Published in final edited form as:

Biom J. 2018 July ; 60(4): 797–814. doi:10.1002/bimj.201700090.

Cox model with interval-censored covariate in cohort studies

Soohyun Ahn¹, Johan Lim², Myunghee Cho Paik², Ralph L. Sacco³, and Mitchell S. Elkind⁴

¹Department of Mathematics, Ajou University, Suwon, Korea

²Department of Statistics, Seoul National University, Seoul, Korea

³Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA

⁴Department of Neurology, Columbia University, NY, New York, USA

Abstract

In cohort studies the outcome is often time to a particular event, and subjects are followed at regular intervals. Periodic visits may also monitor a secondary irreversible event influencing the event of primary interest, and a significant proportion of subjects develop the secondary event over the period of follow-up. The status of the secondary event serves as a time-varying covariate, but is recorded only at the times of the scheduled visits, generating incomplete time-varying covariates. While information on a typical time-varying covariate is missing for entire follow-up period except the visiting times, the status of the secondary event are unavailable only between visits where the status has changed, thus interval-censored. One may view interval-censored covariate of the secondary event status as missing time-varying covariates, yet missingness is partial since partial information is provided throughout the follow-up period. Current practice of using the latest observed status produces biased estimators, and the existing missing covariate techniques cannot accommodate the special feature of missingness due to interval censoring. To handle interval-censored covariates in the Cox proportional hazards model, we propose an available-data estimator, a doubly robust-type estimator as well as the maximum likelihood estimator via EM algorithm and present their asymptotic properties. We also present practical approaches that are valid. We demonstrate the proposed methods using our motivating example from the Northern Manhattan Study.

Keywords

cohort study; Cox model; interval-censored covariates; missing time-varying covariate

Correspondence Myunhee Cho Paik, Department of Statistics, Seoul National University, Seoul, Korea., myunhee chopaik@snu.ac.kr.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information including source code to reproduce the results may be found online in the supporting information tab for this article.

1 | INTRODUCTION

In cohort studies the outcome is often time to a particular event, and subjects are followed at regular intervals. For example, in Northern Manhattan Study, stroke is the event of primary interest, and randomly recruited subjects free of stroke are followed over time. Periodic visits monitor another irreversible event influencing stroke such as diabetes, and a nonnegligible proportion of subjects who are nondiabetic at baseline develop diabetes over the period of follow-up. It is an important clinical question whether the up-to-date diabetic status increases the risk of stroke. Treating the up-to-date diabetic status as a time-varying covariate in Cox proportional hazards models requires continuous assessment over the entire follow-up period, which is practically not feasible since the diabetic status is recorded only at the time of the scheduled visits. When the diabetic status changes between the follow-up visits, the onset time of diabetes as well as diabetic status are interval censored.

The problem of an interval-censored covariate is prevalent in cohort studies when time-varying covariate status changes between the visits. In practice when the time-varying status is not known, the latest observed value is replaced. This naive approach introduces bias and produces distorted coverage probability.

In literature, methods for right or left-censored covariates have been studied in various settings. In the context of linear regression, the use of the censored covariates without any adjustment for the censoring has been shown to lead to a biased estimator of the regression coefficients and a biased test (Austin & Brunner, 2003; Rigobon & Stoker, 2009). Modeling the censoring mechanism parametrically, some authors have taken likelihood approaches (Austin & Hoch, 2004; Gajewski, Nicholson, & Widen, 2009; Heitjan & Rubin, 1990; Sattar, Sinha, & Morris, 2012; Tsimikas, Bantis, & Georgiou, 2012), and others, an estimating equation approach (Rigobon & Stoker, 2009). D'Angelo, Weissfeld, and on behalf of the GenIMS Investigators (2008) replaced the censored covariate values with the estimated conditional expectation under normality assumption, and more recently Atem, Qian, Johnson, and Betensky (2017) proposed a multiple imputation method to accommodate randomly censored covariates. Wang and Feng (2012) utilized the conditional quantiles of the censored covariates without assuming parametric models for censored covariates. In the context of Cox proportional hazards model, Chen, May, Ibrahim, Chu, and Cole (2014) proposed a joint model for longitudinal and survival data that accounts for both missing and left censoring in the longitudinal covariate.

While less attention is paid to interval-censored covariates, some pioneering works have been published. Gomez, Espinal, and Lagakos (2003) studied interval-censored covariates in the linear models, motivated by HIV infection data, and developed an algorithm for the nonparametric maximum likelihood estimation of the regression coefficients and Langohr and Melis (2014) extended the previous work of Gomez et al. (2003) on the estimation and the assessment of the assumptions of linear regression models. Langohr, Gomez, and Muga (2004) presented a formal treatment of the topic in the context of a fully parametric model using likelihood approach. In their model, the interval-censored covariate is assumed to be independent of any other variables. Goggins, Finkelstein, and Zaslavsky (1999) considered interval-censored covariates in the context of Cox model and provided a practically useful

ad-hoc multiple imputation method using randomly generated predicted failure time given the interval-censored time.

One may view the interval-censored covariate as missing time-varying covariate. There is extensive literature on missing time-constant covariate in Cox models. Data are called missing at random (MAR) (Little & Rubin, 1987), when the probability of missingness depends on the observed data, including the complete covariates, the follow-up time and/or the failure status. In that case, likelihood approaches have been presented by Lipsitz and Ibrahim (1998), Chen and Little (1999), and Herring and Ibrahim (2001). Chen (2002) added flexibility by modeling the conditional distribution of the missing covariates given the observed covariates nonparametrically, while avoiding the curse of dimensionality. Among nonlikelihood approaches, Pugh, Robins, Lipsitz, and Harrington (1993) proposed an inverse probability weighted (IPW) estimating equation, and Paik and Tsai (1997), an imputation approach. These approaches require auxiliary model specifications. The former requires correct specification of the missingness mechanism, the latter, the conditional distribution of missing covariates given the observed data. Applying the projection idea of Robins, Rotnitzky, and Zhao (1994), Wang and Chen (2001) proposed an improved IPW partial-likelihood score function, in which the two auxiliary models are parametrically specified. These estimators are doubly robust in a sense that consistency of the resulting estimator holds as long as either of the auxiliary models is specified correctly. Qi, Wang, and Prentice (2005) extended the doubly robust (DR) method to allow nonparametric auxiliary models using kernel techniques.

In applying the methods in the literature for missing time-constant covariates, we face challenges in handling missingness induced by interval censoring. In practice, a convenient solution to the missingness problem is the complete-case (CC) method, which uses only completely observed records. However, if everyone starts from free of the second event, say diabetes, the CC method eliminates records of anyone who experiences diabetes during the follow-up, because they have partially missing diabetes status between the visits. Then, the CC analysis utilizes only the subjects who have no diabetes throughout the followup, and fails to estimate the effect of time-varying diabetic status on stroke. To apply existing methods for missing time-constant covariates for the Cox models, a special feature of missingness generated by interval-censored data calls for the time-varying auxiliary models. This implies, in the case of IPW method, the models for the probability of observation should be estimated at all observed failure times. Therefore despite similarity, existing methods cannot be directly applied to the case of interval-censored covariates.

Dupuy, Grama, and Mesbah (2006) proposed a method to handle missing time-varying covariate in Cox model using a rigorous likelihood treatment. In the setup of Dupuy et al. (2006) all time-varying covariate values at the observed failure times are considered as missing, since the time-varying covariate information of a subject is not generally available at times when other cohort members fail. On the other hand, the interval-censored covariate of the diabetic status provides partial information except for the interval between which the status is changed. The approach by Dupuy et al. (2006) does not allow us to exploit such partial information.

In this article, we propose formal procedures to handle interval-censored covariate generated by an incomplete status of the secondary event in the Cox proportional hazards models. We tackle the problem as missing time-varying covariate with partially available information. We propose a DR estimator as well as the maximum likelihood estimator via EM algorithm and present their asymptotic properties. We borrow the setup of missing time-varying covariates in Cox models but derive the asymptotic properties of the DR estimators by capitalizing on the features of cohort studies, namely, that the missingness is governed by the visiting times scheduled randomly around the targeted visiting time points. The E-step of the maximum likelihood estimator also incorporates the monotone nature of the status of the secondary event. Finally, we apply the proposed methods to the motivating example, the Northern Manhattan Study (NOMAS) and discuss the clinical implication of the analysis.

2 | DESIGN AND AIM OF THE STUDY

The Northern Manhattan Study (NOMAS) is a population-based prospective cohort study designed to evaluate the effects of medical, socio-economic, and other risk factors on the incidence of vascular disease in a stroke-free multiethnic community cohort. A total of 3298 stroke-free participants aged over 39 were recruited from the Northern Manhattan community through random digit dialing (see Banerjee et al., 2012). The average age was 69.2 years, and 62.8% were women; 53.8% Hispanic, 24.7% non-Hispanic black, and 21.4% non-Hispanic white. Data regarding baseline status and risk factors were collected through interviews of participants. Subjects were scheduled to have annual evaluations of risk factors until the end of the study. Some of the risk factors are binary irreversible events such as diabetes. In NOMAS with a median follow-up of 9.46 years, a significant proportion of participants who were free of the irreversible event at baseline experienced the event over the period of follow-up. This provided some information on time to diabetes in addition to time to stroke.

Diabetes is a well-known risk factor for ischemic stroke. Diabetes may increase the risk of stroke through increased atherosclerosis, particularly of the carotid and other neck arteries, endothelial dysfunction, cardiac dysfunction, and clotting abnormalities. While diabetes is an established risk factor for stroke, fewer studies investigated the effect of the severity of diabetes measured by duration. For example, whether the change of diabetic status triggers a stroke or prolonged diabetic condition is a risk factor is not known. This question can be addressed by directly estimating the hazard ratio of the duration of diabetes. Alternatively, one can assess the effect of duration by indirectly estimating the risk of the current diabetes status. If a longer duration increases the risk, the effect of the current diabetic status would be weaker than that of the baseline diabetes since subjects who are diabetic at baseline would have longer diabetic duration. Banerjee et al. (2012) investigated the effect of severity of diabetes directly by using the duration of diabetes as a covariate. However, directly evaluating the effect of duration is challenging mainly due to three hurdles. First, for those who were diabetic at the time of enrolment, the onset time of diabetes needed to be retrospectively ascertained by recall. Retrospective ascertainment of the onset time of diabetes is prone to recall bias. Second, time to diabetes is left truncated in that a subject who developed diabetes but died before the start of the study cannot be sampled in the cohort. Left truncation of diabetes time should be accounted for a valid inference even when

censoring is accounted. Third, the duration is an imperfect measurement even for subjects who were free of diabetes and then developed diabetes during follow-up since we only know the diabetic status at the time of interview and therefore interval censored.

