Published in final edited form as:

Lifetime Data Anal. 2014 July ; 20(3): 424–443. doi:10.1007/s10985-013-9273-5.

Proportional hazards regression in the presence of missing study eligibility information

Qing Pan and

Department of Statistics, George Washington University, Washington, DC 20052, USA

Douglas E. Schaubel

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Qing Pan: qpan@gwu.edu

Abstract

We consider the study of censored survival times in the situation where the available data consist of both eligible and ineligible subjects, and information distinguishing the two groups is sometimes missing. A complete-case analysis in this context would use only subjects known to be eligible, resulting in inefficient and potentially biased estimators. We propose a two-step procedure which resembles the EM algorithm but is computationally much faster. In the first step, one estimates the conditional expectation of the missing eligibility indicators given the observed data using a logistic regression based on the complete cases (i.e., subjects with non-missing eligibility indicator). In the second step, maximum likelihood estimators are obtained from a weighted Cox proportional hazards model, with the weights being either observed eligibility indicators or estimated conditional expectations thereof. Under ignorable missingness, the estimators from the second step are proven to be consistent and asymptotically normal, with explicit variance estimators. We demonstrate through simulation that the proposed methods perform well for moderate sized samples and are robust in the presence of eligibility indicators that are missing not at random. The proposed procedure is more efficient and more robust than the complete case analysis and, unlike the EM algorithm, does not require time-consuming iteration. Although the proposed methods are applicable generally, they would be most useful for large data sets (e.g., administrative data), for which the computational savings outweigh the price one has to pay for making various approximations in avoiding iteration. We apply the proposed methods to national kidney transplant registry data.

Keywords

Cox model; EM algorithm; Missing at random; Posterior expectation; Study eligibility

1 Introduction

Standard missing data methods usually need to be modified in the presence of censored data. Various methods have been proposed to fit proportional hazards models (Cox 1972) in the presence of missing covariates. For example, Lin and Ying (1993) proposed estimating functions where the expectation of each element of the covariate vector is calculated from the non-missing values of that element. Zhou and Pepe (1995) estimated the expected relative risk function given auxiliary information for missing covariates and plugged the expectation into an induced partial likelihood. Paik and Tsai (1997) proposed mean imputation methods for various different missingness mechanisms. Each of these methods targeted some estimating function that mimicked the partial likelihood score function using all available information. However, these methods are not directly applicable to the setting of interest in this report, in which missingness involves information pertaining to whether or not a subject in the available data set is actually eligible for the particular analysis to be conducted. This setting often arises in disease registry databases, which are usually assembled for purposes other than research. Our goal is to filter out the ‘contaminating’ effect of ineligible subjects, while retaining the contribution of subjects who are in fact eligible.

In the data structure of interest, a complete case analysis would discard all subjects with missing eligibility indicators, leading to smaller sample size and resulting in biased estimation unless the missingness mechanism is missing completely at random (MCAR). The EM algorithm can be used to carry out maximum likelihood estimation. However, it requires iterations and therefore may be computationally inconvenient for very large data sets. We propose a two-step procedure which can be viewed as a simplified version of the EM algorithm. In the first step, study eligibility probability is estimated from a logistic regression based on the complete data. Equivalent to an E step, the logistic regression models the conditional mean of the missing eligibility indicator given the covariates, auxiliary information and the outcome. In the second step, subjects with eligibility indicators either observed or missing are pooled, and a weighted Cox model is fitted, with the weights being either the subject’s actual eligibility indicator (when observed) or the subject’s estimated eligibility probability. Equivalent to an M step, the weighted model produces the maximum likelihood estimator for parameters in the group of eligible subjects. Unlike existing approaches based on the EM algorithm (e.g., Chen and Little 1999), our method does not involve potentially time-consuming iterations. A similar idea was used to handle missing covariates (Herring et al. 2004), where missing covariates are replaced by the conditional expectation of the missing covariate given the observed data. Comparing the method proposed in this manuscript and that by Paik and Tsai (1997), there are two major differences. First, Paik and Tsai (1997) considered the setting where the missing information pertains to the covariates, while our methods consider settings where study eligibility data may be missing. Although these problems are connected, they are definitely different and, therefore, each has its own intricacies. Second, although both methods impute the posterior/conditional mean of the missing data, Paik and Tsai (1997) takes the average of the partially missing covariate in a neighborhood with similar event times and levels of observed covariate, while our methods impute from the complete cases using regression.

The proposed method is motivated by the study of kidney transplant data. With the increasing prevalence of end-stage renal disease, clinicians are increasingly opting for transplanting donor kidneys of lesser quality. In particular, an Expanded Criteria Donor (ECD) is a deceased kidney donor characterized by being either (1) age ≥ 60 , or (2) age 50–59 and with at least two of the following three conditions: hypertension, serum creatinine concentration >1.5 mg/dl, or death due to stroke (Port et al. 2002). In the article by Port et al. (2002), which formalized the ECD concept, ECD transplantation increased the graft failure hazard by 70% compared to standard criterion donor (i.e., non-ECD) transplantation. The term ECD has become a well-accepted label of organ quality in the nephrology community.

It is debatable whether ECD kidney transplantation is an effective treatment option. A covariate-adjusted evaluation of ECD graft failure risk (relative to non-ECD transplantation) could be carried out using a Cox model, with the appropriate time origin being the date of transplant, and the end point defined as graft failure or death, whichever occurs first. The coefficient corresponding to the ECD kidney indicator represents the contrast between ECD and non-ECD transplants. However, not all transplant recipients were eligible to receive an ECD kidney. Specifically, at the time of wait-listing, patients indicate (as part of their candidate registration) whether or not they are willing to accept an ECD transplant. Patients who indicate “no” will never receive ECD offers, let alone receive an ECD transplant. It makes sense, then, to compare ECD versus non-ECD transplants among patients willing to receive ECD kidneys. The rationale for excluding ECD-ineligible patients is that they may have generally better health than patients on the ECD wait list. The inclusion of patients unwilling to even consider an ECD transplant would result in a biased comparison between ECD and non-ECD transplants, even after adjustment for available covariates. However, there was no separate ECD wait list before July 2002. Hence, we are unable to explicitly determine from the data whether a transplant recipient registered before July 2002 is eligible for our analysis.

2 Proposed method

We begin by setting up the necessary notation. For subject i ($i = 1, \dots, n$) with event time T_i and censoring time C_i , the observed time and observed event indicator are $\tilde{T}_i = \min(T_i, C_i)$ and $\tilde{\Delta}_i = I(T_i \leq C_i)$, respectively. We define the at-risk indicator $Y_i(t) = I(\tilde{T}_i \geq t)$ and the event

counting process $N_i(t) = I(\tilde{T}_i \leq t, \tilde{\Delta}_i = 1) = \int_0^t dN_i(s)$, with $dN_i(t) = N_i(t) - N_i(t^-)$. Further, we let ξ_i denote the study eligibility indicator, with $\xi_i = 1$ if subject i is in fact eligible for inclusion in the study sample and $\xi_i = 0$ if ineligible. The data at hand (e.g., from an administrative database) consist of study-eligible subjects and subjects who are ineligible. We let $R_i = 1$ if information on study eligibility is available for subject i , and 0 otherwise. We then set $\mathcal{R}_1 = \{i : R_i = 1\}$ and $\mathcal{R}_0 = \{i : R_i = 0\}$. We assume independent censoring, expressed formally as

$$\lim_{\delta \rightarrow 0} \delta^{-1} Pr(t \leq T_i < t + \delta | T_i \geq t, C_i \geq t, Z_i, \xi_i, R_i) = \lim_{\delta \rightarrow 0} \delta^{-1} Pr(t \leq T_i < t + \delta | T_i \geq t, Z_i, \xi_i, R_i).$$

A Cox proportional hazards model is assumed for the subjects of interest (i.e., those with $\xi_i = 1$),

$$\lambda_i(t) \equiv \lambda_i(t|\xi_i=1, Z_i(t)) = \lambda_0(t) e^{\beta'_0 Z_i(t)}. \quad (1)$$

Note that time-varying covariates are assumed to be external; i.e., with paths fixed or decided by external factors (Kalbfleisch and Prentice 2002). The eligibility indicators ξ_i are not observed in $\mathcal{R}_0$. Two practices are common in the presence of missing information:

1. Complete case analysis, which discards the incomplete data $\mathcal{R}_0$.
2. EM algorithm, which replaces missing ξ_i with its posterior mean. Here, the word “posterior” is used in a general sense and refers to the conditional distribution given the observed information.

