_{n,3}(t_2)) &= -E(L_{n,2}(t_1)L_{n,3}(t_3)) \\
E(L_{n,1}(t_2)L_{n,3}(t_1)) &= -E(L_{n,2}(t_2)L_{n,3}(t_1)). \qquad \text{(A.47)}
\end{aligned}$$

From Eqs. (A.43)–(A.47), we obtain the asymptotic covariance function $\sigma(t_1, t_2)$ given by Eq. (A.19). When $t_1 = t_2 = t$, we have the variance function $\sigma(t,t)$ given by Eq. (A.20).

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

$$\begin{aligned} \sigma(t,t) &= \int_0^t \frac{\lambda_0(s)}{s^{(0)}(\boldsymbol{\beta}_0, s)} ds + [\boldsymbol{C}_\mathbf{0}(t)]^T \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \boldsymbol{C}_\mathbf{0}(t) + [\boldsymbol{d}_0(t)]^T \boldsymbol{I}_\mathbf{0}{}^{-1} \boldsymbol{d}_0(t) \\ &\quad - E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) \bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) \left(\alpha(t, X, \boldsymbol{Z}, \boldsymbol{\beta}_0)\right)^2 I(X \le \tau)\right]. \quad \text{(A.48)} \end{aligned}$$

Note that, when $t_1 = t_2 = t$, the third term of Eq. (A.48) is just the right side of Eq. (A.45). Similarly, the fourth term of Eq. (A.48) is just the sum of the right side of Eq. (A.44) and two times the right side of Eq. (A.46). It remains to show that the first two terms of Eq. (A.48) are contributed by the right side of Eq. (A.43). Accordingly for $t_1 = t_2 = t$, write the latter quantity as $T_1 + T_2 + T_3$ and note that

$$\begin{aligned} T_1 &= E\left(\frac{m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)}{(s^{(0)}(\boldsymbol{\beta}_0, X))^2} I(X \le t)\right) = E \int_0^t \frac{1}{(s^{(0)}(\boldsymbol{\beta}_0, s))^2} dN^u(s) \\ &= E \int_0^t \frac{1}{(s^{(0)}(\boldsymbol{\beta}_0, s))^2} Y(s) e^{\boldsymbol{\beta}_0^T Z} \lambda_0(s) ds = \int_0^t \frac{\lambda_0(s)}{s^{(0)}(\boldsymbol{\beta}_0, s)} ds, \end{aligned}$$

which is the first term on the right side of Eq. (A.48). Furthermore, using Eq. (A.9) we have

$$\begin{aligned} T_3 &= E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) [\boldsymbol{C}_\mathbf{0}(t)]^T \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} (\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \boldsymbol{C}_\mathbf{0}(t) I(X \le \tau)\right] \\ &= [\boldsymbol{C}_\mathbf{0}(t)]^T \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} E\left[m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) (\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right] \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \boldsymbol{C}_\mathbf{0}(t) \\ &= [\boldsymbol{C}_\mathbf{0}(t)]^T \boldsymbol{\Sigma}_{\mathrm{C}}^{-1} \boldsymbol{C}_\mathbf{0}(t), \end{aligned}$$

which is the second term on the right side of Eq. (A.48). Finally, the term T_2 is zero as follows:

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

But $E\left[(\boldsymbol{Z}-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,s))Y(s)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right] = \boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0,s)-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,s)s^{(0)}(\boldsymbol{\beta}_0,s) = 0$ and hence, $T_2 = 0$.

A.2 Proof of Theorem 2

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

Defining the normalized sums

$$\begin{aligned} \tilde{\boldsymbol{U}}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= n^{-1/2}\sum_{i=1}^n\boldsymbol{B}_0\tilde{\boldsymbol{I}}_{\mathbf{0}}^{-1}\frac{\xi_i(\delta_i-m(\boldsymbol{W}_i,\boldsymbol{\theta}_0))}{m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}\mathrm{Grad}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0))\\ \tilde{\boldsymbol{U}}_{n,2}(\boldsymbol{\beta}_0) &= n^{-1/2}\sum_{i=1}^n\int_0^\tau\xi_i\left[\boldsymbol{Z}_i-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t)\right]dM_i^u(t)\\ \tilde{\boldsymbol{U}}_{n,3}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= -n^{-1/2}\sum_{i=1}^n\int_0^\tau\xi_i\left[(\delta_i-m(t,\boldsymbol{Z}_i,\boldsymbol{\theta}_0))(\boldsymbol{Z}_i-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t))\right]dN_i(t)\\ \tilde{\boldsymbol{U}}_{n,4}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) &= n^{-1/2}\sum_{i=1}^n\int_0^\tau(1-\xi_i)\left[m(t,\boldsymbol{Z},\boldsymbol{\theta}_0)(\boldsymbol{Z}_i-\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0,t))\right]dM_i(t), \end{aligned}$$

we develop an asymptotic representation given by

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) &= \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}\left[\tilde{\boldsymbol{U}}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)+\tilde{\boldsymbol{U}}_{n,2}(\boldsymbol{\beta}_0)+\sum_{j=3}^4\tilde{\boldsymbol{U}}_{n,j}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right]\\ &\quad + o_p(1). \end{aligned} \tag{A.49}$$

Basically, we will require approximations for $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$, see Eq. (A.28). Under MAR, we have the following representation for the MLE of $\boldsymbol{\theta}$ (cf.

Subramanian [14], or Tsiatis et. al. [39]):

$$
\begin{aligned}
n^{1/2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) &= n^{-1/2}\sum_{i=1}^{n}\tilde{\boldsymbol{I}}_{\mathbf{0}}^{-1}\frac{\xi_i(\delta_i-m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)))}{m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}\mathrm{Grad}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)) \\
&\quad + o_p(1). \qquad (A.50)
\end{aligned}
$$

From Eq. (A.50), it follows that the second term on the right side of Eq. (A.28) produces $\tilde{\boldsymbol{U}}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$. Next write $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)=\boldsymbol{I}_1+\boldsymbol{I}_2$, where

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

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

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

It is straightforward to invoke lemma 2 of Gilbert et al. [42] and show that $\boldsymbol{I}_{12}=o_p(1)$. Let $n\hat{\rho}=\sum_{i=1}^n\xi_i$. Using Eq. (A.4), we obtain

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

It is straightforward to show that

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

The first term of $\boldsymbol{I}_2$ is $\tilde{\boldsymbol{U}}_{n,4}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ and the second term cancels out the second term of $\boldsymbol{I}_{11}$. Eq. (A.49) holds. Asymptotic normality follows by the multivariate central limit theorem.