Our goal is to evaluate the effect of severity of diabetes on stroke. Our strategy is to estimate the effect of the current status of diabetes, which has advantages of circumventing the first two hurdles of the direct approach. The effect of the current status of diabetes on stroke is not reported in the literature. In evaluating the effect of the current status, we still have to handle the third hurdle of the interval-censored covariate. We propose the methods to handle interval-censored covariates in next three Sections.

3 | SETUP AND AVAILABLE DATA ESTIMATOR

Let $T = \min(T^0, C)$ and $\delta = I(T^0 < C)$, where T^0 is the time to failure, in our example, time to stroke, and C is the censoring time. Assume that C is independent of T^0 given $\mathbf{Z}$, a $p \times 1$ covariate vector. Let (t) be the counting process, $(t) = \delta \cdot I(T \leq t)$, and $Y(t)$ be the at-risk indicator, $Y(t) = I(T > t)$. We denote the failure time of the secondary event, in our example, time to onset of diabetes from the entry to the study by X and the counting process of X , by $D(t)$, $D(t) = I(X \leq t)$. If a subject is already diabetic before the entry to the study, then $X < 0$, and $(t) = 1$ for $t \in (0, \tau]$. Let D^* take value 0 if $(X < 0)$, 1, otherwise. Let $\mathbf{J}_i = (J_{i1}, J_{i2}, \dots, J_{ik})$ be the visit times from the entry to the study. We do not observe X_i but if the diabetes occurs, we know the two visit times around the onset. Let I_i be a random interval during which the onset of diabetes occurs. That is, $I_i = (J_{i,j-1}, J_{i,j}]$ if $J_{i,j-1} < X_i \leq J_{i,j}$ where $J_{i,0} = 0$ and $J_{i,k+1} = \tau$. When $D_i^* = 0$, we consider I_i as empty set. We assume

$(T_i, \delta_i, \mathbf{Z}_i, X_i, D_i^*, \mathbf{J}_i)$, $i = 1, \dots, n$, are independently identically distributed, and the censoring time C is independent of $(T^0, X, \mathbf{J}, D^*)$, given $\mathbf{Z}$. Without secondary event, this assumption reduces to the ones commonly used in survival analysis. We also assume that visit times, $\mathbf{J}$ is independent of $(T^0, X, C, D^*, \mathbf{Z})$, and call it the assumption of random visit times (RVT). In practice visit times are scheduled on regular intervals with small random errors, and the RVT assumption is plausible. Due to the assumption of RVT, we have $\mathbf{I}_i \perp (T^0, C, \mathbf{Z})$ given $D^* = 1$, where $\perp$ implies independence. We denote lower and upper values of random interval I_i by U_i and V_i , respectively. When $I_i = (0, V_i]$, that is X_i is right censored, the assumption reduces to the usual independence assumption between failure and censoring times.

For the time to stroke given diabetes status, we consider the proportional hazard model with the hazard function $\lambda(t|\mathbf{Z}, D(t)) = \lambda_0(t)\exp(\mathbf{Z}^T\beta_1 + D(t)\beta_2)$ and the density function is

$$f(T^0|\mathbf{Z}, D(t); \beta) = \lambda(t|\mathbf{Z}, D(t))\exp\left(-\int_0^t \lambda(s|\mathbf{Z}, D(s))ds\right).$$

Our parameter of interest is (β_1, β_2) . If (t) is completely observed, the partial likelihood inference can be used (Anderson & Gill, 1982; Cox, 1975). Since (t) is not fully observed, we need an auxiliary model assumption on time to diabetes. With the hazard function

$h(t|\mathbf{Z}) = h_0(t)\exp(\mathbf{Z}^T\gamma)$ and $H_0(t) = \int_0^t h_0(s)ds$, we postulate the density of X given $D^* = 1$, to be

$$g(X|\mathbf{Z};\gamma) = h(X|\mathbf{Z})\exp\left\{-H_0(X)\exp(\mathbf{Z}^T\gamma)\right\}.$$

We suppress conditioning argument $D^* = 1$ for brevity.

Since $X_j \in I_j$ by definition, (t) is not observed if $t \in I_j$ and $D_j(t)$ is observed if $t \notin I_j$. We denote the observation indicator by $R_j(t) = \mathbb{I}(t \notin I_j)$. Let $P(R_j(t) = 1 | \mathbf{Z}_j, T_j, \delta_j, D_j^*)$ be denoted by $\pi_j(t)$. Most literature on missing covariate in Cox models deals with the case when R_j and π_j do not depend on t . Xu et al. (2010) used time-varying (t) in modeling the time-constant R_j to improve the efficiency of the estimator of β .

A common practice is to use the status evaluated at the closest time prior t (LOCF, last observations carried forward), which leads to a biased estimator. As discussed in Section 1, the CC analysis discards any record with incomplete information, and leads to a degenerate case or extremely inefficient estimator. Alternatively, we can use the available data (AD), where the subjects' contribution to the risk set is dropped when the status is unknown. This approach utilizes partially available information when time-varying status is known as follows:

$$U_{AD}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau R_i(u) \left\{ \frac{\mathbf{Z}_i}{D_i(u)} - \frac{\mathbf{S}_{AD}^{(1)}(\beta, u)}{\mathbf{S}_{AD}^{(0)}(\beta, u)} \right\} dN_i(u),$$

where, for $k = 0, 1$,

$$\mathbf{S}_{AD}^{(k)}(\beta, u) = \frac{1}{n} \sum_{j=1}^n \left[R_j(u) Y_j(u) \left(\frac{\mathbf{Z}_j}{D_j(u)} \right)^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) \right]$$

and $A^{\otimes k}$ is the k -th tensor power of matrix A . The estimating function has a form of the observed minus the expected covariate weighted by the observation indicator, (u) , whenever the failure occurs. The AD method is different from complete-record analysis since it includes the subject in the risk set only when the current status is known. Therefore, the same subject can be in the risk set at one time point but not in the risk set at another time point if the current status is unknown. Ignoring contribution from subjects during t such that $(t) = 0$, the AD method may lose information, but yields a consistent estimator due to the RVT assumption. The efficiency loss can be more damaging when (t) is unknown for the failed subject than for subjects in the risk set. The AD estimator can be computed with minor coding effort using popular software such as SAS and R. As shown in Section 6, there is loss of efficiency using this method, and the loss of efficiency tends to be governed by $\max_t \sum_{i=1}^n R_i(t)/n$.

4 | DOUBLY ROBUST METHOD

We propose a doubly robust method for interval-censored diabetes status. Due to the RVT assumption, I_i is independent of $(\mathbf{T}_i, \delta_i)$ given $D_i^* = 1$, thus $\pi_i(t)$ does not depend on $(\mathbf{T}_i, \delta_i, \mathbf{Z}_i)$. Since $P(R_i(t) = 1 | \mathbf{Z}_i, T_i, \delta_i, D_i^* = 0) = 1$, in practice we need to estimate $P(R_i(t) = 1 | \mathbf{Z}_i, T_i, \delta_i, D_i^* = 1)$, say $\pi_i^*(t)$ whenever failure occurs. Let $t_1 < t_2 < \dots < t_m$ be m distinct failure times and $\mathbf{W}_i = (\mathbf{T}_i, \delta_i, \mathbf{Z}_i)$. For a concrete example, we postulate a parametric logistic model for $\pi^*(t)$ indexed by $\alpha(t)$, which can be obtained by solving

$$\mathbf{U}^\alpha \left\{ \alpha(t) \right\} = \frac{1}{n} \sum_{i=1}^n D_i^* (\mathbf{1}, \mathbf{W}_i^T)^T \left\{ R_i(t) - \pi_i^*(t; \alpha(t)) \right\} = 0 \quad (1)$$

where

$$\pi_i^*(t; \alpha(t)) = \left\{ 1 + \exp(-\mathbf{W}_i^T \alpha(t)) \right\}^{-1}$$

and $t = t_1, t_2, \dots, t_m$. Let $\mathbf{U}^\alpha(\alpha) = [\mathbf{U}^\alpha\{\alpha(t_1)\}^T, \mathbf{U}^\alpha\{\alpha(t_2)\}^T, \dots, \mathbf{U}^\alpha\{\alpha(t_m)\}^T]^T$ and $\hat{\pi}(t) = \pi^*\{t; \hat{\alpha}(t)\}$, if $D^* = 1$, 1, if $D^* = 0$. The inverse probability weighting (IPW) estimating function for β can be constructed with time-varying $\{t, \alpha(t)\}$ as follows:

$$\mathbf{U}_{IPW}^\beta(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{R_i(u)}{\pi_i(u)} \left\{ \begin{pmatrix} \mathbf{Z}_i \\ D_i(u) \end{pmatrix} - \frac{\mathbf{S}_{IPW}^{(1)}(\beta, u)}{\mathbf{S}_{IPW}^{(0)}(\beta, u)} \right\} dN_i(u),$$

where for $k = 0, 1$,

$$\mathbf{S}_{IPW}^{(k)}(\beta, u) = \frac{1}{n} \sum_{j=1}^n \left[\frac{R_j(u)}{\pi_j(u)} Y_j(u) \begin{pmatrix} \mathbf{Z}_j \\ D_j(u) \end{pmatrix}^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) \right].$$

Comparing with $\mathbf{U}_{AD}$, the estimating function $\mathbf{U}_{IPW}^\beta$ obtains the expected term using the inverse probability weighting to represent a sum of covariates from everyone in the risk set. The contribution at each failure time is also weighted by the inverse probability of observation, $R_i(u)/\pi_i(u)$.