An EM algorithm for the above-described set-up would proceed as follows. Another Cox model with different baseline hazards and regression coefficients is assumed for subjects in the ineligible ($\xi_i = 0$) group,

$$\lambda_{i2}(t) \equiv \lambda_i(t|\xi_i=0, Z_i(t)) = \lambda_{02}(t) e^{\beta'_{02} Z_i(t)}. \quad (2)$$

The eligibility indicator ξ_i is assumed to follow a logistic model with covariate vector W_i ,

$$\text{logit}(p_i) = \gamma'_0 W_i, \quad (3)$$

where $p_i = E(\xi_i | W_i) = \text{Pr}(\xi_i = 1 | W_i)$. The covariate vector W_i may contain a subset or all elements of Z_i , as well as additional auxiliary information. The joint likelihood is given by

$$\begin{aligned} L(\tilde{T}, \Delta, \xi, R | Z, W) &= \prod_{i \in \mathcal{R}_1} L_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i=1)^{\xi_i} L_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i=0)^{1-\xi_i} \\ &\times L_3(\xi_i | W_i) \text{Pr}(R_i=1 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i) \prod_{i \in \mathcal{R}_0} L_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i=1)^{\xi_i} L_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i=0)^{1-\xi_i} \\ &\times L_3(\xi_i | W_i) \text{Pr}(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i), \end{aligned}$$

where $\tilde{T} = (\tilde{T}_1, \dots, \tilde{T}_n)'$, with ξ, R, Z and W defined analogously. Under the assumed models, we have

$$\Lambda_i(t) = \int_0^t \lambda_0(s) e^{\beta'_0 Z_i(s)} ds \quad L_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i=1) = \lambda_0(\tilde{T}_i)^{\Delta_i} e^{\beta'_0 Z_i(\tilde{T}_i) \Delta_i} e^{-\Lambda_i(\tilde{T}_i)} \quad (4)$$

$$\Lambda_{i2}(t) = \int_0^t \lambda_{02}(s) e^{\beta'_{02} Z_i(s)} ds \quad L_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i=0) = \lambda_{02}(\tilde{T}_i)^{\Delta_i} e^{\beta'_{02} Z_i(\tilde{T}_i) \Delta_i} e^{-\Lambda_{i2}(\tilde{T}_i)}. \quad (5)$$

Combining (3), (4) and (5), the posterior mean of the missing ξ_i (where $R_i = 0$), conditional on the observed data $(T_i, \Delta_i, Z_i, W_i)$, is given by

$$E(\xi_i | \tilde{T}_i, \Delta_i, Z_i, W_i) = \frac{\lambda_i(\tilde{T}_i)^{\Delta_i} e^{-\Lambda_i(\tilde{T}_i)} e^{\gamma_0' W_i} Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=1)}{\left\{ \lambda_i(\tilde{T}_i)^{\Delta_i} e^{-\Lambda_i(\tilde{T}_i)} e^{\gamma_0' W_i} Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=1) + \lambda_{i2}(\tilde{T}_i)^{\Delta_i} e^{-\Lambda_{i2}(\tilde{T}_i)} Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=0) \right\}} \quad (6)$$

The conditional expectation in (6) depends on unknown parameters $\lambda_0(t)$, $\lambda_{02}(t)$, β_0 , β_{02} and γ_0 . Therefore, the EM algorithm requires iteration between the conditional expectation and maximum likelihood estimators. The computation imposed by such iterations can be intensive and, hence, may be accompanied by slow convergence.

The conditional expectation (6) can be rewritten as

$$E(\xi_i | \tilde{T}_i, \Delta_i, Z_i, W_i) = \text{expit} \left[\Delta_i \log \left\{ \lambda_0(\tilde{T}_i) / \lambda_{02}(\tilde{T}_i) \right\} + (\beta_0 - \beta_{02})' Z_i(\tilde{T}_i) \Delta_i - \Lambda_i(\tilde{T}_i) + \Lambda_{i2}(\tilde{T}_i) + \gamma_0' W_i + \log \left\{ \frac{Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=1)}{Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=0)} \right\} \right]$$

where $\text{expit}(a) = e^a / (1 + e^a)$. This conditional expectation has the same form as the predicted probability from a logistic model with linear predictor,

$$\Delta_i \log \left\{ \frac{\lambda_0(\tilde{T}_i)}{\lambda_{02}(\tilde{T}_i)} \right\} + (\beta_0 - \beta_{02})' Z_i(\tilde{T}_i) \Delta_i - \Lambda_i(\tilde{T}_i) + \Lambda_{i2}(\tilde{T}_i) + \gamma_0' W_i + \log \left\{ \frac{Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=1)}{Pr(R_i=0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i=0)} \right\} \quad (7)$$

Taking advantage of the complete data, $\mathcal{R}_1$, in which ξ_i are observed, we propose an estimator for $E(\xi_i | T_i, \Delta_i, Z_i, W_i)$ that does not require iterations. Specifically, we fit a logistic model using the complete data, $\mathcal{R}_1$, then apply this model to $\mathcal{R}_0$ to predict the posterior mean of missing ξ_i .

Covariates used in estimating $E(\xi_i | T_i, \Delta_i, Z_i, W_i)$ include the covariates in model (3), W_i , the products of the covariates in the Cox models and the event indicator, $Z_i(T_i)$, the event indicator, Δ_i , and the estimated cumulative hazards at the observed time, $\Lambda_i(T_i)$ and $\Lambda_{i2}(T_i)$. The unknown parameters and their functions, γ_0 , $\beta_0 - \beta_{02}$ and $\log \{ \lambda_0(T_i) / \lambda_{02}(T_i) \}$, are treated as regression coefficients and estimated from $\mathcal{R}_1$. Although $\log \{ \lambda_0(T_i) / \lambda_{02}(T_i) \}$ might depend on T_i , when the baseline hazard functions $\lambda_0(t)$ and $\lambda_{02}(t)$ are proportional, $\log \{ \lambda_0(T_i) / \lambda_{02}(T_i) \}$ reduces to an unknown constant, which can be estimated as the coefficient for Δ_i . We assume that the baseline hazard functions for eligible and non-eligible subjects are proportional. If this was found to not hold (e.g., through various residual plots or supplementary model fitting), then departures from proportionality could be accommodated parametrically. When the missingness mechanism is MAR, conditional on $(T_i, \Delta_i, Z_i, W_i)$, the probability of missing the eligibility indicator, $Pr(R_i = 0)$, does not depend on the value of ξ_i . Therefore, $Pr(R_i = 0 | T_i, \Delta_i, Z_i, W_i, \xi_i = 1) = Pr(R_i = 0 | T_i, \Delta_i, Z_i, W_i, \xi_i = 0)$ and $\log \{ Pr(R_i = 0 | T_i, \Delta_i, Z_i, W_i, \xi_i = 1) / Pr(R_i = 0 | T_i, \Delta_i, Z_i, W_i, \xi_i = 0) \} = 0$. In Sect. 4, we evaluate through simulation the robustness of the proposed methods to (1) violations to the proportionality of $\lambda_0(t)$ and $\lambda_{02}(t)$ (2) departures from MAR.