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

$$E\left[\left(\frac{\xi(\delta - m(\boldsymbol{W}, \boldsymbol{\theta}_0))}{m(\boldsymbol{W}, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}, \boldsymbol{\theta}_0)}\right)^2 \text{Grad}(m(\boldsymbol{W}, \boldsymbol{\theta}_0))\text{Grad}^T(m(\boldsymbol{W}, \boldsymbol{\theta}_0))\right] = \tilde{\boldsymbol{I}}_{\boldsymbol{0}}. \tag{A.51}$$

In the moment calculations below, we shall suppress $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$ appearing in the $\tilde{\boldsymbol{U}}_{n,j}$. Martingale integrals in our expectation calculations contribute 0. We will often utilize Eq. (A.4).

$$\begin{aligned}
E\left[\tilde{\boldsymbol{U}}_{n,1}^{\otimes 2}\right] &= \boldsymbol{B}_{\boldsymbol{0}}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}\left(\boldsymbol{B}_{\boldsymbol{0}}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\right)^T = \boldsymbol{B}_{\boldsymbol{0}}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\boldsymbol{B}_{\boldsymbol{0}}^T, \\
E\left[\tilde{\boldsymbol{U}}_{n,2}^{\otimes 2}\right] &= E\left[\int_0^\tau \pi(t, \boldsymbol{Z})(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} Y(t)\lambda_0(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}}dt\right] \\
&= E\left[\int_0^\tau \pi(t, \boldsymbol{Z})(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} d\left\{N^u(t) - M^u(t)\right\}\right] \\
&= E\left[\pi(X, \boldsymbol{Z})m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right], \\
E\left[\tilde{\boldsymbol{U}}_{n,3}^{\otimes 2}\right] &= E\left[\xi(\delta - m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))^2(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right] \\
&= E\left[\pi(X, \boldsymbol{Z})m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)\bar{m}(X, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \le \tau)\right], \\
E\left[\tilde{\boldsymbol{U}}_{n,4}^{\otimes 2}\right] &= E\left[(1-\xi)(m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0))^2(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} d\langle M, M\rangle(t)\right] \\
&= E\left[\int_0^\tau (1 - \pi(t, \boldsymbol{Z}))m(t, \boldsymbol{Z}, \boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t))^{\otimes 2} Y(t)\lambda_0(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}}dt\right].
\end{aligned}$$

Using Eq. (A.4), it follows that

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

Next, noting that δ are MAR and applying Eq. (A.4), we obtain

$$\begin{aligned} & E\left[\tilde{\boldsymbol{U}}_{n,1}\tilde{\boldsymbol{U}}_{n,2}^T\right] \\ = \; & \boldsymbol{B_0}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1} E\left[\frac{\xi\delta(\delta - m(\boldsymbol{W},\boldsymbol{\theta}_0))}{m(\boldsymbol{W},\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W},\boldsymbol{\theta}_0)}\text{Grad}(m(\boldsymbol{W},\boldsymbol{\theta}_0))(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^T I(X \leq \tau)\right] \\ = \; & \boldsymbol{B_0}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1} E\left[\pi(X,\boldsymbol{Z})\text{Grad}(m(\boldsymbol{W},\boldsymbol{\theta}_0))(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^T I(X \leq \tau)\right] \end{aligned}$$

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

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

Also, $E(\tilde{\boldsymbol{U}}_{n,4}\tilde{\boldsymbol{U}}_{n,j}^T) = 0$, $j = 1, 2, 3$. 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

$$\begin{aligned} \boldsymbol{\Sigma}_{\text{M}} = \; & \boldsymbol{\Sigma}_{\text{C}}^{-1}\Big[\boldsymbol{B_0}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\boldsymbol{B_0}^T + E\left\{\pi(X,\boldsymbol{Z})m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right\} \\ & + E\left\{(1 - \pi(X,\boldsymbol{Z}))m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)^2(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right\} \\ & - \left\{\pi(X,\boldsymbol{Z})m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)\bar{m}(X,\boldsymbol{Z},\boldsymbol{\theta}_0)(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, X))^{\otimes 2} I(X \leq \tau)\right\}\Big]\boldsymbol{\Sigma}_{\text{C}}^{-1}. \end{aligned} \tag{A.52}$$

Combining the last three terms inside the square brackets on the right hand side of Eq. (A.52), the asymptotic covariance matrix simplifies yielding Eq. (A.13).

A.2.2 Weak Convergence of Baseline Cumulative Hazard

We introduce the quantities

$$
\begin{aligned}
\tilde{L}_{n,1}(t, \boldsymbol{\beta}_0) &= n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \xi_i \, \alpha(t, s, \boldsymbol{Z}_i, \boldsymbol{\beta}_0) dM_i^u(s), \\
\tilde{L}_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= -n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \xi_i \, (\delta_i - m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)) \, \alpha(t, s, \boldsymbol{Z}_i, \boldsymbol{\beta}_0) dN_i(s), \\
\tilde{L}_{n,3}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i(\delta_i - m(\boldsymbol{W}_i, \boldsymbol{\theta}_0))}{m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i, \boldsymbol{\theta}_0)} \left[[\boldsymbol{d}_0(t)]^T \tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1} \text{Grad}(m(\boldsymbol{W}_i, \boldsymbol{\theta}_0) \right], \\
\tilde{L}_{n,4}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} (1 - \xi_i) m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0) \, \alpha(t, s, \boldsymbol{Z}_i, \boldsymbol{\beta}_0) dM_i(s),
\end{aligned}
$$

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

$$
n^{1/2}(\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda_0(t)) = \tilde{L}_{n,1}(t, \boldsymbol{\beta}_0) + \sum_{j=2}^{4} \tilde{L}_{n,j}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1). \tag{A.53}
$$

Each $\tilde{L}_{n,k}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0), k = 2, 3$ and $\tilde{L}_{n,1}(t, \boldsymbol{\beta}_0)$ are *complete case* normalized sum, same as $L_{n,k}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ and $L_{n,1}(t, \boldsymbol{\beta}_0)$ defined in Section A.1.3, but with ξ_i attached to each summand and $\boldsymbol{I}_{\boldsymbol{0}}$ replaced with $\tilde{\boldsymbol{I}}_{\boldsymbol{0}}$. The proof follows the methods described in Section A.1.3. Specifically, Eq. (A.41) applies and it suffices to derive asymptotic representations for the three quantities on its right hand side. Because of Eq. (A.50) and consistency of $\hat{\boldsymbol{\beta}}$, both $\text{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}_0(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0))$ and $\text{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}_0(\text{t}, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*))$ still converge in probability to $-\boldsymbol{C}_{\boldsymbol{0}}(t)$ and $\boldsymbol{D}_{\boldsymbol{0}}(t)$, respectively, uniformly for $t \in [0, \tau]$. The first term on the right side of Eq. (A.41), namely