Since the available data analysis yields a consistent estimator, the IPW method does not play the usual role of reducing bias. To gain the efficiency from the IPW estimating equation, Robins et al. (1994) proposed to subtract the projection onto the nuisance tangent space. Wang and Chen (2001) implemented the method in the Cox model with missing covariates when missing covariates and weights are time-constant. Xu, Paik, Luo, and Tsai (2009) studied asymptotic behavior of the DR-type estimator in the case of time-constant covariates and time-varying weights. In the case of interval censored covariates, both covariates and

weights are time-varying, and the projection term is slightly different from those of Wang and Chen (2001) or Xu et al. (2009) to accommodate time-varying nature of $\pi(t)$. Let $d M_i(u) = d N_i(u) - d \Lambda_0(u) \exp(\mathbf{Z}_i \beta_1 + D_i(u) \beta_2)$ and $\alpha = (\alpha(t_1)^T, \alpha(t_2)^T, \dots, \alpha(t_m)^T)^T$. By applying the general framework of projection theory, the augmented estimating function, $\mathbf{U}_A^\beta(\beta)$ can be obtained by

$$\mathbf{U}_A^\beta(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{R_i(u)}{\pi_i(u)} \left(\begin{pmatrix} \mathbf{Z}_i \\ D_i(u) \end{pmatrix} - \frac{\mathbf{S}_A^{(1)}(\beta, u)}{\mathbf{S}_A^{(0)}(\beta, u)} \right) d M_i(u) - \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i(\beta),$$

where

$$\mathbf{A}_i(\beta) = \int_0^\tau \frac{R_i(u) - \pi_i(u)}{\pi_i(u)} \left\{ E \left[\begin{pmatrix} \mathbf{Z}_i \\ D_i(u) \end{pmatrix} d M_i(u) \middle| \mathbf{W}_i \right] - \frac{\mathbf{S}_A^{(1)}(\beta, u)}{\mathbf{S}_A^{(0)}(\beta, u)} E[d M_i(u) | \mathbf{W}_i] \right\},$$

and for $k = 0, 1$,

$$\begin{aligned} \mathbf{S}_A^{(k)}(\beta, u) = & \frac{1}{n} \sum_{j=1}^n \left\{ \frac{R_j(u)}{\pi_j(u)} Y_j(u) \begin{pmatrix} \mathbf{Z}_j \\ D_j(u) \end{pmatrix}^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) + \right. \\ & \left. - \frac{R_j(u) - \pi_j(u)}{\pi_j(u)} Y_j(u) \times E \left[\begin{pmatrix} \mathbf{Z}_j \\ D_j(u) \end{pmatrix}^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) \middle| \mathbf{W}_j \right] \right\}. \end{aligned}$$

In using notation $E(\cdot | \mathbf{W})$, we drop the conditioning argument $D^* = 1$. If $D^* = 0$, $D(t) = 1$ for $t \in [0, \tau)$. The estimating function $\mathbf{U}_A^\beta$ subtracts the projection of $\mathbf{U}_{I P W}^\beta$ onto the nuisance tangent space. The resulting estimator is unbiased when either the (u) or $E\{D(u) | \mathbf{W}\}$ is correctly specified.

The augmented estimating function for β , $\mathbf{U}_A^\beta(\beta)$, depends on three nuisance parameters, Λ_0 , α and γ . For clarity we rewrite $\mathbf{U}_A^\beta(\beta) \equiv \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$. First, we replace unknown $d\Lambda_0(t)$ in $\mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ by

$$d\hat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^n \frac{d N_i(t)}{\mathbf{S}_A^{(0)}(\beta, t)}.$$

Second, α can be replaced by the solution of (1). Third, we propose to estimate the projection term involving the conditional expectation of (t) given $\mathbf{W}$ and $D^* = 1$ via parametric modeling of X . Denoting the survivor function of X given $D^* = 1$ by $(\cdot)$ indexed by γ , we estimate γ via maximizing the following likelihood function for interval-censored data,

$$L^{\gamma}(\gamma) = \prod_{i=1}^n \{G(U_i | \mathbf{Z}_i; \gamma) - G(V_i | \mathbf{Z}_i; \gamma)\}, \quad (2)$$

where $G(t | \mathbf{Z}; \gamma) = \exp\{-H_0(t; \gamma_0) \exp(\mathbf{Z}^T \gamma_1)\}$, and $\gamma = (\gamma_0^T, \gamma_1^T)^T$. For a concrete example, we use Weibull model, $H_0(t; \gamma_0) = \gamma_{01} t^{\gamma_{02}}$ and $\gamma_0 = (\gamma_{01}, \gamma_{02})^T$. It is straightforward to extend to semiparametric proportional hazard models. For $U = 0$, X is left censored, for $V = \infty$, X is right censored. Let $\mathbf{U}^{\gamma}(\gamma) = n^{-1} \partial \log L^{\gamma}(\gamma) / \partial \gamma$. The conditional expectation $E[(\cdot) | \mathbf{W}]$ depends on β and γ as well as whether $T \leq t$ or not, as given in the Appendix A.2. Finally let $\hat{\beta}(\hat{\alpha}, \hat{\gamma}, \hat{\Lambda}_0)$ be the solution of

$$\mathbf{U}_A(\beta, \alpha, \gamma) = \begin{pmatrix} \mathbf{U}_A^{\beta}(\beta, \alpha, \gamma, \hat{\Lambda}_0) \\ \mathbf{U}^{\gamma}(\gamma) \\ \mathbf{U}^{\alpha}(\alpha) \end{pmatrix} = 0.$$

Newton–Raphson algorithm is used to compute $(\hat{\beta}, \hat{\gamma}, \hat{\alpha})$.

Theorem 4.1. *Under the regularity conditions, $\hat{\beta}(\hat{\alpha}, \hat{\gamma}, \hat{\Lambda}_0)$ is consistent for the true parameter β , and*

$$n^{1/2}(\hat{\beta}(\hat{\alpha}, \hat{\gamma}, \hat{\Lambda}_0) - \beta) \xrightarrow{d} \mathbf{N}(0, \mathbf{I}_{\beta}^{-1}(\mathbf{V}_{I P W} - \mathbf{V}_1) \mathbf{I}_{\beta}^{-1}),$$

where

$$\mathbf{I}_{\beta} = \int \left[\frac{s^{(2)}(\beta, u)}{s^{(0)}(\beta, u)} - \left(\frac{s^{(1)}(\beta, u)}{s^{(0)}(\beta, u)} \right)^{\otimes 2} \right] s^{(0)}(\beta, u) d\Lambda_0(u),$$

$$\mathbf{V}_{I P W} = E \left[\int 1/\pi(u; \alpha(u)) \left[\left(\frac{\mathbf{Z}_i}{D_i(u)} - \mathbf{e}(\beta, u) \right) dM_i(u) \right]^{\otimes 2} \right],$$

$$\mathbf{V}_1 = E \left[\int \frac{1 - \pi(u, \mathbf{W}_i; \alpha)}{\pi(u, \mathbf{W}_i; \alpha)} \left[E \left[\left(\frac{\mathbf{Z}_i}{D_i(u)} dM_i(u) \middle| \mathbf{W}_i \right) - \mathbf{e}(\beta, u) E[dM_i(u) | \mathbf{W}_i] \right]^{\otimes 2} \right],$$

and $\mathbf{e}(\beta, u) = s^{(1)}(\beta, u)/s^{(0)}(\beta, u)$ with

$$\mathbf{s}^{(k)}(\beta, t) = \mathbb{E} \left[Y_i(t) \begin{pmatrix} \mathbf{Z}_i \\ D_i(t) \end{pmatrix}^{\otimes k} \exp(\mathbf{Z}_i^T \beta_1 + D_i(t) \beta_2) \right].$$

The asymptotic properties of the resulting estimator closely resemble the case Xu et al. (2009) and we present a brief sketch of oofs in Appendix A.1. As in Xu et al. (2009), the asymptotic property of $n^{1/2} \hat{\beta}(\hat{\alpha}, \hat{\gamma}, \hat{\Lambda}_0)$ is the same as that of $n^{1/2} \hat{\beta}(\alpha, \gamma, \Lambda_0)$. The estimator is also doubly robust in that the estimator is consistent when either of the two auxiliary models, $\pi^*(t; \alpha(t))$ or $G(t|\mathbf{Z}, \gamma)$ is correctly specified. Under the RVT assumption, any specified π model includes the true model, and it is guaranteed that the estimator is consistent without correct specification of the model for X . A correct model for X renders semiparametric efficiency of the estimator.

5 | LIKELIHOOD METHOD

We present the maximum likelihood estimator and the EM algorithm when the diabetes status is interval censored. As in Chen and Little (1999) for the missing time-constant covariate for Cox models, we maximize the likelihood for the observed data jointly over parameters (β^T, γ^T) and Λ_0 . A key feature in the EM algorithm is to account for monotonicity of (t) . Let $t_1 < t_2 < \dots < t_m$ be m distinct failure times with $t_0 = 0$. If $(t_{l-1}) = 0$ and $D_i(t_l) = 1$, the diabetes occurred between t_{l-1} and t_l , that is $t_{l-1} < X \leq t_l$. Further, we use notations $D_{il} = (t_l)$, and $\mathbf{D}_i = (D_{i1}, \dots, D_{im})$. There is monotonicity restriction in sequence of 's, and all possible realizations of $\mathbf{D}_i$ has a form $\mathbf{D}_i = (\mathbf{0}_{k-1}, \mathbf{1}_{m-k+1})$, and $\mathbf{D}_i$ carries information at which distinct failure time the status of $D_i(t)$ changes. Since nonparametric maximum likelihood estimator of Λ_0 has jumps at distinct failure times, it is convenient to let $\lambda_0 = ((t_1), (t_2), \dots, \lambda(t_m))^T$. Also let $\theta = (\lambda_0^T, \beta^T, \gamma^T)^T$. The complete data consist of $(T, \delta, \mathbf{Z}, X)$, and the observed data $(T, \delta, \mathbf{Z}, I)$. If complete data are available, we can maximize $L(\theta; T, \delta, \mathbf{D}, \mathbf{Z}, I)$, where

$$\begin{aligned} L(\theta; T, \delta, \mathbf{D}, \mathbf{Z}, I) &= \prod_{i=1}^n L_i(\theta; T_i, \delta_i, \mathbf{D}_i, \mathbf{Z}_i, I_i) = \\ &= \prod_{i=1}^n [\tilde{f}(T_i, \delta_i | \mathbf{D}_i; \beta, \lambda_0) \times P(\mathbf{D}_i | \mathbf{Z}_i, I_i; \gamma)], \end{aligned}$$

where

$$\begin{aligned} \tilde{f}(T_i, \delta_i | \mathbf{D}_i; \beta, \lambda_0) &= \left\{ \lambda_0(T_i) \exp(\mathbf{Z}_i^T \beta_1 + D_i(T_i) \beta_2) \right\}^{\delta_i} \times \\ &\times \exp \left\{ - \sum_{l=1}^m Y_i(t_l) \lambda_0(t_l) \exp(\mathbf{Z}_i^T \beta_1 + D_i(t_l) \beta_2) \right\}. \end{aligned}$$