In (7), the covariates $W_i, Z_i(T_i)$ are observed. However, $\Lambda_i(T_i)$ and $\Lambda_{i2}(T_i)$ need to be estimated. The observed time, the log-transformed observed time, and their combinations have been used by previous authors in predicting the posterior mean of missing covariates in various survival settings; e.g., van Buuren et al. (1999), Clark and Altman (2003), Barzi and Woodward (2004). White and Royston (2009) compared various forms of subject-level cumulative hazards estimators in predicting the posterior mean of binary missing covariates; including no adjustment, logarithm of observed time, observed time, square of observed time, Nelson-Aalen type hazard estimator, and Breslow-Aalen type hazard estimator. In simulations, all forms, except no-adjustment, were found to perform reasonably well, leading to small bias in the Cox regression coefficient and coverage probabilities between 0.93 and 0.96. The cumulative hazards at the observed time points depend on both the baseline hazards, $\lambda_0(t)$ and $\lambda_{02}(t)$, and the regression coefficients, β_0 and β_{02} . Traditionally, we would iterate between the parameter estimates and the conditional expectations of missingness indicators, which requires no less computation than the EM algorithm. To estimate the posterior mean of the missing eligibility indicator in one step, we propose two approximations for the subject-level cumulative hazards. The first one is nonparametric, the Nelson-Aalen estimator of the marginal cumulative hazard. The second estimator is semiparametric, using Breslow estimators based on Cox models (1) and (2) fitted to the complete cases in the eligible ($\xi_1 = 1$) and ineligible ($\xi_1 = 0$) groups, respectively. The nonparametric estimator intended to serve as an approximation and, strictly speaking, would be the correct choice only when the eligible and ineligible samples were two homogeneous groups. The approximation is reduced in accuracy to the extent that the groups differ from each other, and to the extent that β is away from the null. From this perspective, the semiparametric estimator is valid much more generally, but requires a MAR assumption since it is based on complete cases. However, the effects of small biases in one covariate in predicting the conditional mean of the missing variable is unlikely to greatly affect the Cox model estimates (White and Royston 2009), meaning that the proposed methods should perform well in many practical settings. We examine the robustness of both one-step approximations through simulation in Sect. 4.

Based on the one-step posterior mean estimates, a two-step procedure is proposed for inference on the survival outcomes in the target population. In the first step, a logistic model is fitted to the complete set $\mathcal{R}_1$ and the probability of being eligible ($\xi_i = 1$) conditional on the observed data is modeled as

$$\text{logit}(\tilde{p}_i) = \theta'_0 X_i$$

where $p_i(\theta_0) = E(\xi_i | T_i, Z_i, W_i) = \Pr(\xi_i = 1 | T_i, Z_i, W_i)$ and $X_i = (W_i, Z_i(T_i), \hat{\Lambda}_i(T_i), \hat{\Lambda}_{i2}(T_i))$. The maximum likelihood estimator of θ_0, θ , are solutions to the score equation of the logistic model

$$U_L(\theta) = \sum_{i=1}^n U_{L_i}(\theta) \quad U_{L_i}(\theta) = R_i X_i \{\xi_i - \tilde{p}_i(\theta)\}.$$

Then, $p_i \tilde{\theta} = \text{expit}(\theta' X_i)$ is the maximum likelihood estimator for $p_i \tilde{\theta}$ by the invariance property of maximum likelihood estimators.

In the second step, we replace the missing ξ_i in $\mathcal{R}_0$ with the posterior mean estimate $p_i \tilde{\theta}$ and fit a weighted Cox proportional hazards model, where the weights are ξ_i if $i \in \mathcal{R}_1$ or $p_i \tilde{\theta}$ if $i \in \mathcal{R}_0$. The score equation for the weighted Cox model on the pooled data of $\mathcal{R}_1$ and $\mathcal{R}_0$ is given by $U(\beta, \theta) = 0$, where

$$U(\beta, \theta) = \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(\beta, \theta; t)\} \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} dN_i(t), \quad (8)$$

with

$$\begin{aligned} \bar{Z}(\beta, \theta; t) &= \frac{S^{(1)}(\beta, \theta; t)}{S^{(0)}(\beta, \theta; t)} \\ S^{(0)}(\beta, \theta; t) &= n^{-1} \sum_{i=1}^n \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} Y_i(t) e^{\beta' Z_i(t)} \\ S^{(1)}(\beta, \theta; t) &= n^{-1} \sum_{i=1}^n \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} Y_i(t) e^{\beta' Z_i(t)} Z_i(t). \end{aligned}$$

The corresponding weighted cumulative baseline hazard estimator is then given by

$$\hat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} dN_i(s)}{S^{(0)}(\beta, \theta; s)}.$$

3 Asymptotic properties

In this section, we describe the asymptotic properties of the parameter estimators from the weighted Cox model fitted at the second step of the proposed methods. We assume the following regularity conditions:

- a. $\{T_i, C_i, Z_i, \xi_i, X_i, R_i\}$ are independent and identically distributed for $i = 1, \dots, n$.
- b. $Pr(T_i \sim \tau) > 0$ for a constant $\tau > 0$.
- c. $\int_0^{\tau} d\Lambda_0(t) < \infty$.
- d. $|X_{i:l}| < \infty$ and $\int_0^{\tau} d|Z_{i:m}(s)| < \infty$ almost surely, where l and m refer to the l th and m th element of X_i and Z_i , respectively.
- e. Positive-definiteness of the matrices, $A(\beta, \theta)$ and $B(\theta)$, where

$$\begin{aligned}
A(\beta, \theta) &= E \left[\int_0^\tau \{Z_i(t) - \bar{z}(\beta, \theta; t)\}^{\otimes 2} \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} Y_i(t) e^{\beta' Z_i(t)} d\Lambda_0(t) \right], \\
s^{(1)}(\beta, \theta; t) &= \lim_{n \rightarrow \infty} S^{(1)}(\beta, \theta; t), \\
s^{(0)}(\beta, \theta; t) &= \lim_{n \rightarrow \infty} S^{(0)}(\beta, \theta; t), \\
\bar{z}(\beta, \theta; t) &= \frac{s^{(1)}(\beta, \theta; t)}{s^{(0)}(\beta, \theta; t)}, \\
B(\theta) &= E[R_i X_i \tilde{p}_i(\theta) \{1 - \tilde{p}_i(\theta)\} X_i'],
\end{aligned}$$

with $a^{\otimes 2} = aa'$.

The independent and identically distributed condition (a) applies to many practical settings and simplifies the asymptotic development. Recall that the covariate X_i includes the cumulative hazard estimator, such that $\hat{\Lambda}_i(\tilde{T}_i)$ and $\hat{\Lambda}_{i2}(\tilde{T}_i)$ are naturally correlated across subjects. However, $\hat{\Lambda}_i(\tilde{T}_i)$ and $\hat{\Lambda}_{i2}(\tilde{T}_i)$ converge to $\Lambda_i(\tilde{T}_i)$ and $\Lambda_{i2}(\tilde{T}_i)$, respectively, such that the dependence would become small for large n . Condition (b) is required for the identifiability of the event process parameters on $(0, \tau]$. In practice, τ is usually taken as the largest $\tilde{T}_i$ to avoid extrapolating inferences to regions without data. Conditions (c) and (d) can be relaxed, but at the expense of additional technical considerations. Note that the boundedness condition, (d), applies to all elements of Z_i and X_i . Assumption (e) is typically satisfied upon reasonable specification of the respective covariate vectors.

With respect to the proposed regression parameter estimator, we have the following result.