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

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

$$\begin{aligned}
\tilde{E}_1(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\xi_i\delta_i}{S^{(0)}(\boldsymbol{\beta}_0, s)}dN_i(s), \\
\tilde{E}_2(t) &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\xi_i(\delta_i - m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))}{S^{(0)}(\boldsymbol{\beta}_0, s)}dN_i(s) \\
\tilde{E}_3(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{(1-\xi_i)m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)}{S^{(0)}(\boldsymbol{\beta}_0, s)}dN_i(s).
\end{aligned}$$

Using Eq. (A.4) followed by an application of Lenglart's inequality, $\tilde{E}_1(t)$ can be further decomposed as the following sum plus $o_p(n^{-1/2})$:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\xi_i}{s^{(0)}(\boldsymbol{\beta}_0, s)}dM_i^u(s) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{(\xi_i - \hat{\rho})}{S^{(0)}(\boldsymbol{\beta}_0, s)}Y_i(s)\lambda_0(s)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}ds + \hat{\rho}\Lambda_0(t). \qquad \text{(A.54)}
\end{aligned}$$

Write $\tilde{E}_{11}(t)$ for the first term of Eq. (A.54). Apply lemma 2 of Gilbert et al. [42] to obtain

$$\tilde{E}_2(t) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\xi_i(\delta_i - m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))}{s^{(0)}(\boldsymbol{\beta}_0, s)}dN_i(s) + o_p(n^{-1/2}). \qquad \text{(A.55)}$$

Likewise, we can also show that

$$\begin{aligned}
\tilde{E}_3(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{(1-\xi_i)'m(s, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))}{s^{(0)}(\boldsymbol{\beta}_0, s)}dM_i(s) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\xi_i - \hat{\rho}}{S^{(0)}(\boldsymbol{\beta}_0, s)}Y_i(s)\lambda_0(s)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}ds \\
&\quad + (1-\hat{\rho})\Lambda_0(t) + o_p(n^{-1/2}). \qquad \text{(A.56)}
\end{aligned}$$

Write $\tilde{E}_{31}(t)$ for the first term of Eq. (A.56). Note that the second terms of Eq. (A.54) and Eq. (A.56) cancel out. Because of Eq. (A.49), the second term of Eq. (A.41) contributes four expressions. The *second* expression combined with $n^{1/2}\tilde{E}_{11}(t)$ gives $\tilde{L}_{n,1}(t, \boldsymbol{\beta}_0)$. The *third* expression combined with $n^{1/2}\tilde{E}_2(t)$ gives $\tilde{L}_{n,2}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1)$. The *first* expression combined with the third term on the right side of Eq. (A.41) gives

$\tilde{L}_{n,3}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$. The *fourth* expression combined with $n^{1/2}\tilde{E}_{31}(t)$ gives $\tilde{L}_{n,4}(t,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, and Eq. (A.53) holds. Finite dimensional convergence follows by the multivariate central limit theorem and tightness can be verified as described before.

Write $\alpha_{t_1,t_2}(X,\boldsymbol{Z},\boldsymbol{\beta}_0) = \alpha(t_1,X,\boldsymbol{Z},\boldsymbol{\beta}_0)\alpha(t_2,X,\boldsymbol{Z},\boldsymbol{\beta}_0)$. For the covariance function, analogous to the calculations given in Section A.1.3, the following expressions can be verified.

$$E(\tilde{L}_{n,1}(t_1)\tilde{L}_{n,1}(t_2)) = E\left[\pi(X,\boldsymbol{Z})m(X,\boldsymbol{Z},\boldsymbol{\theta}_0)\alpha_{t_1,t_2}(X,\boldsymbol{Z},\boldsymbol{\beta}_0)I(X\le\tau)\right], \quad \text{(A.57)}$$

$$E(\tilde{L}_{n,2}(t_1)\tilde{L}_{n,2}(t_2)) = E\left[\pi(X,\boldsymbol{Z})V(\delta|X,\boldsymbol{Z})\alpha_{t_1,t_2}(X,\boldsymbol{Z},\boldsymbol{\beta}_0)I(X\le\tau)\right], \quad \text{(A.58)}$$

$$E(\tilde{L}_{n,3}(t_1)\tilde{L}_{n,3}(t_2)) = [\boldsymbol{d}_0(t_1)]^T\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}[\boldsymbol{d}_0(t_2)] = [\boldsymbol{d}_0(t_1)]^T\tilde{\boldsymbol{I}}_{\boldsymbol{0}}^{-1}[\boldsymbol{d}_0(t_2)], \quad \text{(A.59)}$$

$$E(\tilde{L}_{n,4}(t_1)\tilde{L}_{n,4}(t_2)) = E\left[(1-\pi(X,\boldsymbol{Z}))m^2(X,\boldsymbol{Z},\boldsymbol{\theta}_0)\alpha_{t_1,t_2}(X,\boldsymbol{Z},\boldsymbol{\beta}_0)I(X\le\tau)\right] \quad \text{(A.60)}$$

$$E(\tilde{L}_{n,1}(t_1)\tilde{L}_{n,2}(t_2)) = E(\tilde{L}_{n,1}(t_2)\tilde{L}_{n,2}(t_1)) = -E(\tilde{L}_{n,2}(t_1)\tilde{L}_{n,2}(t_2)), \quad \text{(A.61)}$$

$$E(\tilde{L}_{n,1}(t_1)\tilde{L}_{n,3}(t_2)) = -E(\tilde{L}_{n,2}(t_1)\tilde{L}_{n,3}(t_2)), \quad \text{(A.62)}$$

$$E(\tilde{L}_{n,1}(t_2)\tilde{L}_{n,3}(t_1)) = -E(\tilde{L}_{n,2}(t_2)\tilde{L}_{n,3}(t_1)). \quad \text{(A.63)}$$

Since $\tilde{L}_{n,4}(\cdot,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ is orthogonal to $\tilde{L}_{n,j}(\cdot,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0), j = 1,2,3,$ we obtain from Eqs. (A.57)–(A.63) the final expression of the limiting covariance function $\sigma_M(t_1,t_2)$ given by Eq. (A.21).