We denote the smallest failure time from which D_{ij} is missing by

$$t_{a_i+1} = \min \left\{ t_j | t_j \geq U_i, 1 \leq j \leq m \right\} \text{ and the largest failure time, } t_{b_i} = \max \left\{ t_j | t_j < V_i, 1 \leq j \leq m \right\}.$$

Then, D_{ij} can be written as

$$D_{il} = \begin{cases} 0 & l = 1, 2, \dots, a_i \\ \text{unknown} & l = a_i + 1, \dots, b_i \\ 1 & l = b_i + 1, \dots, m. \end{cases}$$

We let the observed part of $\mathbf{D}_i$ by $\mathbf{D}_i^o$ and missing part of $\mathbf{D}_i$ by $\mathbf{D}_i^m = (D_{ia_i+1}, \dots, D_{ib_i})$. Due

to monotonicity of $\mathbf{D}_i$, there are only $b_i - a_i + 1$ possible realizations of $\mathbf{D}_i^m$, which can be expressed as $\mathbf{d}_i^k = (\mathbf{0}_k, \mathbf{1}_{b_i - a_i - k})$, for $k = 0, \dots, b_i - a_i$. Since $D_i(t) = I(X_i \leq t)$,

$P(\mathbf{D}_i | \mathbf{Z}_i, I_i; \gamma) = P(\mathbf{D}_i^m | \mathbf{Z}_i, I_i; \gamma)$. Then the observed likelihood function is

$$\prod_{i=1}^n \sum_{k=0}^{b_i - a_i} \left[\tilde{f}(T_i, \delta_i | \mathbf{D}_i^o, \mathbf{D}_i^m = \mathbf{d}_i^k; \beta, \lambda_0) P(\mathbf{D}_i^m = \mathbf{d}_i^k | \mathbf{Z}_i, I_i; \gamma) \right],$$

where

$$\begin{aligned} P(\mathbf{D}_i^m = \mathbf{d}_i^0 | \mathbf{Z}_i; \gamma) &= P(U_i < X_i \leq t_{a_i+1} | \mathbf{Z}_i) = \\ &= G(U_i | \mathbf{Z}_i; \gamma) - G(t_{a_i+1} | \mathbf{Z}_i; \gamma), \end{aligned}$$

$$\begin{aligned} P(\mathbf{D}_i^m = \mathbf{d}_i^k | \mathbf{Z}_i; \gamma) &= P(t_{j-1} < X_i \leq t_j | \mathbf{Z}_i) = \\ &= G(t_{j-1} | \mathbf{Z}_i; \gamma) - G(t_j | \mathbf{Z}_i; \gamma), \end{aligned}$$

for $k = 1, \dots, b_i - a_i - 1$, and

$$\begin{aligned} P(\mathbf{D}_i^m = \mathbf{d}_i^{b_i - a_i} | \mathbf{Z}_i; \gamma) &= P(t_{b_i} < X_i \leq V_i | \mathbf{Z}_i) = \\ &= G(t_{b_i} | \mathbf{Z}_i; \gamma) - G(V_i | \mathbf{Z}_i; \gamma). \end{aligned}$$

The resulting score function has a form

$$\sum_{i=1}^n \sum_{k=0}^{b_i - a_i} \left[\left\{ \frac{\partial}{\partial \theta} \log \tilde{f}(T_i, \delta_i | \mathbf{D}_i^o, \mathbf{D}_i^m = \mathbf{d}_i^k; \beta, \lambda_0) \right\} \times P(\mathbf{D}_i^m = \mathbf{d}_i^k | T_i, \delta_i, \mathbf{Z}_i, I_i; \gamma) \right], \quad (3)$$

where

$$\begin{aligned} P(\mathbf{D}_i^m = \mathbf{d}_i^k | T_i, \delta_i, \mathbf{Z}_i, I_i; \theta) &= \\ &= \frac{\tilde{f}(T_i, \delta_i | \mathbf{D}_i^O, \mathbf{D}_i^m = \mathbf{d}_i^k; \beta, \lambda_0) P(\mathbf{D}_i^m = \mathbf{d}_i^k | \mathbf{Z}_i, I_i; \gamma)}{\sum_{j=1}^{b_i - a_i} \tilde{f}(T_i, \delta_i | \mathbf{D}_i^O, \mathbf{D}_i^m = \mathbf{d}_i^j; \beta, \lambda_0) P(\mathbf{D}_i^m = \mathbf{d}_i^j | \mathbf{Z}_i, I_i; \gamma)}. \end{aligned}$$

Furthermore, let D_{ik}^m be the k -th element of $\mathbf{D}_i^m$. Then

$$\begin{aligned} P(D_{ik}^m = 1 | T_i, \delta_i, \mathbf{Z}_i, I_i; \theta) &= \\ &= E(D_{a_i+k} | T_i, \delta_i, \mathbf{Z}_i, I_i; \theta) = \\ &= \frac{\sum_{j=1}^k \tilde{f}(T_i, \delta_i | \mathbf{D}_i^O, \mathbf{D}_i^m = \mathbf{d}_i^j; \beta, \lambda_0) P(\mathbf{D}_i^m = \mathbf{d}_i^j | \mathbf{Z}_i; \gamma)}{\sum_{j=1}^{b_i - a_i} \tilde{f}(T_i, \delta_i | \mathbf{D}_i^O, \mathbf{D}_i^m = \mathbf{d}_i^j; \beta, \lambda_0) P(\mathbf{D}_i^m = \mathbf{d}_i^j | \mathbf{Z}_i; \gamma)}. \end{aligned}$$

The conditional expectation shows that the whole array of $\mathbf{D}_i^m$ is evaluated not the individual missing value at $D_{ik}(t)$ to take into account irreversible nature of the covariate. Explicit forms of the conditional mean are given in Appendix A.2, which show that the conditional mean depends on whether t is smaller than T or not, due to time dependence of $D(t)$.

The first m elements of equation (3) gives a closed form solution for $\hat{\lambda}_0$,

$$d\hat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^n \frac{d N_i(t)}{S_L^{(0)}(\beta, t)} \quad (4)$$

where

$$\begin{aligned} S_L^{(k)}(\beta, u) &= \frac{1}{n} \sum_{j=1}^n \left\{ R_j(u) Y_j(u) \left(\frac{\mathbf{Z}_j}{D_j(u)} \right)^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) + \right. \\ &\quad \left. + (1 - R_j(u)) Y_j(u) \times E \left[\left(\frac{\mathbf{Z}_j}{D_j(u)} \right)^{\otimes k} \exp(\mathbf{Z}_j^T \beta_1 + D_j(u) \beta_2) | \mathbf{Q}_j \right] \right\} \end{aligned}$$

with $\mathbf{Q}_j = (T_j, \delta_j, \mathbf{Z}_j, I_j)$. Plugging in (4) into the score function (3) corresponding to the derivative with respect to β , we obtain

$$U_L^{\beta}(u) = \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \left\{ R_i(u) \left(\frac{\mathbf{Z}_i}{D_i(u)} \right) + (1 - R_i(u)) \left(\frac{\mathbf{Z}_i}{E[D_i(u) | \mathbf{Q}_i]} \right) - \frac{\mathbf{S}_L^{(1)}(\beta, u)}{\mathbf{S}_L^{(0)}(\beta, u)} \right\} dN_i(u).$$

We bring attention to the fact that $\mathbf{Q}_i$ is different from $\mathbf{W}_i$ in that $\mathbf{Q}_i$ includes I_i .

The resulting estimator is consistent and asymptotically normal and we obtained the variance of the estimator via Louis' formula (1982), while the conditional variance of the score was computed via Monte Carlo method by sampling from $\mathbf{d}_i^k$, $k = 0, \dots, b_i - a_i$ with probability $P(\mathbf{D}_i^m = \mathbf{d}_i^k | T_i, \delta_i, \mathbf{Z}_i; \gamma)$.

6 | SIMULATION STUDIES

We conducted simulation studies to evaluate finite sample properties of the proposed DR estimator and the MLE. A common practice is to impute the covariate measured in last previous time, which we refer to the LOCF estimator. We thus compared (i) the AD estimator, (ii) the LOCF estimator imputing observed $D(t)$ in the closest previous time, (iii) the DR estimator, (iv) the MLE described in Section 4, and finally (v) also evaluate full data estimator (when there are no missing $D(t)$) for the correctly specified model. We conducted 500 replications for sample size of $n = 300$ for all settings.

The unobserved X is simulated from two distributions; (i) the Weibull distribution $\gamma_{01} = 0.05$, $\gamma_{02} = 3$, $\gamma_1 = 1$, whose explicit form is given in (2), and (ii) the log-logistic distribution with scale parameter with survivor function $G(t) = \{1 + 0.005 \exp(Zt^2)\}^{-1}$, where, in both, the covariate Z is from $N(0, 1)$. In the study below, we assumed that X is from a Weibull model for both cases to assess the performance, when the X model is correctly specified versus misspecified.

We generated thirty intervisit times by adding uniformly distributed noises to a constant, and set their cumulative sums as visiting times, J_j , $j = 1, 2, \dots, 30$. We let $U_i = \max\{J_j : J_j < X_i, 10^{-3}\}$ and $V_i = \min\{J_j : J_j > X_i, 999\}$ for the i -th subject for interval censoring time. The average missing rate over 500 replications was either approximately 30% or 40% for both Weibull and log-logistic model. The missing rate for a single dataset was $\max_t \sum_{i=1}^n R_i(t)/n$ as defined at the end of Section 2.

For the stroke time T , we assumed the proportional hazard model with the hazard rate $\lambda(t) = \lambda_0(t) \cdot \exp(Z\beta + D(t)\beta_2)$, where $D(t) = 1$ if $t \leq X$ and 0 otherwise. We set $\lambda_0(t) = 0.1$ and $(\beta_1, \beta_2) = (-\log(2), \log(2))$. The censoring time C was generated from an exponential distribution with proper rate parameter so that the resulting censoring proportion is about 20%.