Theorem 1 *Under conditions (a) to (e), $\hat{\beta}$ converges almost surely to β_0 . Furthermore, $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $A(\beta_0, \theta_0)^{-1} \Sigma(\beta_0, \theta_0) A(\beta_0, \theta_0)^{-1}$ as $n \rightarrow \infty$, where*

$$\begin{aligned}
\Sigma(\beta, \theta) &= E\{\psi_i(\beta, \theta)^{\otimes 2}\}, \\
\psi_i(\beta, \theta) &= \int_0^\tau \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} \{Z_i(t) - \bar{z}(\beta, \theta; t)\} dM_i(\beta, \theta; t) + G(\beta, \theta) B(\theta)^{-1} U_{Li}, \\
dM_i(\beta, \theta; t) &= dN_i(t) - Y_i(t) e^{\beta' Z_i(t)} d\Lambda_0(t) \\
G(\beta, \theta) &= E \left[\int_0^\tau \{Z_i(t) - \bar{z}(\beta, \theta; t)\} (1 - R_i) \zeta_i(\theta) dM_i(\beta, \theta; t) \right] \\
\zeta_i(\theta) &= \frac{\partial \tilde{p}_i(\theta)}{\partial \theta} = X_i' \tilde{p}_i(\theta) (1 - \tilde{p}_i).
\end{aligned}$$

Similar to the proof of the EM algorithm, the parameter estimates $\hat{\beta}$ maximizing the conditional expectation of the logarithm of the joint likelihood given the observed data also maximize the marginal likelihood of the observed data. The consistency of $\hat{\beta}$ for β_0 from model (1) can be derived from the convergence of $\hat{\theta}$ to θ_0 , the Strong Law of Large Numbers and the Continuous Mapping Theorem, Taylor expansions, and the Empirical Central Limit Theorem (Pollard 1990; van der Vaart and Wellner 1996; van der Vaart 2000). Details are

provided in the Appendix. Note that the Martingale Central Limit Theorem (Fleming and Harrington 1991; Kalbfleisch and Prentice 2002) is not applicable since the weights $p_i \tilde{\gamma}(\theta)$ are not predictable due to W_i .

We summarize the essential asymptotic results for the baseline cumulative hazard estimator in Theorem 2.

Theorem 2 Under conditions (a) to (e), $\Lambda_0(\hat{\beta}, \hat{\theta}; t)$ converges uniformly to $\Lambda_0(t)$ for $t \in [0, \tau]$ and $n^{1/2}\{\Lambda_0(\hat{\beta}, \hat{\theta}; t) - \Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function,

$$\Omega(s, t) = E\{\phi_i(\beta_0, \theta_0; s)\phi_i(\beta_0, \theta_0; t)\},$$

for $(s, t) \in [0, \tau] \times [0, \tau]$, where $\phi_i(\beta, \theta; t) = \phi_{i1}(\beta, \theta; t) + \phi_{i2}(\beta, \theta; t) + \phi_{i3}(\beta, \theta; t)$, with

$$\begin{aligned} b(\beta, \theta; t) &= E[(1 - R_i)\zeta_i(\theta)Y_i(t)e^{\beta'Z_i(t)}], \\ k(\beta, \theta; t) &= E\left[\int_0^t \left\{ \frac{(1 - R_i)\zeta_i(\theta)}{S^{(0)}(\beta, \theta; s)} + \frac{\{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\}\hat{b}'(\beta, \theta; s)}{\{S^{(0)}(\beta, \theta; s)\}^2} \right\} dN_i(s)\right], \\ \phi_{i1}(\beta, \theta; t) &= k'(\beta, \theta; t)B(\theta)^{-1}U_{Li}(\theta), \\ h(\beta, \theta; t) &= \int_0^t \bar{Z}(\beta, \theta; s)d\hat{\Lambda}_0(\beta, \theta; s), \\ \phi_{i2}(\beta, \theta; t) &= h'(\beta, \theta; t)A(\beta, \theta)^{-1}\psi_i(\beta, \theta), \\ \phi_{i3}(\beta, \theta; t) &= \int_0^t \frac{dM_i(\beta, \theta; s)}{s^{(0)}(\beta, \theta; s)}. \end{aligned}$$

4 Simulation studies

We generated four covariates. The first covariate Z_{i1} followed a Bernoulli(0.5) distribution and serves as the factor of chief interest. Adjustment covariates (Z_{i2}, Z_{i3}, Z_{i4}) are intended to represent degrees of a specific condition. They are generated through a variate $U_i \sim \text{Unif}(0, 4)$, with $Z_{i2} = I\{U_i \in [0, 1]\}$, $Z_{i3} = I\{U_i \in (1, 2]\}$ and $Z_{i4} = I\{U_i \in (2, 3]\}$; with the highest level, (3,4], serving as the reference group. The eligibility indicators follow a logistic model $\text{logit}(p_i) = -3 + Z_{i1} + 2Z_{i2} + Z_{i3} + 0.5Z_{i4}$. About 80% of the eligibility indicators ξ_i equal 1 and the corresponding event hazard rates follow a Cox model $\lambda_i(t) = .025 \exp(\beta Z_{i1} + 2Z_{i2} + Z_{i3} + 0.5Z_{i4})$; hence, the parameter of interest is β . The remaining 20%, with $\xi_i = 0$, follow another Cox model $\lambda_{i2}(t) = .0125 \exp(Z_{i1} + Z_{i2} + 0.5Z_{i3} + Z_{i4})$. Censoring time C_i follows uniform distribution on $[0, 40]$, leading to censoring percentages around 35 and 40% in the $\xi_i = 1$ and $\xi_i = 0$ groups respectively, when $\beta = 0.5$. Three sample sizes, $n = 100$, $n = 200$ and $n = 300$, are examined. The eligibility indicators are MCAR and three missing percentages are evaluated: 25, 50, 75%. The parameters in the first model where $\xi_i = 1$ are of interest, especially the regression coefficient for Z_{i1} , β .

The posterior means of the missing ξ_i values are estimated from a logistic model fitted on the complete cases, whose covariates include $Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}$, i , $Z_{i1} \cdot i, Z_{i2} \cdot i, Z_{i3} \cdot i, Z_{i4} \cdot i, \hat{\Lambda}_i(T_i)$ and $\hat{\Lambda}_{i2}(T_i)$. The cumulative hazard estimates, $\hat{\Lambda}_i(T_i)$ and $\hat{\Lambda}_{i2}(T_i)$, are obtained either nonparametrically or semiparametrically based on the complete cases. The observed ξ_i values and estimated $\hat{E}(\xi_i | T_i, i, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4})$ are used as weights in a Cox model with covariates Z_{i1}, Z_{i2}, Z_{i3} and Z_{i4} . Table 1 reports and compares the bias, empirical standard deviations (ESD), asymptotic standard errors (ASE) and coverage probabilities (CP) of regression coefficients estimates from the complete case analysis and those from the proposed two-step procedure. The relative efficiency (RE) of the proposed estimates versus the complete case analysis is reported in the last column, which is calculated as the ratios of the empirical variances from the complete case analysis and those from the proposed method. From Table 1, both the complete-case analysis and the semiparametric version of the proposed procedure are generally unbiased, with CP close to the nominal level. However, the nonparametric version has slightly larger bias and CP from 90 to 97%. Also, the sandwich-type ASE tends to be larger than the ESD. Nevertheless, the proposed procedure leads to considerable efficiency gains, which tend to increase with increasing percent missingness. Comparing the two versions of the proposed estimator (non- and semiparametric cumulative hazard estimators), that based on the semiparametric cumulative hazard appears to perform slightly better in terms of bias and coverage probability; although the efficiency gains (relative to the complete-case analysis) are slightly less pronounced.