APPENDIX B

ASYMPTOTIC JUSTIFICATION OF GMB

We assume all the regularity conditions given in the Appendix A. In addition, we will assume that $\hat{\boldsymbol{\theta}}$ is strongly consistent, see theorem 2.1 and corollary 2.2 of Dikta [13]. We will also need a strengthening of condition **A.2** in the first project:

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

$$\sup_{\boldsymbol{\beta}\in\mathcal{B},\ t\in[0,\tau_H)} ||\boldsymbol{S}^j(\boldsymbol{\beta}, t) - \boldsymbol{s}^j(\boldsymbol{\beta}, t)|| \;=\; o(1) \text{ a.s.}$$

In Subsection A.1.2 we obtain an alternate representation for $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ needed for the proofs. Note that Eq. A.27 gave an asymptotic representation for $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ that was helpful for proving efficiency in their proposition 1. Here, however, we derive an alternate, simpler, representation that we will need in our proofs, namely

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{\Sigma}_{\mathrm{C}}^{-1}[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)] + o_p(1), \tag{B.1}$$

where $\boldsymbol{\Sigma}_{\mathrm{C}}$ is the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{C}} - \boldsymbol{\beta}_0)$ and

$$\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \;=\; n^{-1/2}\sum_{i=1}^{n} \frac{\delta_i - m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)}{m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i, \boldsymbol{\theta}_0)} \boldsymbol{B_0}\boldsymbol{I_0}^{-1}\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i, \boldsymbol{\theta}_0)), \tag{B.2}$$

$$\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \;=\; n^{-1/2}\sum_{i=1}^{n} \int_0^{\tau_H} m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)(\boldsymbol{Z}_i - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, u))dM_i(u). \tag{B.3}$$

Note that $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$ is the Cox partial likelihood estimator of $\boldsymbol{\beta}$ and $\boldsymbol{B_0}$ is given by Eq. (B.4) below. To derive representation (B.1), it suffices to show from Eq.(A.28) that $\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(n^{1/2})$. From Eqs.(3.2) and (A.3), we have

$$\begin{aligned}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \;=\;& \sum_{i=1}^{n} \int_0^{\tau_H} m(t, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)(\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t))dM_i(t) \\ &+ \sum_{i=1}^{n} \int_0^{\tau_H} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, t))Y_i(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}\lambda_0(t)dt.\end{aligned}$$

The first term equals $\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + o_p(1)$, the negligibility of the remainder term following from an application of Lemma 2 of Gilbert et al. [42]; see the calculation of $\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ in Section A.1.2 for a similar calculation. The second term is algebraically 0.

For facilitating our derivations, we will also need the following quantities:

$$\begin{aligned}
\boldsymbol{B_0} &= E\left[\int_0^{\tau_H} [\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, t)][\text{Grad}(m(\boldsymbol{W}, \boldsymbol{\theta}_0))]^T dN(t)\right], &\text{(B.4)}\\
\boldsymbol{E}_0(t, \boldsymbol{\beta}_0) &= E\left[\int_0^{t} \frac{\exp(\boldsymbol{\beta}_0^T \boldsymbol{z}_0)}{s^{(0)}(\boldsymbol{\beta}_0, u)}[\boldsymbol{z}_0 - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, u)] m(u, \boldsymbol{Z}, \boldsymbol{\theta}_0) dN(u)\right], &\text{(B.5)}\\
\boldsymbol{F}_0(t, \boldsymbol{\beta}_0) &= E\left[\int_0^{t} \frac{\exp(\boldsymbol{\beta}_0^T \boldsymbol{z}_0)}{s^{(0)}(\boldsymbol{\beta}_0, u)}\text{Grad}_{\boldsymbol{\theta}} m(u, \boldsymbol{Z}, \boldsymbol{\theta}_0) dN(u)\right], &\text{(B.6)}\\
a(t, u) &= \frac{I(X \le t) e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}}{s^{(0)}(\boldsymbol{\beta}_0, u)} + \boldsymbol{E}_0^T(t, \boldsymbol{\beta}_0)\boldsymbol{\Sigma}_{\mathrm{C}}^{-1}(\boldsymbol{Z} - \bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, u)), &\text{(B.7)}\\
\boldsymbol{b}^T(t) &= \boldsymbol{F}_0^T(t, \boldsymbol{\beta}_0) + \boldsymbol{E}_0^T(t, \boldsymbol{\beta}_0)\boldsymbol{\Sigma}_{\mathrm{C}}^{-1}\boldsymbol{B_0}, &\text{(B.8)}\\
\sigma^2(t, \boldsymbol{z}_0) &= E\left[m^2(\boldsymbol{W}, \boldsymbol{\theta}_0) a^2(t, X) I(X < \tau_H)\right] + \boldsymbol{b}^T(t)\boldsymbol{I_0}^{-1}\boldsymbol{b}(t). &\text{(B.9)}
\end{aligned}$$

B.1 Limiting Distribution of the Conditional Cumulative Hazard Function

Henceforth, where convenient, we shall suppress the appearance of $\boldsymbol{z}_0$. For example, we write $\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{z}_0)$ simply as $\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$. Note, therefore, that $\Lambda(t, \boldsymbol{z}_0) \equiv \Lambda(t) = \exp(\boldsymbol{\beta}_0^T \boldsymbol{z}_0)\Lambda_0(t)$, where $\Lambda_0(t)$ is the baseline cumulative hazard function. Let $\boldsymbol{\theta}^*$ ($\boldsymbol{\beta}^*$) denote a value on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ ($\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$). Also, let $\text{Grad}_{\boldsymbol{\beta}}\left(\hat{\Lambda}(t, \boldsymbol{\beta}, \boldsymbol{\theta})\right)$ and $\text{Grad}_{\boldsymbol{\theta}}\left(\hat{\Lambda}(t, \boldsymbol{\beta}, \boldsymbol{\theta})\right)$ denote the vector of partial derivatives of $\hat{\Lambda}(t, \boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, respectively.

Taylor's expansion of $\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$ yields the three terms, denoted for easy reference by T_1, T_2 and T_3, on the right hand side (RHS) of the following equation:

$$\begin{aligned}
\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \Lambda(t) &= \left(\hat{\Lambda}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \Lambda(t)\right) + \left\langle \text{Grad}_{\boldsymbol{\beta}}\left(\hat{\Lambda}(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0)\right), \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right\rangle \\
&\quad + \left\langle \text{Grad}_{\boldsymbol{\theta}}\left(\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)\right), \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\rangle. &\text{(B.10)}
\end{aligned}$$

Let $\Lambda_0^*(t) = \int_0^t I\{\sum_{i=1}^n Y_i(x) > 0\}\lambda_0(x)dx$. From page 300 of Fleming and Harrington [16],

$$\Lambda_0^*(t) - \Lambda_0(t) = o_p(n^{-1/2}), \qquad \text{(B.11)}$$

uniformly for $t \in [0, \tau_H)$. The first term on the RHS of Eq. (B.10) can be expressed as

$$T_1 = \sum_{i=1}^n \int_0^t \frac{m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}}{nS^{(0)}(\boldsymbol{\beta}_0, u)} dN_i(u) - e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}\Lambda_0^*(t) + e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}\left(\Lambda_0^*(t) - \Lambda(t)\right),$$

which, on applying Eq. (A.3) and then Eq. (B.11) twice, equals

$$T_1 = \sum_{i=1}^n \int_0^t \frac{m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}}{nS^{(0)}(\boldsymbol{\beta}_0, u)} dM_i(u) + o_p(n^{-1/2}).$$