Table 1 shows the results for the correct specification of the model for diabetes. All estimates except for the LOCF show negligible bias, and also are close to the FULL, without any missing data. The LOCF estimator, which replaces missing (t) with that evaluated at the closest previous observation time, was significantly biased for β_2 . Since the AD estimator is

consistent, we recommend the AD over the LOCF estimator in practice. The MLE is the most efficient, but the DR estimator also improves the efficiency over the AD estimator.

Table 2 shows the performance of the estimators when the model for diabetes is misspecified. The true distribution of X is from the log-logistic but the specified model is the Weibull. The LOCF estimate is still biased. We find that the bias of the EM estimator increases as the model for diabetes is misspecified. The bias of the EM estimate becomes particularly large in β_2 and its magnitude increases as the missing proportion is increased. The DR estimator shows smaller bias than the EM estimate, and again substantial improvement in efficiency from the AD estimator, reducing 17% of the variance for $\hat{\beta}_2$ and 25% to 31% for $\hat{\beta}_1$.

Source code (R code) to reproduce the results of Table 1 for 31% missing proportion is available as Supporting Information on the journal's web page.

7 | ANALYSIS OF NORTHERN MANHATTAN STUDY

As described in Section 2, the NOMAS is a population-based prospective cohort study of stroke-free 3298 subjects from a multiethnic community. Our analysis included 3021 participants after deleting subjects with the ethnic origin other than white, black, and Hispanic, and with missing other key covariates. At the time of this analysis, 231 had an ischemic stroke out of 3021 subjects. The median follow-up was 9.46 years. The NOMAS collected diabetes status through interviews at scheduled annual visits yielding the interval-censored diabetes status. Six hundred fifty-five subjects (21.7 %) had diabetes at baseline, and additionally, 321 subjects (10.6 %) developed diabetes during the follow-up. The median duration of two visits between which diabetes occurred, $V_i - U_i$ was 1.3 year. Figure 1 shows the number of missing $D(t)$ at each failure time in the risk set, $\sum_{i=1}^n (1 - R_i(t))Y_i(t)$. Figure 2 shows Kaplan–Meier curves of three groups, who had diabetes at baseline, who developed diabetes during follow-up, and who stayed nondiabetic throughout the follow-up. The estimated survival function was distinctly worse for those who were diabetic at baseline than the other two and the best for those who stayed nondiabetic during the follow-up. Subjects who developed diabetes had slightly worse survival later in the follow-up. We applied methods discussed in Sections 3–5 to evaluate the effect of the severity of diabetes through the current status of diabetes.

Table 3 shows the four estimates described in Section 6. To compute the DR estimate and the MLE, the model for the time to diabetes are estimated assuming a Weibull distribution. We did not include the duration retrospectively ascertained for the prevalent diabetics. The estimated γ_1 (standard errors) for Black, Hispanic, gender, and age were $\{0.385(0.221), 1.080(0.189), 0.081(0.116), -0.008(0.006)\}$ with $\hat{\gamma}_{01} = 0.006(0.003)$ $\hat{\gamma}_{02} = 1.129(0.071)$, indicating Black and Hispanic had significantly higher risk of developing diabetes than White.

In the stroke model, two race-ethnicity, Black (BLACK), Hispanic (HISP), along with Man (MALE), Moderate alcohol consumption (etmod), Any cardiac disease (NCAD), Systolic

(SYSTOLIC), Former smoker (SMOKER1), Current smoker (SMOKER2), HDL (LHDL), and LDL (LLDL), $D(t)$ and interaction between race and systolic blood pressure are included as covariates. All five methods yielded similar results, identifying age, Male, NCAD, SMOKER2, and (t) as significant predictors of stroke. Although not reported in the Table, when (t) was replaced by the baseline diabetes status, the coefficient estimate (standard error) for the baseline diabetes was .930 (.139), giving hazard ratio around 2.5. The estimate of (t) shown in Table 3 are only slightly smaller than that of the baseline diabetes. The EM estimator could be biased when the Weibull assumption for the time to the secondary event is violated, but the difference between the EM and the DR estimators was negligible possibly because the effect of (t) was dominated by subjects who were diabetic at baseline as shown in Figure 2. For the prevalent diabetics, (t) is fully observed throughout the follow-up period and is not affected by the assumption on the failure time of the secondary event. Since the AD and DR estimators are unbiased under the assumptions stated earlier, we focus on the results from AD and DR. The coefficient estimates for the current status of diabetes were similar between the two, yielding hazard ratio of 2.4, but the relative efficiency of AD to DR estimate was only 66% (.113²/.139²). This suggests that the use of information from interval-censored diabetes status increases the power to detect any effect of current diabetic status on stroke.

Table 4 shows the analysis only among the subjects who do not have diabetes at baseline ($n = 2366$). The coefficient estimates for the covariates other than (t) were similar to the analysis including subjects who have diabetes at baseline. As for $D(t)$, the EM estimate was substantially larger than the rest. This may be due to a violation of the assumption for the time to the secondary event. Compared to the analysis including the prevalent diabetics shown in Table 3 where the effect of $D(t)$ was mainly driven by diabetic subjects at baseline, this analysis reflects the effect of the current status among those who newly developed diabetes. In this analysis, the effect of $D(t)$ is driven by the interval-censored variable, and the EM estimates are sensitive to the conditional expectation of missing $D(t)$ obtained based on distributional assumption. We focus on the results based on the AD and DR methods that yield unbiased estimators. As in Table 3, the point estimates of $D(t)$ from the AD and DR methods were similar, yielding hazard ratio of 1.4, but the relative efficiency of the AD to DR estimate was 47% (.235²/.344²). The estimated hazard ratio of $D(t)$ among those who were free of diabetes, 1.4, was smaller than the hazard ratio 2.4, when the prevalent diabetic were included, suggesting that the newly diagnosed diabetic subjects are less likely to develop a stroke than those with longer duration of diabetes. This supports an increased risk of diabetes as the disease advances.

8 | CONCLUSION AND DISCUSSION

In this paper, we develop a doubly robust and maximum likelihood estimators with EM algorithm to handle interval-censored covariates in Cox proportional hazards model for cohort studies. Many cohort studies have irreversible secondary events that can be used as covariates and have subjects who already experienced the event at baseline. Our proposed method can be applied broadly in this setup. The main contribution is to handle uniqueness of the missingness induced by interval-censoring that has not been explored in the literature.

Asymptotic results are obtained using the feature of random scheduling for cohort studies, and numerical studies show that the proposed methods perform as anticipated.

Our results show that widely used method LOCF can be seriously biased especially for the coefficient of the current status since it ignores the change that might have happened during the unobserved interval. On the other hand, the proposed AD is consistent by withdrawing its contribution when the current status is unknown. The AD method can be a viable option in practice when the RVT assumption is reasonable. Since most prospective studies have scheduled visit plans, adhering to the protocol would help avoid violating the RVT assumption. The assumption of RVT can be relaxed so that $R(t)$ may depend on (T, δ) , which leads $\pi(t)$ as a function of (T, δ) . This requires a new asymptotic derivation and could be a future research topic.

The maximum likelihood estimator is the most efficient estimator when the assumptions are correctly specified. When the assumption for the secondary event is violated, the maximum likelihood estimator for the current status can suffer from bias while the DR method is robust. The AD estimator is unbiased due to the assumption of RVT, but the efficiency relative to the DR estimator is lost by failing to utilize available information. In the NOMAS example, the AD and DR estimators show similar point estimates but the relative efficiency was slightly lower than 50% in Table 4, which implies that by using the DR method instead of the AD, a better precision can be obtained with only one half of the sample size used in the AD method.

Parametric assumption for diabetes time X can be extended to a Cox model, using methods such as Goetghebeur and Ryan (2000), Betensky, Lindsey, Ryan, and Wand (2002), or Heller (2011). Since the conditional expectations shown in Appendix A.2 do not depend on the hazard function for X but only on the cumulative hazard function, asymptotic properties can be obtained similarly by taking into account for the estimated baseline cumulative hazard function of X nonparametrically.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) for Paik (No. 2013R1A2A2A01067262), and National Institute of Health of United States for Sacco and Elkind (No. 6R01NS029993–23).

Funding information

National Research Foundation of Korea, Grant/Award Number: 2013R1A2A2A01067262; National Institute of Health, USA, Grant/Award Number: 6R01 NS029993–23

APPENDIX

A.1 Sketch of the proof of Theorem 1

In below, for notational simplicity, we let $\mathbf{Z}(t)^T = (\mathbf{Z}^T, D(t))$. The following regularity conditions are needed in the proofs.

C1. $\Lambda(\tau) < \infty$.

C2. $P(Y(\tau) = 1) > 0$.

C3. $\int_0^t \{s^{(2)}(\beta, t)/s^{(0)}(\beta, t) - (s^{(2)}(\beta, t)/s^{(0)}(\beta, t))^{\otimes 2}\} s^{(0)}(\beta, t) d\Lambda_0(t)$ is positive definite.

C4. $\mathbf{Z}$ is time independent and bounded.

C5. $\pi(t) - \varepsilon > 0$ for any $t \in [0, \tau]$.

C6. $\mathbf{I}_{22} = E[-\mathbf{U}\gamma/\gamma]$ is positive definite.

C7. $\mathbf{I}_{33} = E[-\mathbf{U}\alpha/\alpha]$ is positive definite.

C8. $\sup_{t \in [0, \tau]} \left\| n^{-1/2} \sum_{i=1}^n (R_i(t) - \pi_i(t))/\pi_i(t) \partial E[\mathbf{Z}_i(t) dM_i(t) | \mathbf{W}_i] / \partial \Lambda_0(t) \right\| = O_p(1)$.

Step 1: Asymptotic normality of $n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ By theorem III.1 of Anderson and Gill (1982), it follows that

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| \mathbf{S}_A^{(k)}(\beta, t) - s^{(k)}(\beta, t) \right\| \xrightarrow{a.s.} 0. \quad (\text{A.1})$$

for $k = 0, 1, 2$, where the parameter space $\mathcal{B}$ is a compact neighborhood of the true parameter β , $s^{(k)}(\beta, t)$ is bounded and $s^{(0)}(\beta, t)$ is bounded away from zero on $[0, \tau] \times \mathcal{B}$, and $s^{(k)}(\beta, t)$ is an equicontinuous family at β .