The logistic model in the second step consistently estimates the conditional expectations of the missing eligibility indicators, under the assumptions that the two baseline hazards functions $\lambda_0(t)$ and $\lambda_{02}(t)$ are proportional. The nonparametric estimators of $\hat{\Lambda}_i(T_i)$ and $\hat{\Lambda}_{i2}(T_i)$ require independently and identically distributed T_i and eligibility status which is MCAR, while the semiparametric $\hat{\Lambda}_i(T_i)$ and $\hat{\Lambda}_{i2}(T_i)$ from Cox models on the complete cases require only MAR. Robustness of the proposed method in the presence of assumption violations is examined in Table 2. The parameter setup in the first row of Table 1 is used. We examine the case when $\lambda_0(t)$ is a constant while $\lambda_{02}(t)$ is piecewise constant, leading to different baseline hazards ratios in different time intervals (nonPH). We also examined two MAR scenarios where missing probabilities are generated as functions of either Z_{i1} (MAR1) or T_i (MAR2). Three MNAR cases are generated where $\log\{Pr(R_i | \xi_i = 1)/Pr(R_i | \xi_i = 0)\}$ is a constant (MNAR1), or a function of Z_{i1} (MNAR2) or T_i (MNAR3). Under the nonPH scenario, the performance of the two-step method with random weights is almost unaffected. When the missing mechanism is not MAR, the bias is nonzero and the ASE tends to be larger than the ESD. However, the bias and discrepancy in variance estimator are small and the coverage probabilities are between 0.95 and 0.97. This is not the case with complete case analysis. When the missing probability depends on the event time T_i (MNAR3), the complete case analysis is severely biased. Besides, throughout all the scenarios examined, the proposed method exhibits considerable efficiency gains over the complete case analysis.

5 Application

This study used data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and

Transplantation Network (OPTN), and has been described elsewhere (SRTR 2007). The Health Resources and Services Administration (HRSA), US Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

Our objective is to compare ECD transplants versus non-ECD transplants among end-stage-renal-disease patients willing to receive an ECD kidney. The outcome of interest was graft failure, with death included as a cause of graft failure (as is typically done in such studies). We include in our study transplant recipients who initially registered on the kidney transplant wait list between 01/01/2000 and 12/31/2010 and received a transplant before 12/31/2010. We excluded patients age <18 , repeat transplants, and living-donor transplants. In total, the study population included $n = 85,025$ transplant recipients. However, 20,174 registration forms before 07/01/2002 do not record the patient's eligibility for ECD transplantation. It is worth pointing out that 2,931 subjects with missing wait list membership actually received an ECD transplant, thus they fall into the willing-to-accept-ECD group. In summary, 17,243 (20%) of ECD eligibility indicators are missing. The remaining 67,782 patients (80%) indicated whether or not they were willing to receive an ECD transplant. Among such patients, a total of 38,354 (45%) were in fact listed for ECD transplant. Among the ECD wait listed candidates, 16,001 received ECD transplant, while 22,353 received a non-ECD transplant. Among subjects on the ECD wait list, 10,054 recipients experienced graft failure, which includes 5,614 ECD recipients and 4,440 non-ECD recipients. Among the 29,428 patients restricting their choices to non-ECD kidneys only, 4,802 experienced graft failure. There were 7,575 graft failure cases among the 17,243 subjects with missing ECD eligibility indicators. Note that missing eligibility status was related only to calendar period. In this sense, the MAR assumption seems reasonable in this analysis.

Adjustment covariates for each of the Cox models included adjusting for age, gender, race, body mass index (BMI), history of cardiovascular disease, drug treated Chronic Obstructive Pulmonary Disease (COPD), Hepatitis B Virus (HBV) positive, primary diagnosis, previous malignancy, previous Intensive Care Unit (ICU) stay, previous hospital stay but not in the ICU, functional status, calendar year of transplant, and number of years on the wait list prior to transplantation. For the proposed method, the logistic model adjusts for the covariates set in the Cox model (i.e., $W_i = Z_i$), plus the non- and semiparametric cumulative hazard estimates, censoring indicator and the interactions between the censoring indicator and Z_i .

We list in Table 3 the results of four Cox models for the kidney transplant data set. The first three models are based on ECD-eligible patients. The complete case analysis discards the 17,243 subjects with missing ECD-eligibility indicators, while the proposed two-step procedure incorporates partial information from these subjects. The fourth model is based on a different study population; specifically, all ECD and non-ECD recipients, irrespective of their willingness (or lack thereof) to accept ECD kidneys. Both versions of the proposed two-step estimator provide hazard ratio estimates smaller than the complete case analysis, with the smallest being from the semiparametric version ($\exp\{\hat{\beta}_E\} = 1.35$). Both versions of the two-step method have much smaller standard errors than the complete case analysis; corresponding to 65 and 58% efficiency gains for the non- and semiparametric versions, respectively. Based on the semiparametric version of the two-step method, ECD (relative to

non-ECD) transplantation increases the graft failure hazard by 35%, with covariate-adjusted hazard ratio of $\exp\{\hat{\beta}_E\} = 1.35$ ($p < 0.0001$). In contrast the ignore-eligibility analysis yields $\exp\{\hat{\beta}_E\} = 1.56$ ($p < 0.0001$). The larger hazard ratio in the ignore-eligibility analysis may be due to the fact that subjects off the ECD wait list generally had better health at the time of listing and, consequently, were also healthier at the time of transplant; i.e., in ways not captured by Z_i . The discrepancy between the two estimators is noteworthy given their standard errors. Comparing the two proposed methods, since the semiparametric estimator should be accurate in the presence of strong adjustment covariates effects, one would tend to trust it more in this application.

With respect to the adjustment covariates, based on the proposed two-step procedure, females, patients age 55, Hispanics, Asians, and patients with BMI [20,30] have significantly lower graft failure hazard. History of angina, peripheral vascular disease, drug treated COPD and hospitalization at time of wait listing are associated with increased risk. In terms of primary diagnosis, diabetes and hypertension were each associated with significantly increased graft failure rates relative to polycystic kidney disease. Each additional year on the wait list leads to a 3% higher post-transplant hazard (data not tabulated).

6 Discussion

The proposed methods were mostly motivated by large data sets, such as the registry data analyzed in Sect. 5. Unlike related methods in the existing literature, our method uses a one-step estimator of the posterior mean. That is, we estimate the conditional expectation of the missing eligibility indicators through a logistic regression based on the complete data. Our two-step models do not need iterations and simplify the computation greatly. Various approximations are used, in order to simplify computations. In particular, for the cumulative hazard function, we use either a Nelson-Aalen estimator or a Breslow estimator, with the accuracy of the former likely to decrease in the presence of strong covariates effect on the hazard.

Schaubel and Cai (2006) study recurrent event from multiple categories with partially missing event categories. The contribution of an event with missing category to each event category in the score function is weighted by the probability of the event belonging to the category. However, there are major differences between recurrent event and univariate event. First, a subject is at risk for events in all categories in their report while only eligible subjects are at risk for the events of interest in our manuscript. Second, in the case of recurrent events, the event history provides a good approximation of the hazard/rate function, which is not available in the univariate survival case.

The procedures proposed in this report have the potential to be generalized to other types of missing information and outcomes other than univariate survival times.

Acknowledgments

This work was supported in part by National Institutes of Health Grant 5R01-DK070869. The authors thank the Associate Editor and two Reviewers for constructive comments which strengthened the article. The study was

approved by the Internal Review Board at George Washington University. The data reported here have been supplied by the Minneapolis Medical Research Foundation (MMRF) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the SRTR or the US Government.