Applying Lenglart's inequality to the main term on the RHS of the above equation, we obtain

$$T_1 = \hat{\Lambda}(t, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \Lambda(t) = \sum_{i=1}^n \int_0^t \frac{m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0)e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}}{ns^{(0)}(\boldsymbol{\beta}_0, u)} dM_i(u) + o_p(n^{-1/2}). \qquad \text{(B.12)}$$

By the consistency of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, and the strong law of large numbers, it follows that

$$\begin{aligned} &\mathrm{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0)) \\ = &\frac{1}{n}\sum_{i=1}^n \int_0^t \frac{e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}(\boldsymbol{z}_0 - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}_0, u))}{S^{(0)}(\boldsymbol{\beta}_0, u)} m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0) dN_i(u) + o_p(1), \end{aligned} \qquad \text{(B.13)}$$

$$\begin{aligned} &\mathrm{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)) \\ = &\frac{1}{n}\sum_{i=1}^n \int_0^t \frac{e^{\boldsymbol{\beta}_0^T \boldsymbol{z}_0}\mathrm{Grad}_{\boldsymbol{\theta}}(m(u, \boldsymbol{Z}_i, \boldsymbol{\theta}_0))}{S^{(0)}(\boldsymbol{\beta}_0, u)} dN_i(u) + o_p(1). \end{aligned} \qquad \text{(B.14)}$$

Now apply conditions **A.2** and **A.3**, to conclude from Eqs. (B.13) and (B.14) that

$$\mathrm{Grad}_{\boldsymbol{\beta}}(\hat{\Lambda}(t, \boldsymbol{\beta}^*, \boldsymbol{\theta}_0)) = \boldsymbol{E}_0(t, \boldsymbol{\beta}_0) + o_p(1), \qquad \text{(B.15)}$$

$$\mathrm{Grad}_{\boldsymbol{\theta}}(\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}, \boldsymbol{\theta}^*)) = \boldsymbol{F}_0(t, \boldsymbol{\beta}_0) + o_p(1). \qquad \text{(B.16)}$$

From Eqs. (B.1) and (B.15), it follows that the second term on the RHS of Eq. (B.10) equals

$$T_2 \;=\; \boldsymbol{E}_0^T(t,\boldsymbol{\beta}_0)\boldsymbol{\Sigma}_{\mathrm{C}}^{-1}[\boldsymbol{U}_{n,1}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)+\boldsymbol{U}_{n,2}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]+o_p(n^{-1/2}). \qquad \text{(B.17)}$$

From Eqs. (A.23) and (B.16), it follows that the third term on the RHS of Eq. (B.10) equals

$$\begin{aligned} T_3 \;&=\; \sum_{i=1}^{n}\frac{\delta_i - m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}{m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)\bar{m}(\boldsymbol{W}_i,\boldsymbol{\theta}_0)}\boldsymbol{F}_0^T(t,\boldsymbol{\beta}_0)\boldsymbol{I}_{\boldsymbol{0}}{}^{-1}\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i,\boldsymbol{\theta}_0)) \\ &+\quad o_p(n^{-1/2}). \qquad \text{(B.18)}\end{aligned}$$

Combining the dominant term on the RHS of Eq. (B.12) and the second dominant term on the RHS of Eq. (B.17), and using Eqs. (B.3) and (B.7), we get $L_{n,1}(t)$ given by Eq. (3.5). Likewise, combining the dominant term of Eq. (B.18) and the first dominant term of Eq. (B.17), and using Eqs. (B.2) and (B.8), we get $L_{n,2}(t)$ given by Eq. (3.6).

We now compute the covariance function of the limiting Gaussian process. We have

$$\begin{aligned} & E\left(L_{n,1}(t_1)L_{n,1}(t_2)\right) \\ =\;& E\int_0^{\tau_H} m^2(u,\boldsymbol{Z},\boldsymbol{\theta}_0)a(t_1,u)a(t_2,u)\,d\left\langle M(u),M(u)\right\rangle \\ =\;& E\int_0^{\tau_H} m^2(u,\boldsymbol{Z},\boldsymbol{\theta}_0)a(t_1,u)a(t_2,u)\frac{Y(u)\lambda_0(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}}}{m(u,\boldsymbol{Z},\boldsymbol{\theta}_0)}du \qquad [\text{by Eq. } (A.5)] \\ =\;& E\int_0^{\tau_H} m^2(u,\boldsymbol{Z},\boldsymbol{\theta}_0)a(t_1,u)a(t_2,u)\,d[N(u)-M(u)] \\ =\;& E\left[m^2(\boldsymbol{W},\boldsymbol{\theta}_0)a(t_1,X)a(t_2,X)I(X<\tau_H)\right]. \qquad \text{(B.19)}\end{aligned}$$

It is straightforward to show that

$$E(L_{n,2}(t_1)L_{n,2}(t_2)) \;=\; \boldsymbol{b}^T(t_1)\boldsymbol{I}_0^{-1}\boldsymbol{b}(t_2) \qquad \text{(B.20)}$$

Using iterated conditional expectation with conditioning by $\boldsymbol{W}$, the cross product term is 0:

$$E(L_{n,1}(t_1)L_{n,2}(t_2)) = E(L_{n,1}(t_2)L_{n,2}(t_1)) = 0$$

The final form of the covariance function of the limiting Gaussian process is given by

$$C(t_1,t_2) = E\left[m^2(\boldsymbol{W},\boldsymbol{\theta}_0)a(t_1,X)a(t_2,X)I(X<\tau_H)\right] + \boldsymbol{b}^T(t_1)\boldsymbol{I}_0^{-1}\boldsymbol{b}(t_2). \quad \text{(B.21)}$$

When $t_1 = t_2 = t$ the above reduces to the variance function $V(t)$, given by

$$V(t) = C(t,t) = E\left[m^2(\boldsymbol{W},\boldsymbol{\theta}_0)a^2(t,X)I(X<\tau_H]\right) + \boldsymbol{b}^T(t)\boldsymbol{I}_0^{-1}\boldsymbol{b}(t). \quad \text{(B.22)}$$

B.2 Large-sample Justification of the GMB

Note that $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, the partial likelihood estimator of $\boldsymbol{\beta}$, is strongly consistent, see Tsiatis [17]. Asymptotic validity of the GMB requires strong consistency of $\hat{\boldsymbol{\beta}}$, not proved in the Section A. Here, in Subsection B.2.1, we first show the strong consistency of $\hat{\boldsymbol{\beta}}$. We then show, in Subsection B.2.2, that the GMB is asymptotically valid.