By replacing $dN_i(t)$ in $\mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ with $dM_i(t)$, $n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ is equivalent to

$$\begin{aligned} n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) &= \\ &= n^{-1/2} \sum_{i=1}^n \int \frac{R_i(u)}{\pi(u, \mathbf{W}_i; \alpha)} \left\{ \mathbf{Z}_i(u) - \mathbf{S}_A^{(1)}(\beta, u)/\mathbf{S}_A^{(0)}(\beta, u) \right\} dM_i(u) + \\ &\quad - n^{-1/2} \sum_{i=1}^n \int \frac{R_i(u) - \pi(u, \mathbf{W}_i; \alpha)}{\pi(u, \mathbf{W}_i; \alpha)} \left[E[\mathbf{Z}_i(u) dM_i(u) | \mathbf{W}_i] - \mathbf{S}_A^{(1)}(\beta, u)/\mathbf{S}_A^{(0)}(\beta, u) E[dM_i(u) | \mathbf{W}_i] \right], \end{aligned}$$

which is decomposed into

$$n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) = n^{-1/2} \mathbf{u}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) - \mathbf{B}_1 - \mathbf{B}_2,$$

Where

$$\begin{aligned} \mathbf{u}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) &= \sum_{i=1}^n \int \frac{R_i(u)}{\pi(u, \mathbf{W}_i; \alpha)} \{ \mathbf{Z}_i(u) - \mathbf{e}(\beta, u) \} d\mathbf{M}_i(u) + \\ &\quad - \sum_{i=1}^n \int \frac{R_i(u) - \pi(u, \mathbf{W}_i; \alpha)}{\pi(u, \mathbf{W}_i; \alpha)} [E[\mathbf{Z}_i(u) d\mathbf{M}_i(u) | \mathbf{W}_i] - \mathbf{e}(\beta, u) E[d\mathbf{M}_i(u) | \mathbf{W}_i]] \end{aligned}$$

with $\mathbf{e}(\beta, u) = \mathbf{s}^{(1)}(\beta, u)/\mathbf{s}^{(0)}(\beta, u)$, and

$$\mathbf{B}_1 = \int \left\{ \mathbf{s}_A^{(1)}(\beta, u)/\mathbf{s}_A^{(0)}(\beta, u) - \mathbf{e}(\beta, u) \right\} d\bar{\mathbf{M}}_n,$$

$$\mathbf{B}_2 = \int \left\{ \mathbf{s}_A^{(1)}(\beta, u)/\mathbf{s}_A^{(0)}(\beta, u) - \mathbf{e}(\beta, u) \right\} d\bar{\mathbf{M}}_n^*,$$

with

$$d\bar{\mathbf{M}}_n = n^{-1/2} \sum_{i=1}^n \frac{R_i(t)}{\pi(u, \mathbf{W}_i; \alpha)} d\mathbf{M}_i(t)$$

and

$$d\bar{\mathbf{M}}_n^* = n^{-1/2} \sum_{i=1}^n \frac{R_i(u) - \pi(u, \mathbf{W}_i; \alpha)}{\pi(u, \mathbf{W}_i; \alpha)} \mathbf{e}[d\mathbf{M}_i(u) | \mathbf{W}_i].$$

Applying (A.1) and the proof of Theorem 1 in Qi et al. (2005), $\mathbf{B}_1 \rightarrow 0$ and $\mathbf{B}_2 \rightarrow 0$ in probability. This implies that $n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ can be approximated by the sum of iid random variables. By the central limit theorem, we have

$$n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) \xrightarrow{d} N(0, \mathbf{V}_{IPW} - \mathbf{V}_1), \quad (\text{A.2})$$

where $\mathbf{V}_{IPW}$ and $\mathbf{V}_1$ are those defined in the theorem.

Applying (A.1) and Lemma 1 in Qi et al. (2005), we have

$$\sup_{\beta \in \mathcal{B}} \left\| -\frac{1}{n} \partial \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) / \partial \beta^T - \mathbf{I}_\beta \right\| \xrightarrow{a.s.} 0. \quad (\text{A.3})$$

The weak convergence of $n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ implies that $n^{-1} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) \xrightarrow{a.s.} 0$. $\mathbf{I}_\beta$ is positive definite by condition (C3). From the same argument as in the proof of Theorem 2 of

Foutz (1977), $\hat{\beta}_A(\alpha, \gamma, \Lambda_0)$ exists and is unique in $\mathcal{B}$ with probability converging to 1 as $n \rightarrow \infty$, and $\hat{\beta}_A(\alpha, \gamma, \Lambda_0) \xrightarrow{P} \beta$.

Finally, by the Taylor expansion of $n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)$ at the true parameter β , we can get

$$n^{1/2}(\hat{\beta}_A(\alpha, \gamma, \Lambda_0) - \beta) = \left(-\frac{1}{n} \frac{\partial \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0)}{\partial \beta^T} \bigg|_{\beta = \beta^*} \right)^{-1} n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0), \quad (\text{A.4})$$

where β^* lies between true parameter β and $\hat{\beta}_A(\alpha, \gamma, \Lambda_0)$. The asymptotic normality of $n^{1/2}(\hat{\beta}_A(\alpha, \gamma, \Lambda_0) - \beta)$ follows from (A.2), (A.3), and (A.4).

Step 2: Asymptotic equivalence between $n^{1/2} \hat{\beta}(\alpha, \gamma, \hat{\Lambda}_0)$ and $n^{1/2} \hat{\beta}(\alpha, \gamma, \Lambda_0)$

Now we show asymptotic equivalence between $n^{1/2} \hat{\beta}(\alpha, \gamma, \hat{\Lambda}_0)$ and $n^{1/2} \hat{\beta}(\alpha, \gamma, \Lambda_0)$ following steps of Xu et al. (2009). We first find asymptotic distribution of $\hat{\beta}(\alpha, \gamma, \hat{\Lambda}_0)$. Using (A.1) and Lemma 1 of Qi et al. (2005) we have

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| \hat{\Lambda}_0(\beta, t) - \Lambda_0(\beta, t) \right\| \xrightarrow{a.s.} 0. \quad (\text{A.5})$$

Denoting

$$\mathbf{a}_n(t) = n^{-1} \sum_{j=1}^n \left[\frac{\{R_{j(t)} - \pi_{j(t)}\} / \pi_{j(t)} Y_{j(t)} \times}{\times \left\{ \hat{E}[\mathbf{Z}_{j(t)} \otimes^k \exp(\mathbf{Z}_{j(t)}^T \beta) | \mathbf{W}_{j(t)}] - E[\mathbf{Z}_{j(t)} \otimes^k \exp(\mathbf{Z}_{j(t)}^T \beta) | \mathbf{W}_{j(t)}] \right\}} \right],$$

we have

$$\hat{S}_A^{(k)}(\beta, t) = \mathbf{s}_A^{(k)}(\beta, t) - \mathbf{a}_n^{(k)}(t).$$

Then

$$\mathbf{a}_n^{(k)}(t) = n^{-1} \sum_{j=1}^n \frac{\{R_{j(t)} - \pi_{j(t)}\}}{\pi_{j(t)}} Y_{j(t)} \frac{\partial \mathbf{q}^{(k)}(t, \mathbf{W}_{j(t)})}{\partial \Lambda_0(T_{j(t)})} \{\hat{\Lambda}_0(T_{j(t)}) - \Lambda_0(T_{j(t)})\} + o_p(1),$$

where $\mathbf{q}^{(k)}(t, \mathbf{W}_{j(t)}) = \mathbf{e}[\mathbf{Z}_{j(t)} \otimes^k \exp(\mathbf{Z}_{j(t)}^T \beta) | \mathbf{W}_{j(t)}]$ and is given in the Section A.2. By conditions (C1), (C4), (C5), and (A.5), $\mathbf{a}_n^{(k)}(t)$ converges to 0 in probability uniformly over $t \in [0, \tau]$ and $\beta \in \mathcal{B}$. Applying (A.1), we have

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| \widehat{\mathbf{S}}_A^{(k)}(\beta, t) - \mathbf{s}^{(k)}(\beta, t) \right\| \xrightarrow{p} 0, \quad (\text{A.6})$$

for $k = 0, 1, 2$. Using similar steps in Xu et al. (2009) along with (A.5), (A.6), Lemma 1 in Qi et al. (2005), and condition (C8), we can show that

$$n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \widehat{\Lambda}_0) = n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) + o_p(1).$$

By the central limit theorem,

$$n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \widehat{\Lambda}_0) \xrightarrow{d} N(0, \mathbf{V}_{IPW} - \mathbf{V}_1). \quad (\text{A.7})$$

From (A.1), we can first establish

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| \frac{\partial \mathbf{s}_A^{(k)}(\beta, t)}{\partial \beta^T} - \mathbf{s}^{(k+1)}(\beta, t) \right\| \xrightarrow{p} 0,$$

for $k = 0, 1$ and after simple calculation, show

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \left\| -\frac{1}{n} \frac{\partial \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \widehat{\Lambda}_0)}{\partial \beta^T} - \mathbf{I}_\beta \right\| \xrightarrow{p} 0.$$

Therefore

$$n^{1/2}(\widehat{\beta}_A(\alpha, \gamma, \widehat{\Lambda}_0) - \beta) = \mathbf{I}_\beta^{-1} n^{-1/2} \mathbf{U}_A^\beta(\beta, \alpha, \gamma, \Lambda_0) + o_p(1),$$

and resulting asymptotic distribution of $n^{1/2} \widehat{\beta}_A(\alpha, \gamma, \widehat{\Lambda}_0)$ is the same as that of $n^{1/2} \widehat{\beta}_A(\alpha, \gamma, \Lambda_0)$.