Appendix

Proof of Theorem 1 The logarithm of the complete joint likelihood is

$$\begin{aligned}\ell_J = & \sum_{i=1}^n \{ \xi_i l_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i) \\ & + (1 - \xi_i) l_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i) \} + l_3(\xi_i | W_i) \\ & + \sum_{i \in \mathcal{R}_1} \log \{ Pr(R_i = 1 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i) \} \\ & + \sum_{i \in \mathcal{R}_0} \log \{ Pr(R_i \\ & = 0 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i) \}.\end{aligned}$$

For the subset $\mathcal{R}_0$, ξ_i is missing. The joint log-likelihood can be decomposed into the sum of log-marginal-likelihood ℓ_M and log-conditional-likelihood ℓ_C , where

$$\begin{aligned}\ell_M = & \sum_{i \in \mathcal{R}_1} \xi_i l_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i) + (1 - \xi_i) l_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i) + l_3(\xi_i | W_i) + \log \{ Pr(R_i = 1 | \tilde{T}_i, \Delta_i, Z_i, W_i, \xi_i) \} \\ & + \sum_{i \in \mathcal{R}_0} [\log \{ \sum_{\xi_i=0}^1 L_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i) \} - \log \{ \sum_{\xi_i=0}^1 L_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i) \} \\ & + \log \{ \sum_{\xi_i=0}^1 L_3(\xi_i | W_i) \}] \\ \ell_C = & \sum_{i \in \mathcal{R}_1} \{ \xi_i l_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i) + (1 - \xi_i) l_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i) \} - \sum_{i \in \mathcal{R}_1} \log \left[\sum_{\xi_i=0}^1 \{ L_1(\tilde{T}_i, \Delta_i | Z_i, \xi_i)^{\xi_i} L_2(\tilde{T}_i, \Delta_i | Z_i, \xi_i)^{1-\xi_i} L_3(\xi_i | W_i) \} \right]\end{aligned}$$

The only form of variable ξ_i in ℓ_J is itself. Therefore, when we replace the missing ξ_i in ℓ_J with its posterior expectation $p_i = E(\xi_i | \tilde{T}_i, \Delta_i, Z_i, W_i)$, we get the expectation of ℓ_J conditional on the observed data $\tilde{T}_i, \Delta_i, Z_i, W_i$, denoted by $E(\ell_J | \tilde{T}_i, \Delta_i, Z_i, W_i)$. Take one element of β , β_k , as an example. It will be seen from the following that the β_k value, $\hat{\beta}_k$, maximizing $E(\ell_J | \tilde{T}_i, \Delta_i, Z_i, W_i)$ also maximizes ℓ_M .

$$\begin{aligned}
& \frac{\partial E(\ell_J | \tilde{T}_i, \Delta_i, Z_i, W_i)}{\partial \beta_k} \\
&= \frac{\partial E\{\log(L_J | \tilde{T}_i, \Delta_i, Z_i, W_i)\}}{\partial \beta_k} \\
&= \frac{\partial E\{\log(L_C | \tilde{T}_i, \Delta_i, Z_i, W_i) + (\ell_M | \tilde{T}_i, \Delta_i, Z_i, W_i)\}}{\partial \beta_k} \\
&= \frac{\partial \sum_{\xi_i=0}^1 \{\log(L_C) L_C\} + \ell_M}{\partial \beta_k} \\
&= \sum_{\xi_i=0}^1 \left\{ \frac{\partial L_C}{\partial \beta_k} / L_C \right. \\
&\quad \left. * L_C \right\} + \frac{\partial \ell_M}{\partial \beta_k} = \sum_{\xi_i=0}^1 \left\{ \frac{\partial L_C}{\partial \beta_k} \right\} \\
&\quad + \frac{\partial \ell_M}{\partial \beta_k} = \frac{\partial \sum_{\xi_i=0}^1 L_C}{\partial \beta_k} \\
&\quad + \frac{\partial \ell_M}{\partial \beta_k} = \frac{\partial 1}{\partial \beta_k} + \frac{\partial \ell_M}{\partial \beta_k} = 0 + \frac{\partial \ell_M}{\partial \beta_k}.
\end{aligned}$$

Here L_J and L_C denote the joint and conditional likelihoods, respectively. In summary, the parameter estimates $\hat{\beta}, \hat{\lambda}_0(t)$ that maximize $E(\ell | \tilde{T}_i, \Delta_i, Z_i, W_i)$ also satisfy $\ell_M' / \beta = 0_{p \times 1}$ and $\ell_M' / \lambda_0(t) = 0_{q \times 1}$ where p is the dimension of β and q is the number of unique event time.

We further prove the consistency of $\hat{\beta}, \hat{\lambda}_0(t)$ from the proposed weighted Cox models. The log-likelihood of the weighted Cox model can be written as

$$\ell(\beta, \theta) = \sum_{i=1}^n \int_0^\tau \{R_i \xi_i + (1 - R_i) \tilde{p}_i(\theta)\} [\beta' Z_i(t) - \log\{S^{(0)}(\beta, \theta; t)\}] dN_i(t).$$

Define the average difference between the log profile likelihood at β, θ and that at the true parameter β_0, θ_0 as $D(\beta, \theta) = n^{-1} \{\ell(\beta, \theta) - \ell(\beta_0, \theta_0)\}$, which can be decomposed into four parts $D(\beta) = \sum_{k=1}^4 D_k(\beta)$, where

$$\begin{aligned}
D_1(\beta, \theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau \{\tilde{p}_i(\theta) - \tilde{p}_i(\theta_0)\} (1 - R_i) \beta' Z_i(t) dN_i(t) \\
D_2(\beta, \theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau \{R_i \xi_i + \tilde{p}_i(\theta_0) (1 - R_i)\} (\beta - \beta_0)' Z_i(t) dN_i(t) \\
D_3(\beta, \theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau -\{\tilde{p}_i(\theta) - \tilde{p}_i(\theta_0)\} (1 - R_i) \log\{S^{(0)}(\beta, \theta; t)\} dN_i(t) \\
D_4(\beta, \theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau -\{R_i \xi_i + \tilde{p}_i(\theta_0) (1 - R_i)\} \log\left\{\frac{S^{(0)}(\beta, \theta; t)}{S^{(0)}(\beta_0, \theta_0; t)}\right\} dN_i(t).
\end{aligned}$$

By assumption (e), X_i and Z_i are bounded almost surely; the increment $dN_i(t)$ is bounded by one since at most one event would happen for each subject. When $\theta = \hat{\theta}$, the Maximum Likelihood Estimator of θ_0 , $\hat{\theta} \xrightarrow{a.s.} \theta_0$ as $n \rightarrow \infty$. Then we apply Taylor Series expansion to $D_1(\beta, \hat{\theta})$ and get

$$D_1(\hat{\beta}, \hat{\theta}) = n^{-1} \sum_{i=1}^n \int_0^\tau (1 - R_i) \zeta(\theta_*) \beta' Z_i(t) dN_i(t) (\hat{\theta} - \theta_0) + o_p(1),$$

where $\zeta(\theta)$ is defined in Theorem 1 and θ_* lies between θ_0 and $\hat{\theta}$. By the Strong Law of Large Numbers (SLLN), $D_1(\hat{\beta}, \hat{\theta})$ converges almost surely to zero as $n \rightarrow \infty$. Since $S^{(0)}(\hat{\beta}, \hat{\theta}; t) \xrightarrow{a.s.} s^{(0)}(\beta_0, \theta_0; t)$ from the continuous mapping theorem, $S^{(0)}(\hat{\beta}, \hat{\theta}; t)$ is also bounded under assumed conditions. We apply the same procedure used in $D_1(\hat{\beta}, \hat{\theta})$ and reach a similar conclusion for $D_3(\hat{\beta}, \hat{\theta}) \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$.