B.2.1 Strong Consistency of $\hat{\boldsymbol{\beta}}$

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

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

Since $\breve{\boldsymbol{l}}_n(\boldsymbol{\beta}_0,\hat{\boldsymbol{\theta}})$ is free of $\boldsymbol{\beta}$, it follows that $\hat{\boldsymbol{\beta}}$ maximizes

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

We first show that $\|\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{l}}(\boldsymbol{\beta},\boldsymbol{\theta}_0)\| = o(1)$ almost surely (a.s.), where

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

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

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

and instead show that the following two equations hold:

$$\|\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)\| = o(1) \quad \text{a.s.,} \tag{B.23}$$

$$\|\tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0) - \tilde{\boldsymbol{l}}(\boldsymbol{\beta},\boldsymbol{\theta}_0)\| = o(1) \quad \text{a.s.} \tag{B.24}$$

First, Eq. (B.24) follows from strong law of large numbers. To prove Eq. (B.23), Taylor's expansion of $\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ yields

$$\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0) = [\boldsymbol{l}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0) - \tilde{\boldsymbol{l}}_n(\boldsymbol{\beta},\boldsymbol{\theta}_0)] + \left(\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{l}_n(\boldsymbol{\beta},\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \tag{B.25}$$

We apply conditions **AA.2** and **A.3** to deduce that the first term of Eq. (B.25) is $o(1)$ a.s. To show that the second term is also $o(1)$ a.s. for each $\boldsymbol{\beta} \in \mathcal{B}$, we shall assume that $p = 1$. Note that condition **D.1** implies that $|\langle \text{Grad}_{\boldsymbol{\theta}}(m(x,\boldsymbol{z},\boldsymbol{\theta}^*)), (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\rangle| \leq kM(x,\boldsymbol{z})||\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0||$. Then,

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

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

Since the random concave function $\boldsymbol{l}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})$ converges pointwise to the concave function $\tilde{\boldsymbol{l}}(\boldsymbol{\beta},\boldsymbol{\theta}_0)$ a.s., theorem 10.8 of Rockafellar [45] guarantees that the convergence a.s. is uniform in $\boldsymbol{\beta} \in [-M, M]$. Also, the limit function $\tilde{\boldsymbol{l}}(\boldsymbol{\beta}, \boldsymbol{\theta}_0)$ has unique (global) maximum at $\boldsymbol{\beta}_0$, see Section A.1.1. We now follow the concluding part of the strong consistency proof given by Tsiatis [17]. Fix a $\delta > 0$ and note that for a δ-neighborhood around $\boldsymbol{\beta}_0$ we have $\tilde{\boldsymbol{l}}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \tilde{\boldsymbol{l}}(\boldsymbol{\beta}, \boldsymbol{\theta}_0) \geq 0$. From the uniform convergence proved above, it follows that $\boldsymbol{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) - \boldsymbol{l}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\theta}})$ converges a.s. to $\tilde{\boldsymbol{l}}(\boldsymbol{\beta}, \boldsymbol{\theta}_0) - \tilde{\boldsymbol{l}}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. This implies that for almost all realizations there exists n_0, depending on the realization, such that for all $n \geq n_0$ any $\boldsymbol{\beta}$ on the boundary of the δ-neighborhood cannot be a local maximum. In turn, since $\boldsymbol{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ is continuous and differentiable over the interval $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \delta$, there must be a local maximum in the interior. That is, $d\,\boldsymbol{l}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})/d\boldsymbol{\beta} = 0$, which is satisfied by $\hat{\boldsymbol{\beta}}$. This argument can be repeated for each shrinking δ to obtain a consistent sequence $\hat{\boldsymbol{\beta}}$ converging a.s. to $\boldsymbol{\beta}_0$.

B.2.2 Asymptotic Justification

Recall from Section 3.2 that $\hat{\mathbb{H}}^*(t) = \hat{L}^*_{n,1}(t) + \hat{L}^*_{n,2}(t)$, where the RHS quantities are defined by Eqs. (3.8) and (3.9). Also, note that $\hat{\boldsymbol{E}}_0(t)$, a strongly consistent estimate of $\boldsymbol{E}_0(t, \boldsymbol{\beta}_0)$, can be obtained by replacing $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ in Eq. (B.13) with $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, respectively. Likewise, $\hat{\boldsymbol{F}}_0(t, \hat{\boldsymbol{\beta}})$ can be obtained from Eq. (B.14). Furthermore, strong consistent estimates of $\boldsymbol{B_0}$ and $\boldsymbol{b}^T(t)$ can be obtained from Eqs. (B.4) and (B.8), respectively. Finally, $\hat{a}_i(t, u)$ will be obtained by replacing $\boldsymbol{\beta}_0$, $s^{(0)}(\boldsymbol{\beta}_0, u)$, $\boldsymbol{E}_0(t, \boldsymbol{\beta}_0)$, $\boldsymbol{\Sigma}_{\mathrm{C}}$ and $\bar{\boldsymbol{z}}(\boldsymbol{\beta}_0, u)$ in Eq. (B.7) with $\hat{\boldsymbol{\beta}}$, $S^{(0)}(\hat{\boldsymbol{\beta}}, u)$, $\hat{\boldsymbol{E}}_0(t)$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{C}}$ and $\bar{\boldsymbol{Z}}(\hat{\boldsymbol{\beta}}, u)$, respectively.

Let $\mathbb{P}_{\mathrm{G}}$, $\mathbb{E}_{\mathrm{G}}$, $\mathrm{Cov}_{\mathrm{G}}$, $\mathrm{Var}_{\mathrm{G}}$ be the probability measure, expectation, covariance, and variance with respect to G, that is, conditioned on the sample $(X_i, \delta_i, \boldsymbol{Z}_i)_{1 \leq i \leq n}$. We have

$$\begin{aligned} \mathrm{Cov}_{\mathrm{G}}(\hat{\mathbb{H}}^*(t_1), \hat{\mathbb{H}}^*(t_2)) &= \mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,1}(t_1)\hat{L}^*_{n,1}(t_2)) + \mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,2}(t_1)\hat{L}^*_{n,2}(t_2)) \\ &\quad + \mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,1}(t_1)\hat{L}^*_{n,2}(t_2)) + \mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,1}(t_2)\hat{L}^*_{n,2}(t_1)), \end{aligned} \quad \text{(B.26)}$$