Step 3. Asymptotic equivalence between $\widehat{\beta}_A(\widehat{\alpha}, \widehat{\gamma}, \Lambda_0)$ and $\widehat{\beta}_A(\alpha, \gamma, \Lambda_0)$

We establish asymptotic equivalence between $\widehat{\beta}_A(\widehat{\alpha}, \widehat{\gamma}, \Lambda_0)$ and $\widehat{\beta}_A(\alpha, \gamma, \Lambda_0)$. Let

$$\theta^T = (\beta^T, \gamma^T, \alpha^T),$$

$$\mathbf{U}_A(\theta) = \begin{pmatrix} \mathbf{U}_A^\beta(\theta) \\ \mathbf{U}_A^\gamma(\theta) \\ \mathbf{U}_A^\alpha(\theta) \end{pmatrix}, \quad \mathbf{I} = \mathbf{E} \left[-\frac{\partial \mathbf{U}_A(\theta)}{\partial \theta^T} \right] = \begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} & \mathbf{I}_{13} \\ \mathbf{I}_{12}^T & \mathbf{I}_{22} & \mathbf{I}_{23} \\ \mathbf{I}_{13}^T & \mathbf{I}_{23}^T & \mathbf{I}_{33} \end{pmatrix},$$

and $\mathbf{I}^{jk}$ denotes (j, k) -th submatrix of $\mathbf{I}^{-1}$. By the Taylor expansion, we have

$$n^{1/2}(\hat{\beta}_A(\hat{\alpha}, \hat{\gamma}, \Lambda_0) - \beta) = \mathbf{I}^{11} n^{-1/2} \mathbf{U}_A^\beta(\theta) + \mathbf{I}^{12} n^{-1/2} \mathbf{U}_A^\gamma(\theta) + \mathbf{I}^{13} n^{-1/2} \mathbf{U}^\alpha(\alpha) + o_p(1).$$

We can show that $\mathbf{I}^{12} = 0$ since $\sup_{t \in [0, \tau], \theta \in \Theta} \|n^{-1} \partial \hat{\mathbf{S}}_A^{(k)}(\theta, t) / \partial \gamma^T\| \xrightarrow{p} 0$, and consequently

$$\sup_{t \in [0, \tau], \theta \in \Theta} \left\| -n^{-1} \partial \mathbf{U}_A^\beta(\theta) / \partial \gamma^T \right\| \xrightarrow{p} 0,$$

by (A.1) and Lemma 1 in Qi et al. (2005). Similarly, $\mathbf{I}^{13} = 0$ because

$$\sup_{t \in [0, \tau], \theta \in \Theta} \left\| \partial \hat{\mathbf{S}}_A^{(k)}(\theta, t) / \partial \alpha^T \right\| \xrightarrow{p} 0, \text{ and thus}$$

$$\sup_{t \in [0, \tau], \theta \in \Theta} \left\| -n^{-1} \partial \mathbf{U}_A^\beta(\theta) / \partial \alpha^T \right\| \xrightarrow{p} 0.$$

From (A.3) and $\sup_{\theta \in \Theta} \left\| -n^{-1} \partial \mathbf{U}_A^\gamma(\theta) / \partial \beta^T \right\| \xrightarrow{p} 0$ using the same approach, we get $\mathbf{I}^{11}$ is upper left submatrix of $\mathbf{I}_\beta^{-1}$. Thus $\hat{\beta}_A(\hat{\alpha}, \hat{\gamma}, \Lambda_0)$ has the same asymptotic results as $\hat{\beta}_A(\alpha, \gamma, \Lambda_0)$.

Step 4: Repetitions on asymptotic equivalence:

Using the same approach for $\hat{\beta}_A(\alpha, \gamma, \Lambda_0)$ and $\hat{\beta}_A(\hat{\alpha}, \hat{\gamma}, \Lambda_0)$, we can establish the asymptotic behavior for $\hat{\beta}_A(\hat{\alpha}, \hat{\gamma}, \Lambda_0)$.

A.2 Computation of conditional expectations

A.2.1 Computation of $E[D(t)|\mathbf{W}]$ for the EM and DR

We show the conditional expectation $E[D(t)|\mathbf{W} = (T, \delta, \mathbf{Z})]$ in Section 3. The subscript i is omitted for brevity. First we let $E[D_i(t)|\mathbf{W}_i] = P(D_i(t) = 1|\mathbf{W}_i) = q_i(t)$. Then

$E[\exp(D_i(t)\beta_2)|\mathbf{W}_i]$, $E[D_i(t)\exp(D_i(t)\beta_2)|\mathbf{W}_i]$, $E[D_i^2(t)\exp(D_i(t)\beta_2)|\mathbf{W}_i]$ can be computed accordingly.

$$\begin{aligned}
E[D(t)|W] &= P[X \leq t | (T, \delta), \mathbf{Z}] = \frac{\int_0^t \tilde{f}(T, \delta | x, \mathbf{Z}) g(x | \mathbf{Z}; \gamma) dx}{\int_0^\infty \tilde{f}(T, \delta | x, \mathbf{Z}) g(x | \mathbf{Z}; \gamma) dx} = \\
&= \frac{\int_0^t g(x | \mathbf{Z}; \gamma) [\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + D(T) \beta_2)]^\delta \exp\left\{-\int_0^T \lambda_0(\mathbf{S}) \exp(\mathbf{Z}^T \beta_1 + D(\mathbf{S}) \beta_2) d\mathbf{S}\right\} dx}{\int_0^\infty g(x | \mathbf{Z}; \gamma) [\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + D(T) \beta_2)]^\delta \exp\left\{-\int_0^T \lambda_0(\mathbf{S}) \exp(\mathbf{Z}^T \beta_1 + D(\mathbf{S}) \beta_2) d\mathbf{S}\right\} dx} \\
&\equiv q(t).
\end{aligned}$$

(A.8)

If $T < t$, the subject is not in the risk set and this type of calculation is not necessary. If $t = T$, both the numerator and denominator of (A.8) have the factor $[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1)]^\delta$, which cancels out in computing $q(t)$. Thus, the numerator and denominator of (A.8) are, respectively,

$$q_n(t) = \exp(\delta \beta_2) \cdot \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \cdot \int_0^t g(x | \mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx$$

and

$$\begin{aligned}
q_d(t) &= \exp(\delta \beta_2) \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \int_0^T g(x | \mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx + \\
&\quad + \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1)\right\} \int_T^\infty g(x | \mathbf{Z}; \gamma) dx.
\end{aligned}$$

A.2.2 Computation of $E(D(t)|Q)$ for the EM and DR—We show $E[D(t)|Q]$ needed be evaluated for computing the EM and DR. Subscript i is omitted for brevity. Note $E[D(t)|Q] = 0$ if $t < u$, and $E[D(t)|Q] = 1$ if $v < t$. Thus, it suffices to consider the case $u \leq t \leq v$.

$$\begin{aligned}
E[D(t)|Q] &= P[X \leq t | T, \delta, \mathbf{Z}, I] = \frac{\int_u^t \tilde{f}(T, \delta | x, \mathbf{Z}, I) g(x | \mathbf{Z}) dx}{\int_u^v \tilde{f}(T, \delta | x, \mathbf{Z}, I) g(x | \mathbf{Z}) dx} = \\
&= \frac{\int_u^t g(x | \mathbf{Z}; \gamma) [\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + D(T) \beta_2)]^\delta \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s) \beta_2) ds\right\} dx}{\int_u^v g(x | \mathbf{Z}; \gamma) [\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + D(T) \beta_2)]^\delta \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s) \beta_2) ds\right\} dx} \\
&\equiv q(t).
\end{aligned}$$

First, if $T < u$,

$$q(t) = \frac{\int_u^t g(x|\mathbf{Z}; \gamma) \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1) ds\right\} dx}{\int_u^v g(x|\mathbf{Z}; \gamma) \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1) ds\right\} dx} = \frac{\int_u^t g(x|\mathbf{Z}; \gamma) dx}{\int_u^v g(x|\mathbf{Z}; \gamma) dx}.$$

Second, if $v < T$

$$q(t) = \frac{\int_u^t g(x|\mathbf{Z}; \gamma) \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s)\beta_2) ds\right\} dx}{\int_u^v g(x|\mathbf{Z}; \gamma) \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s)\beta_2) ds\right\} dx},$$

Third, if $u < t < v$ and $T < t$, this type of observation again is not appeared in the risk set at time t . Thus, the computation is not necessary.

Finally, if $u < T$, $t < v$ and $t \leq T$ $q(t) = q_{2n}/q_{2d}$ where

$$\begin{aligned} q_{2n} &= \int_u^t g(x|\mathbf{Z}; \gamma) \left[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2) \right]^\delta \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s)\beta_2) ds\right\} = \\ &= \left[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2) \right]^\delta \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \times \\ &\times \int_u^t g(x|\mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx \end{aligned}$$

and

$$\begin{aligned} q_{2d}(t) &= \int_u^v g(x|\mathbf{Z}; \gamma) \left[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + D(T)\beta_2) \right]^\delta \exp\left\{-\int_0^T \lambda_0(s) \exp(\mathbf{Z}^T \beta_1 + D(s)\beta_2) ds\right\} = \\ &= \left[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2) \right]^\delta \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \times \\ &\times \int_u^T g(x|\mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx + \\ &+ \left[\lambda_0(T) \exp(\mathbf{Z}^T \beta_1) \right]^\delta \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1)\right\} \int_T^v g(x|\mathbf{Z}; \gamma) dx. \end{aligned}$$

Therefore,

$$q(t) = \frac{\exp(\delta\beta_2) \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \cdot \int_u^t g(x|\mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx}{\exp(\delta\beta_2) \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1 + \beta_2)\right\} \int_u^T g(x|\mathbf{Z}; \gamma) \exp\left\{\Lambda_0(x) \exp(\mathbf{Z}^T \beta_1) (\exp(\beta_2) - 1)\right\} dx + \exp\left\{-\Lambda_0(T) \exp(\mathbf{Z}^T \beta_1)\right\} \int_T^v g(x|\mathbf{Z}; \gamma) dx}$$

REFERENCES

Atem FD, Qian J, Johnson KA, & Betensky RA (2017). Linear regression with a randomly censored covariate: Application to an Alzheimer's study. *Applied Statistics*, 66, part2, 313–328. [PubMed: 28239197]