Regarding $D_2(\hat{\beta}, \hat{\theta})$, $\int_0^\tau \beta'_0 Z_i(t) dN_i(t)$ is bounded and $R_i \xi_i + p_i(\theta_0)(1 - R_i)$ is between 0 and 1. Using the SLLN,

$$D_2(\beta, \theta) \xrightarrow{a.s.} E \left\{ \int_0^\tau \{R_i \xi_i + \tilde{p}_i(\theta_0)(1 - R_i)\} (\beta - \beta_0)' Z_i(t) dN_i(t) \right\} \equiv d_2(\beta, \theta). \text{ Similarly,}$$

$$D_4(\beta, \theta) \xrightarrow{a.s.} E \left[\int_0^\tau - \{R_i \xi_i + \tilde{p}_i(\theta_0)(1 - R_i)\} \log \{s^{(0)}(\beta, \theta; t) / s^{(0)}(\beta_0, \theta_0; t)\} dN_i(t) \right] \equiv d_4(\beta, \theta).$$

Combining results given above regarding $D_k(\beta)$ ($k = 1, 2, 3, 4$),

$D(\beta, \theta) \xrightarrow{a.s.} d_2(\beta, \theta) + d_4(\beta, \theta) \equiv d(\beta, \theta)$. When the set containing β_0 is compact, then $D(\beta, \hat{\theta}) \xrightarrow{a.s.} d(\beta, \hat{\theta})$ uniformly on the compact set. In addition, $d(\beta) / \beta|_{\beta=\beta_0} = 0_{p \times 1}$ and $^2 d(\beta) / \beta \beta' = -A(\beta, \hat{\theta})$ is negative definite by condition (f). Therefore, $d(\beta, \theta)$ has a unique maximizer $\beta = \beta_0$. When n is large enough, $D(\beta, \hat{\theta})$ will also have a unique maximizer $\beta = \hat{\beta}$ because $D(\beta, \hat{\theta}) / \beta|_{\beta=\hat{\beta}} = 0_{p \times 1}$. When $n \rightarrow \infty$, $\hat{\beta} \xrightarrow{a.s.} \beta_0$.

Using a Taylor expansion around $\beta = \beta_0$, $n^{1/2}(\hat{\beta} - \beta_0) = \hat{A}(\beta_*, \hat{\theta})^{-1} n^{-1/2} U(\beta_0, \hat{\theta})$ where β_* lies between $\hat{\beta}$ and β_0 in $\mathcal{R}^p$ and

$$\hat{A}(\beta, \theta) = n^{-1} \sum_{i=1}^n \left[\int_0^\tau \{Z_i(t) - \bar{z}(\beta, \theta; t)\}^{\otimes 2} \{\xi_i R_i + \tilde{p}_i(\theta)(1 - R_i)\} Y_i(t) e^{\beta' Z_i(t)} d\Lambda_0(t) \right].$$

Using the fact that $\hat{\theta} \xrightarrow{a.s.} \theta_0$, $\hat{\beta} \xrightarrow{a.s.} \beta_0$, the continuous mapping theorem and SLLN,

$$\hat{A}(\hat{\beta}, \hat{\theta}) \xrightarrow{a.s.} A(\beta_0, \theta_0).$$

Hence $n^{1/2}(\hat{\beta} - \beta_0) = A(\beta_0, \theta_0)^{-1} n^{-1/2} U(\beta_0, \theta_0) + o_p(1)$. Through the definition of $Z(\beta, \theta; t)$ and with some basic algebra,

$$U(\beta, \hat{\theta}) = \sum_{i=1}^n \int_0^\tau \{\xi_i R_i + \tilde{p}_i(\hat{\theta})(1 - R_i)\} \{Z_i(t) - \bar{Z}(\beta, \hat{\theta}; t)\} dM_i(\beta, \hat{\theta}; t).$$

We set $U(\beta, \hat{\theta}) = U_1(\beta, \hat{\theta}) + U_2(\beta, \hat{\theta})$, where

$$\begin{aligned} U_1(\beta, \hat{\theta}) &= \sum_{i=1}^n \int_0^\tau \{\xi_i R_i + \tilde{p}_i(\theta_0)(1 - R_i)\} \{Z_i(t) - \bar{Z}(\beta, \hat{\theta}; t)\} dM_i(\beta, \hat{\theta}; t) \\ U_2(\beta, \hat{\theta}) &= \sum_{i=1}^n \int_0^\tau \{(\tilde{p}_i(\hat{\theta}) - \tilde{p}_i(\theta_0))(1 - R_i)\} \{Z_i(t) - \bar{Z}(\beta, \hat{\theta}; t)\} dM_i(\beta, \hat{\theta}; t). \end{aligned}$$

By the strong convergence of $\hat{\theta}$ to θ_0 , the fact that

$$\left\| n^{-1/2} \sum_{i=1}^n \int_0^\tau \{\xi_i R_i + \tilde{p}_i(\theta_0)(1 - R_i)\} \{\bar{Z}(\beta_0, \theta_0; t) - \bar{z}(\beta_0, \theta_0; t)\} dM_i(\beta_0, \theta_0; t) \right\| \xrightarrow{P} 0_{p \times 1},$$

and the continuous mapping theorem,

$$n^{-1/2} U_1(\beta_0, \hat{\theta}) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \{\xi_i R_i + \tilde{p}_i(\theta_0)(1 - R_i)\} \{Z_i(t) - \bar{z}(\beta_0, \theta_0; t)\} dM_i(\beta_0, \theta_0; t) + o_p(1). \quad (9)$$

With respect to $U_2(\beta_0)$, from the logistic regression that produces $\hat{\theta}$,

$$n^{1/2}(\hat{\theta} - \theta_0) = \hat{B}(\theta_0)^{-1} n^{-1/2} U_{Li}(\theta_0) + o_p(1), \quad (10)$$

where $\widehat{B}(\theta) = n^{-1} \sum_{i=1}^n R_i X_i \tilde{p}_i(\theta) \{1 - \tilde{p}_i(\theta)\} X_i'$. By the SLLN, $\hat{B}(\theta_0) \xrightarrow{a.s.} B(\theta_0)$. Replacing $\hat{\theta}$ and $Z(\beta_0, \hat{\theta}; t)$ with their asymptotic expectations,

$$n^{-1/2} U_2(\beta_0, \hat{\theta}) = n^{-1/2} \hat{G}(\beta_0, \theta_0) B(\theta_0)^{-1} \sum_{i=1}^n R_i U_{Li}(\theta_0) + o_p(1). \quad (11)$$

Here, $\hat{G}(\beta, \theta) = n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{z}(\beta, \theta; t)\} (1 - R_i) \zeta_i(\theta) dM_i(\beta, \theta; t)$. Again

$\hat{G}(\beta_0, \theta_0) \xrightarrow{a.s.} G(\beta_0, \theta_0)$ by SLLN. Combining (9) and (11),

$$n^{-1/2} U(\beta_0, \hat{\theta}) = n^{-1/2} \sum_{i=1}^n \psi_i(\beta_0, \theta_0) + o_p(1),$$

where $\psi_i(\beta, \theta)$ and $G(\beta, \theta)$ are defined in Theorem 1. Essentially, $U(\beta_0, \hat{\theta})$ can be viewed as the sum of independently and identically distributed mean-zero random vectors, and by Multivariate Central Limit Theorem (van der Vaart 2000),

$n^{-1/2}U(\hat{\beta}, \hat{\theta}) \xrightarrow{D} n(0_{p \times 1}, \Sigma(\beta_0, \theta_0))$. Finally, we derive the asymptotic distribution of $\hat{\beta}$ through Slutsky's Theorem (Sen and Singer 1993).

The proof of Theorem 2 is available upon request from the first author.