where

$$
\begin{aligned}
\mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,1}(t_1)\hat{L}^*_{n,1}(t_2)) &= \frac{1}{n}\sum_{i=1}^{n} m^2(X_i,\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}})\hat{a}_i(t_1,X_i)\hat{a}_i(t_2,X_i)I(X_i<\tau_H),\\
\mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,2}(t_1)\hat{L}^*_{n,2}(t_2)) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\delta_i-m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}{m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})\bar{m}(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}\right)^2(\hat{\boldsymbol{b}}^T(t_1)\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}})\\
&\quad\times\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})))(\hat{\boldsymbol{b}}^T(t_2)\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}})\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}}))),\\
\mathbb{E}_{\mathrm{G}}(\hat{L}^*_{n,1}(t_1)\hat{L}^*_{n,2}(t_2)) &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\delta_i-m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}}))m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}{m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})\bar{m}(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}})}\hat{\boldsymbol{b}}^T(t_2)\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}})\\
&\quad\times\,\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i,\hat{\boldsymbol{\theta}}))\hat{a}_i(t_1,X_i)I(X_i<\tau_H).
\end{aligned}
$$

Strong consistency of $\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}}),\hat{\boldsymbol{E}}_0(t),\hat{\boldsymbol{F}}_0(t),\hat{\boldsymbol{B}}_{\boldsymbol{0}}$ and $\hat{\boldsymbol{b}}(t)$, coupled with condition **AA.2** and the strong law of large numbers ensures that, for almost all samples, the first two terms on the RHS of Eq. (B.26) converge to the terms on the RHS of Eqs. (B.19) and (B.20), respectively. The same argument, followed by an application of iterated conditional expectation with conditioning by $\boldsymbol{W}$, also ensures that the two cross-moment terms in Eq. (B.26) are each zero. Therefore, the process $\hat{\mathbb{H}}^*(\cdot)$ has the same limiting covariance structure given by Eq. (B.21).

It remains to show that the process $\hat{\mathbb{H}}^*(\cdot)$ converges weakly to a zero-mean Gaussian process. We shall verify Lindeberg's condition and tightness. Recall that $\hat{\mathbb{H}}^*(t)=L^*_{n,1}(t)+L^*_{n,2}(t)$, where $L^*_{n,1}(t)$ and $L^*_{n,2}(t)$ are given by Eqs. (3.8) and (3.9), respectively. Combining the terms on the RHS of Eqs. (3.8) and (3.9), we can write $\hat{\mathbb{H}}^*(t)=n^{-1/2}\sum_{i=1}^{n}k_i(t)G_i\equiv\sum_{i=1}^{n}H_i^*(t)$. Let $s_n^2:=\mathrm{Var}(\sum_{i=1}^{n}H_i^*(t))$, so that $s_n^2=\sum_{j=1}^{n}k_j^2(t)/n$. To verify Lindeberg's condition we need to show that, for almost all sample sequences,

$$
\frac{1}{ns_n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathrm{G}}\left[k_i^2(t)G_i^2\ I\left(|k_i(t)G_i|\geq n^{1/2}\epsilon s_n\right)\right]\ \longrightarrow\ 0 \text{ as } n\to\infty.
$$

Without loss of generality, we can assume that $s_n^2=1$, so it suffices to prove that

$$
\frac{1}{n}\sum_{i=1}^{n}k_i^2(t)\mathbb{E}_{\mathrm{G}}\left[G^2\ I\left(|k_i(t)G|\geq n^{1/2}\epsilon\right)\right]\ \longrightarrow\ 0 \text{ as } n\to\infty. \tag{B.27}
$$

Since the left side of Eq. (B.27) is bounded between 0 and $\sum_{i=1}^{n} a_i/n$, where

$$a_i = k_i^2(t)\mathbb{E}_{\mathrm{G}}\left[G^2\ I\left(|k_i(t)G| \geq i^{1/2}\epsilon\right)\right],$$

we apply Cesaro means to instead show that $a_n \to 0$ as $n \to \infty$. By problem 2 on page 46 of Chung [46], however, it is enough to show that for almost all sample sequences $\mathbb{P}_{\mathrm{G}}\left(|k_n(t)G| \geq n^{1/2}\epsilon\right) \longrightarrow 0$ as $n \to \infty$. This follows by applying Chebyshev's inequality and the boundedness of $E(k_n^2(t)G^2)$ for all n.

Recall that $\mu_4 = 3\mu_2^2$, where μ_i is the ith central moment of the normal distribution. To verify tightness, apply formula (30) on p. 52 of Shorack and Wellner [44]: For $t_1 < t_2$,

$$\begin{aligned}
\overline{\lim_{n\to\infty}}\,\mathbb{E}_{\mathrm{G}}\left[\hat{\mathbb{H}}^*(t_2) - \hat{\mathbb{H}}^*(t_1)\right]^4 &= \overline{\lim_{n\to\infty}}\,\mathbb{E}_{\mathrm{G}}\left[n^{-1/2}\sum_{i=1}^{n}(k_i(t_2) - k_i(t_1))G_i\right]^4 \\
&= \overline{\lim_{n\to\infty}}\ 3\left[\frac{1}{n}\sum_{i=1}^{n}(k_i(t_2) - k_i(t_1))^2\right]^2 \\
&\leq \overline{\lim_{n\to\infty}}\left[\frac{K}{n}\sum_{i=1}^{n}(k_i(t_2) - k_i(t_1))^2\right]^2. \qquad \text{(B.28)}
\end{aligned}$$

where $K \geq \sqrt{3}$. From Eq. (B.5), it will be convenient to define, for $t_1 < t_2$,

$$\hat{\boldsymbol{E}}_0\left((t_1, t_2]\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\exp(\hat{\boldsymbol{\beta}}^T\boldsymbol{z}_0)(\boldsymbol{z}_0 - \bar{\boldsymbol{Z}}(\hat{\boldsymbol{\beta}}, X_i))}{S^{(0)}(\hat{\boldsymbol{\beta}}, X_i)}m(X_i, \boldsymbol{Z}_i, \hat{\boldsymbol{\theta}})I(t_1 < X_i \leq t_2).$$

The quantities $\hat{\boldsymbol{F}}_0\left((t_1, t_2], \hat{\boldsymbol{\beta}}\right)$, $\hat{\boldsymbol{b}}\left((t_1, t_2]\right)$, $\hat{a}\left((t_1, t_2], u\right)$ are likewise defined as above, based on Eqs. (B.6)–(B.8). It can now be shown that the RHS of inequality (B.28) equals

$$\begin{aligned}
&\frac{K}{n}\sum_{i=1}^{n}(k_i(t_2) - k_i(t_1))^2 \\
= &\frac{K}{n}\sum_{i=1}^{n}\Bigg[m(X_i, \boldsymbol{Z}_i, \hat{\boldsymbol{\theta}})\hat{a}_i\left((t_1, t_2], X_i\right)I(X_i < \tau_H) \\
&+ \frac{\delta_i - m(\boldsymbol{W}_i, \hat{\boldsymbol{\theta}})}{m(\boldsymbol{W}_i, \hat{\boldsymbol{\theta}})\bar{m}(\boldsymbol{W}_i, \hat{\boldsymbol{\theta}})}\hat{\boldsymbol{b}}^T\left([t_1, t_2)\right)\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}})\mathrm{Grad}_{\boldsymbol{\theta}}(m(\boldsymbol{W}_i, \hat{\boldsymbol{\theta}}))\Bigg]^2,
\end{aligned}$$

which equals the RHS of Eq. (B.22), after replacing the first argument of a and $\boldsymbol{b}$ there with $(t_1, t_2]$. Therefore, the LHS of Eq. (B.28) is finite. Tightness is verified.