- Austin PC, & Brunner LJ (2003). Type I error inflation in the presence of a ceiling effect. *The American Statistician*, 57, 97–104.
- Austin PC, & Hoch JJ (2004). Estimating linear regression models in the presence of a censored independent variable. *Statistics in Medicine*, 23, 411–429. [PubMed: 14748036]
- Banerjee C, Paik MC, Rundek CT, Mora-McLaughlin, Vieira JR, Sacco RL, & Elkind MSV (2012). Duration of diabetes and risk of ischemic stroke: The northern manhattan study. *Stroke*, 43, 1212–1217. [PubMed: 22382158]
- Betensky RA, Lindsey JC, Ryan LM, & Wand MP (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine*, 21, 263–275. [PubMed: 11782064]
- Chen HY (2002). Double-semiparametric method for missing covariates in cox regression models. *Journal of the American Statistical Association*, 97, 565–576.
- Chen HY, & Little RJA (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 94, 896–908.
- Chen Q, May RC, Ibrahim JG, Chu H, & Cole SR (2014). Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. *Statistics in Medicine*, 33, 4560–4576. [PubMed: 24947785]
- D'Angelo G, Weissfeld L, & on behalf of the GenIMS Investigators. (2008). An index approach for the cox model with left censored covariates. *Statistics in Medicine*, 27, 4502–4514. [PubMed: 18407573]
- Dupuy J, Grama I, & Mesbah M (2006). Asymptotic theory for the cox model with missing time-dependent covariate. *The Annals of Statistics*, 34, 903–924.
- Foutz R (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72, 147–148.
- Gajewski BJ, Nicholson N, & Widen JE (2009). Predicting hearing threshold in nonresponsive subjects using a log-normal Bayesian linear model in the presence of left-censored covariates. *Statistics in Biopharmaceutical Research*, 1, 137–148.
- Goetghebuer E, & Ryan L (2000). Semiparametric regression analysis of interval-censored data. *Biometrics*, 56, 1139–1144. [PubMed: 11129472]
- Goggins WB, Finkelstein DM, & Zaslavsky AM (1999). Applying the cox proportional hazards model when the change time of a binary time-varying covariate is interval censored. *Biometrics*, 55, 445–451. [PubMed: 11318198]
- Gomez G, Espinal A, & Lagakos SW (2003). Inference for a linear regression model with an interval-censored covariate. *Statistics in Medicine*, 22, 409–425. [PubMed: 12529872]
- Heitjan DF, & Rubin DB (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, 85, 304–314.
- Heller G (2011). Proportional hazards regression with interval censored data using an inverse probability weight. *Lifetime Data Analysis*, 17, 373–385. [PubMed: 21191653]
- Herring AH, & Ibrahim JG (2001). Likelihood-based methods for missing covariates in the cox proportional hazards model. *Journal of the American Statistical Association*, 96, 292–302.
- Langohr K, Gomez G, & Muga R (2004). A parametric survival model with an interval-censored covariate. *Statistics in Medicine*, 23, 3159–3175. [PubMed: 15449329]
- Langohr K, & Melis GG (2014). Estimation and residual analysis with R for a linear regression model with an interval-censored covariate. *Biometrical Journal*, 56(5), 867–885. [PubMed: 25103399]
- Lipsitz SR, & Ibrahim JG (1998). Estimating equations with incomplete categorical covariates in the cox model. *Biometrics*, 54, 1002–1013. [PubMed: 9750248]
- Little RJA, & Rubin DB (1987). *Statistical analysis with missing data* New York, NY: John Wiley and Sons.
- Louis TA (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society Series B*, 44, 226–233.
- Paik MC, & Tsai W (1997). On using the cox proportional hazards models with missing covariates. *Biometrika*, 84, 579–593.

- Pugh M, Robins J, Lipsitz S, & Harrington D (1993). Inference in the cox proportional hazards model with missing covariate data. Technical report 758Z Division of Biostatistics, Dana-Farber Cancer Institute, Boston, MA.
- Qi L, Wang CY, & Prentice RL (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 100, 1250–1263.
- Rigobon R, & Stoker TM (2009). Bias from censored regressors. *Journal of Business and Economic Statistics*, 27, 340–353.
- Robins JM, Rotnitzky A, & Zhao LP (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846–866.
- Sattar A, Sinha SK, & Morris NJ (2012). A parametric survival model when a covariate is subject to left-censoring. *Journal of Biometrics and Biostatistics*, Supple 3(2). 10.4172/2155-6180.S3-002.
- Tsimikas JV, Bantis LE, & Georgiou SD (2012). Inference in generalized linear regression models with a censored covariate. *Computational Statistics and Data Analysis*, 56, 1854–1868.
- Wang CY, & Chen HY (2001). Augmented inverse probability weighted estimator for cox missing covariate regression. *Biometrics*, 57, 414–419. [PubMed: 11414564]
- Wang HJ, & Feng X (2012). Multiple imputation for m-regression with consored covariates. *Journal of the American Statistical Association*, 107, 194–204.
- Xu Q, Paik MC, Luo Z, & Tsai WY (2009). Reweighting estimators for cox regression with missing covariates. *Journal of the American Statistical Association*, 104, 1155–1167.

FIGURE 1.
The number of missing $D(t)$ at each failure time in the risk set

FIGURE 2.

Kaplan–Meier curves for those who had diabetes at baseline (red) versus those without diabetes at baseline but who developed diabetes during follow-up (blue) versus those who never developed diabetes (green)

TABLE 1

Correctly specified diabetes model: simulation comparison of various estimators of the Cox model parameter $\beta = (-\log(2), \log(2))$ based on 500 replications with a sample size of 300, 22% censoring proportion

Missing Prop.	Method	$\hat{\beta}_1$	RB	MSE	CP	$\hat{\beta}_2$	RB	MSE	CP
31%	AD	-0.688	0.007	0.008	0.962	0.676	0.025	0.130	0.966
	LOCF	-0.646	0.067	0.008	0.898	0.285	0.588	0.222	0.600
	DR	-0.690	0.005	0.007	0.962	0.652	0.060	0.127	0.950
	EM	-0.692	0.002	0.006	0.972	0.688	0.007	0.097	0.976
	FULL	-0.692	0.002	0.006	0.964	0.688	0.008	0.056	0.968
44%	AD	-0.704	0.016	0.010	0.944	0.736	0.062	0.230	0.944
	LOCF	-0.642	0.073	0.008	0.900	0.193	0.722	0.307	0.454
	DR	-0.705	0.018	0.009	0.940	0.715	0.031	0.225	0.928
	EM	-0.704	0.016	0.007	0.950	0.735	0.060	0.147	0.942
	FULL	-0.701	0.011	0.006	0.962	0.705	0.017	0.068	0.936

LOCF: last observation carried forward; RB: relative bias $\left(= |E(\hat{\beta}) - \beta|/|\beta| \right)$; MSE : mean squared error; CP: coverage probability.

TABLE 2

Misspecified diabetes models: simulation comparison of various estimators of the Cox model parameter $\beta = (-\log(2), \log(2))$ based on 500 replications with a sample size of 300, 22 % censoring proportion

Missing Prop.	Method	$\hat{\beta}_1$	RB	MSE	CP	$\hat{\beta}_2$	RB	MSE	CP
31%	AD	-0.701	0.012	0.008	0.958	0.686	0.010	0.083	0.958
	LOCF	-0.659	0.049	0.007	0.938	0.420	0.394	0.128	0.938
	DR	-0.694	0.001	0.006	0.964	0.663	0.043	0.069	0.964
	EM	-0.696	0.005	0.006	0.956	0.654	0.057	0.056	0.956
43%	AD	-0.698	0.006	0.010	0.950	0.694	0.001	0.101	0.956
	LOCF	-0.648	0.065	0.008	0.894	0.360	0.481	0.164	0.678
	DR	-0.691	0.004	0.007	0.962	0.659	0.050	0.084	0.980
	EM	-0.696	0.004	0.007	0.946	0.640	0.077	0.067	0.946

LOCF: last observation carried forward; RB: relative bias $(= |E(\hat{\beta}) - \beta|/|\beta|)$; MSE : mean squared error; CP: coverage probability.

TABLE 3

NOMAS data example: comparison of various estimators with all subjects ($n = 3021$). S.E: standard error

Method	AD	S.E.	LOCF	S.E	DR	S.E	EM	S.E
BLACK	-1.368	1.431	-1.106	1.413	-1.151	1.242	-1.108	1.415
HISP	-3.248	1.301	-3.175	1.275	-3.188	1.119	-3.182	1.275
MALE	0.462	0.153	0.423	0.151	0.421	0.128	0.425	0.151
AGE	0.052	0.008	0.054	0.008	0.053	0.006	0.054	0.008
Etmod	-0.448	0.163	-0.419	0.160	-0.422	0.132	-0.414	0.160
NCAD	0.442	0.144	0.411	0.143	0.412	0.120	0.407	0.143
SYSTOLIC	-0.006	0.008	-0.005	0.008	-0.005	0.007	-0.005	0.008
SMOKER1	0.209	0.156	0.187	0.154	0.182	0.131	0.183	0.154
SMOKER2	0.548	0.194	0.500	0.192	0.499	0.154	0.497	0.192
LHDL	0.003	0.005	0.002	0.005	0.002	0.004	0.003	0.005
LLDL	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
SYSTOLIC*BLACK	0.011	0.010	0.009	0.010	0.010	0.009	0.009	0.010
SYSTOLIC*HISP	0.023	0.009	0.023	0.009	0.023	0.008	0.023	0.009
$D(t)$	0.884	0.139	0.849	0.137	0.855	0.113	0.905	0.138

TABLE 4

NOMAS data example: comparison of various estimators with those who do not have diabetes at baseline ($n = 2366$). S.E: standard error

Method	AD	S.E.	LOCF	S.E	DR	S.E	EM	S.E
BLACK	-0.171	1.672	0.154	1.648	0.135	1.242	0.135	1.648
HISP	-3.339	1.514	-3.280	1.468	-3.279	1.102	-3.260	1.462
MALE	0.523	0.201	0.451	0.197	0.450	0.140	0.462	0.197
AGE	0.058	0.010	0.061	0.010	0.060	0.006	0.061	0.010
Etmod	-0.512	0.206	-0.465	0.200	-0.467	0.155	-0.461	0.200
NCAD	0.425	0.191	0.372	0.187	0.373	0.140	0.368	0.187
SYSTOLIC	-0.004	0.009	-0.003	0.009	-0.003	0.007	-0.003	0.009
SMOKER1	0.371	0.200	0.329	0.194	0.325	0.138	0.325	0.194
SMOKER2	0.585	0.258	0.498	0.254	0.499	0.180	0.496	0.254
LHDL	0.010	0.006	0.007	0.006	0.007	0.004	0.008	0.006
LLDL	0.000	0.003	-0.001	0.002	-0.001	0.002	-0.001	0.002
SYSTOLIC*BLACK	0.003	0.012	0.001	0.011	0.001	0.009	0.001	0.011
SYSTOLIC*HISP	0.023	0.010	0.023	0.010	0.023	0.008	0.022	0.010
$D(t)$	0.366	0.344	0.300	0.342	0.323	0.235	0.589	0.309