References

- Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology*. 2004; 160:34–45. [PubMed: 15229115]
- Chen HY, Little RJ. Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*. 1999; 94:896–908.
- Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *Journal of Clinical Epidemiology*. 2003; 56:28–37. [PubMed: 12589867]
- Cox D. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B*. 1972; 34:187–220.
- Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. New York: Wiley; 1991.
- Herring AH, Ibrahim JG, Lipsitz SR. Non-ignorable missing covariate data in survival analysis: a case-study of an International Breast Cancer Study Group trial. *Journal of the Royal Statistical Society Series C*. 2004; 53:293–310.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Hoboken, New Jersey: Wiley; 2002.
- Lin DY, Ying Z. Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*. 1993; 88:1341–1349.
- Paik MC, Tsai WY. On using the Cox proportional hazards model with missing covariates. *Biometrika*. 1997; 84:579–593.
- Pollard, D. *Empirical Processes: Theory and Applications*. Hayward, CA: Institute of Mathematical Statistics; 1990.
- Port FK, Bragg-Gresham JL, Metzger RA, Dykstra DM, Gillespie BW, Young EW, Delmonico FL, Wynn JJ, Merion RM, Wolfe RA, Held PJ. Donor characteristics associated with reduced graft survival: An approach to expanding the pool of donor kidneys. *Transplantation*. 2002; 74:1281–1286. [PubMed: 12451266]
- Schaubel DE, Cai J. Rate/mean regression for multiple-sequence recurrent event data with missing event category. *Scandinavian Journal of Statistics*. 2006; 33:191–207.
- Scientific Registry of Transplant Recipients. The 2006 SRTR Report on the State of Transplantation. *Am J Transplant*. 2007; 5((2)):1317–1433. 5.
- Sen, PK.; Singer, JM. *Large Sample Methods in Statistics*. New York: Chapman and Hall; 1993.
- van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*. 1999; 18:681–694. [PubMed: 10204197]
- van der Vaart, AW. *Asymptotic Statistics*. Melbourne: Cambridge; 2000.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. New York: Springer; 1996.
- White IR, Royston P. Imputing missing covariate values for the Cox model. *Statistics in Medicine*. 2009; 28:1982–1998. [PubMed: 19452569]
- Zhou H, Pepe MS. Auxiliary covariate data in failure time regression. *Biometrika*. 1995; 82:139–149.

Table 1

Simulation results: complete case versus two-step procedure

β	n	$P(R_i = 0)$	Complete case				Two-step procedure						
			Bias	ESD	ASE	CP	Bias	ESD	ASE	CP	RE		
Nonparametric $\hat{\Lambda}_1(\tilde{T}_i)$ and $\hat{\Lambda}_2(\tilde{T}_i)$													
0.5	300	0.25	0.009	0.199	0.196	0.95	-0.007	0.168	0.175	0.97	1.40		
1.0	300	0.25	0.010	0.200	0.196	0.96	-0.043	0.164	0.172	0.96	1.49		
2.0	300	0.25	0.032	0.218	0.219	0.95	-0.136	0.186	0.218	0.90	1.36		
0.5	300	0.25	0.009	0.199	0.196	0.95	-0.007	0.168	0.175	0.97	1.40		
0.5	300	0.50	0.013	0.249	0.243	0.95	-0.018	0.173	0.187	0.96	2.07		
0.5	300	0.75	0.028	0.379	0.363	0.94	-0.042	0.206	0.308	0.93	3.38		
0.5	300	0.25	0.009	0.009	0.199	0.95	-0.007	0.168	0.175	0.97	1.40		
0.5	200	0.25	0.023	0.243	0.244	0.95	0.005	0.205	0.221	0.97	1.41		
0.5	100	0.25	0.053	0.372	0.363	0.95	0.021	0.317	0.432	0.96	1.38		
Semiparametric estimators of $\hat{\Lambda}_1(\tilde{T}_i)$ and $\hat{\Lambda}_2(\tilde{T}_i)$													
0.5	300	0.25	0.012	0.188	0.196	0.96	0.010	0.168	0.176	0.96	1.25		
1.0	300	0.25	0.024	0.203	0.197	0.95	0.017	0.178	0.184	0.94	1.30		
2.0	300	0.25	0.039	0.229	0.219	0.94	0.014	0.204	0.266	0.95	1.26		
0.5	300	0.25	0.012	0.188	0.196	0.96	0.010	0.168	0.176	0.96	1.25		
0.5	300	0.50	0.013	0.256	0.246	0.94	0.015	0.193	0.197	0.94	1.76		
0.5	300	0.75	0.011	0.403	0.363	0.93	-0.003	0.232	0.314	0.91	3.02		
0.5	300	0.25	0.012	0.188	0.196	0.96	0.010	0.168	0.176	0.96	1.25		
0.5	200	0.25	0.021	0.253	0.245	0.95	0.017	0.220	0.244	0.96	1.32		
0.5	100	0.25	0.031	0.394	0.364	0.94	0.026	0.341	0.421	0.95	1.34		

ESD empirical standard deviation, ASE average asymptotic standard error, CP coverage probability, RE relative efficiency (empirical)

Table 2
Simulation results: evaluation of robustness of two-step method: non-PH baselines, MAR and MNAR

Violation	Complete case				Two-step procedure					
	Bias	ESD	ASE	CP	Bias	ESD	ASE	CP	RE	
Nonparametric $\hat{\Lambda}_1(T_i)$ and $\hat{\Lambda}_2(T_i)$										
nonPH	-0.005	0.206	0.196	0.94	-0.016	0.175	0.175	0.94	1.39	
MAR1	0.028	0.245	0.240	0.95	-0.014	0.174	0.191	0.96	1.98	
MAR2	-0.149	0.190	0.182	0.86	0.025	0.180	0.207	0.97	1.11	
MNAR1	0.001	0.189	0.189	0.95	-0.022	0.163	0.179	0.96	1.34	
MNAR2	0.015	0.194	0.189	0.94	-0.032	0.165	0.189	0.96	1.38	
MNAR3	-0.225	0.206	0.200	0.79	0.052	0.197	0.407	0.97	1.09	
Semiparametric $\hat{\Lambda}_1(T_i)$ and $\hat{\Lambda}_{2S}(T_i)$										
nonPH	0.009	0.202	0.197	0.94	0.013	0.178	0.178	0.95	1.29	
MAR1	0.026	0.257	0.241	0.94	0.020	0.194	0.196	0.94	1.75	
MAR2	-0.150	0.190	0.184	0.85	0.049	0.182	0.217	0.96	1.09	
MNAR1	0.003	0.191	0.190	0.95	0.002	0.175	0.181	0.95	1.19	
MNAR2	0.006	0.196	0.190	0.95	-0.010	0.176	0.192	0.96	1.24	
MNAR3	-0.233	0.203	0.201	0.78	0.069	0.213	0.286	0.96	0.91	

ESD Empirical standard deviation, ASE average asymptotic standard error, CP coverage probability, RE relative efficiency (empirical)

Table 3

Analysis of SRTR data: estimated ECD effect based on various methods

Analysis	Complete-case	Proposed two-step	method	Ignore-eligibility
$\Lambda_i(\tilde{T}_i)$ and $\Lambda_{i2}(\tilde{T}_i)$	n/a	Nonparametric	Semiparametric	n/a
Sample	Patients known to be on ECD wait list	Patients not known to be off the ECD wait list		All wait-listed patients
weight	$R_i \times \xi_i$	$R_i \xi_i + (1 - R_i) p_i(\tilde{\theta})$		1
n	38,354	55,597		85,025
$\hat{\beta}_E$	0.41	0.38	0.30	0.44
$SE(\hat{\beta}_E)$	0.0216	0.0168	0.0173	0.0162
p value	<0.0001	<0.0001	<0.0001	<0.0001
$\exp\{\hat{\beta}_E\}$	1.51	1.46	1.35	1.56
ARE	1	1.65	1.58	n/a

ARE asymptotic relative efficiency, relative to the complete-case analysis *n/a* not applicable