BIBLIOGRAPHY

[1] M.J. Zhang and P.J. Klein. Confidence bands for the difference of two survival curves under proportional hazards model. *Lifetime Data Analysis*, 7:243–254, 2001.

[2] E.R. Dickson, P.M. Grambsch, T.R. Fleming, L.D. Fisher, and A. Langworthy. Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology*, 10:1–7, 1989.

[3] D. Burr and H. Doss. Confidence bands for the median survival time as a function of the covariates in the cox model. *Journal of the American Statistical Association*, 88:1330–1340, 1993.

[4] D.Y. Lin, T.R. Fleming, and L.J. Wei. Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81:73–81, 1994.

[5] T.L. Lai and Z. Su. Confidence intervals for survival quantiles in the cox regression model. *Lifetime Data Analysis*, 12:407–419, 2006.

[6] D.M. Dabrowska and W.T. Ho. Confidence bands for comparison of transition probabilities in a markov chain model. *Lifetime Data Analysis*, 6:5–21, 2000.

[7] G. Wei and D.E. Schaubel. Estimating cumulative treatment effects in the presence of nonproportional hazards. *Biometrics*, 64:724–732, 2008.

[8] P.B. Gilbert, L.J. Wei, M.R. Kosorok, and J.D. Clemens. Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics*, 58:773–780, 2002.

[9] Y. Shen and S.C. Cheng. Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–1100, 1999.

[10] G. Yin and J. Hu. Two simulation methods for constructing confidence bands under the additive risk model. *Journal of Biopharmaceutical Statistics*, 14:389–402, 2004.

[11] J. Lee and S. Hyun. Confidence bands for the difference of two survival functions under the additive risk model. *Journal of Applied Statistics*, 38:785–797, 2011.

[12] S. Subramanian and P. Zhang. Confidence bands for survival functions under semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 143:219–226, 2013.

[13] G. Dikta. On semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 66:253–279, 1998.

[14] S. Subramanian. The missing censoring-indicator model of random censorship. *Handbook of Statistics Advances in Survival Analysis. Eds. N. Balakrishnan and C.R.Rao.*, 23:123–141, 2004.

[15] D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34:187–220, 1972.

[16] T. R. Fleming and D.P. Harrington. *Counting Processes and Survival analysis.* Wiley, New York, NY, 2005.

[17] A.A. Tsiatis. A large sample study of cox's regression model. *Annals of Statistics*, 9:93–108, 1981.

[18] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10:1100–1120, 1982.

[19] N.E. Breslow. Contribution to the discussion on the paper by dr cox, regression and life tables. *Journal of the Royal Statistical Society: Series B*, 34:216–217, 1972.

[20] W.L. Hall and J.A. Wellner. Confidence bands for a survival curve from censored data. *Biometrika*, 67:133–143, 1980.

[21] V.N. Nair. Confidence bands for survival functions with censored data: A comparative study. *Technometrics*, 26:265–275, 1984.

[22] M.G. Akritas. Bootstrapping the kaplan–meier estimator. *Journal of the American Statistical Association*, 81:1032–1038, 1986.

[23] M. Yuan. Semiparametric censorship model with covariates. *Test*, 14:1–26, 2005.

[24] J. A. Koziol and S. B. Green. A cramer-von mises statistic for randomly censored data. *Biometrika*, 63:465–474, 1976.

[25] S. Subramanian. Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazard. *Journal of Statistical Planning and Inference*, 84:81–94, 2000.

[26] E. L. Lehmann and G. Casella. *Theory of Point Estimation.* Springer, New York, NY, 1998.

[27] M. Siotani. Some applications of loewner's ordering on symmetric matrices. *Annals of the Institute of Statistical Mathematics*, 19:245–259, 1967.

[28] C.R. Rao. *Linear Statistical Inference and Its Applications.* Wiley, New York, NY, 1973.

[29] C. Liu and Q. Wang. Semiparametric estimation for regression coefficients in the cox model with failure indicators missing at random. *Statistica Sinica*, 20:1125–1142, 2010.

[30] L.A. Escobar and W.Q. Meeker. Assessing influence in regression analysis with censored data. *Biometrics*, 48:507–528, 1992.

[31] P.H. Rossi, R.A. Berk, and K.J. Lenihan. *Money, Work and Crime: Some Experimental Results.* Academic Press, San Diego, CA, 1980.

[32] J. Fox. *An R and S Plus Companion to Applied Regression*. Sage Publications Ltd, London, 2002.

[33] D. Collett. *Modelling Binary Data*. CRS Press, Boca Raton, FL, 2002.

[34] G. Dikta. Asymptotically efficient estimation under semi-parametric random censorship models. *Journal of the American Statistical Association*, 101:521–530, 2006.

[35] P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, New York, NY, 1993.

[36] D.Y. Lin, L.J. Wei, and Z. Ying. Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80:557–572, 1993.

[37] J.P. Klein and M.L. Moeschberger. *Survival analysis techniques for censored and truncated data*. Springer, New York, NY, 2003.

[38] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semiparametric models*. Springer, New York, NY, 1993.

[39] A.A. Tsiatis, M. Davidian, and B. Mcneney. Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika*, 89:238–244, 2002.

[40] A. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, NY, 1998.

[41] M.D. Perlman. On the strong consistency of approximate maximum likelihood estimates. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:263281, 1972.

[42] P.B. Gilbert, I.W. McKeague, and Y. Sun. The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy. *Biostatistics*, 9:263–276, 2008.

[43] T.M. Apostol. *Mathematical Analysis*. Addison Wesley, Reading, MA, 1974.

[44] G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley, New York, NY, 1986.

[45] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[46] K.L. Chung. *A Course in Probability Theory*. Academic Press, San Diego, CA, 2001.