Lifetime Data Analysis (2022) 28:335–355
https://doi.org/10.1007/s10985-022-09550-y

A new approach to estimation of the proportional hazards model based on interval-censored data with missing covariates

Ruiwen Zhou[1] · Huiqiong Li[2] · Jianguo Sun[1] · Niansheng Tang[2]

Received: 24 March 2021 / Accepted: 3 February 2022 / Published online: 29 March 2022

Abstract

This paper discusses the fitting of the proportional hazards model to interval-censored failure time data with missing covariates. Many authors have discussed the problem when complete covariate information is available or the missing is completely at random. In contrast to this, we will focus on the situation where the missing is at random. For the problem, a sieve maximum likelihood estimation approach is proposed with the use of I-spline functions to approximate the unknown cumulative baseline hazard function in the model. For the implementation of the proposed method, we develop an EM algorithm based on a two-stage data augmentation. Furthermore, we show that the proposed estimators of regression parameters are consistent and asymptotically normal. The proposed approach is then applied to a set of the data concerning Alzheimer Disease that motivated this study.

Keywords Case II interval-censored data · EM algorithm · Missing at random · Sieve approach

1 Introduction

It is well-known that the proportional hazards model is one of the most commonly used models for regression analysis of failure time data, and a great deal of literature has been established for fitting it to right-censored or interval-censored data. By interval-censored data, we mean that the failure time of interest is observed only to belong to an interval instead of being known exactly, and it is apparent that they include right-censored data as a special case (Sun 2006). Among others, the fields that often generate

✉ Huiqiong Li
lihuiqiong@ynu.edu.cn

[1] Department of Statistics, University of Missouri, Columbia, MO 65211, USA

[2] Department of Statistics, Yunnan University, Kunming 650091, China

interval-censored data include demographical, epidemiological, financial, medical and sociological studies. In the following, we will discuss the fitting of the proportional hazards model to interval-censored failure time data when some covariates may have missing observations.

As discussed by many authors, missing data can arise due to many circumstances and in general, their analysis highly depends on the censoring mechanism (Little and Rubin 2002). For the situation, a naive approach is the so-called complete-case (CC) method, which bases the analysis only on the complete part of the data or throw away the subjects with missing information. It is apparent that the CC method not only may be inefficient but also could yield biased estimation when the missing data mechanism depends on the observed data (Lipsitz et al. 1994; Little and Rubin 2002; Qi et al. 2005). Instead of the CC method, some alternatives could be the multiple imputation procedure and the estimating equation approach. As pointed out by many authors, when the missing is missing at random (MAR), the focus of this paper, the maximum likelihood approach may be preferred or should be used.

Several maximum likelihood methods have been proposed for regression analysis of right-censored failure time data with missing covariates under the proportional hazards model when the missing is MAR (Chen et al. 2002; Chen and Little 1999; Zhou and Pepe 1995). However, it does not seem to exist an established approach for interval-censored data when some covariates may be missing at random except Wen and Lin (2011), who proposed a semiparametric maximum likelihood estimation procedure under the proportional hazards model. However, they only considered current status data or case I interval-censored data, a special case of the general interval-censored data discussed here. In the following, we will consider the estimation of the proportional hazards model when one faces case II interval-censored data with missing covariates and propose a sieve maximum likelihood estimation approach. The method can be easily implemented and makes use of I-spline functions to approximate the underlying cumulative hazard function in the model.

The remainder of this paper is organized as follows. We will begin in Sect. 2 with introducing the model and assumptions that will be used throughout the paper and then present the resulting likelihood function. The proposed sieve maximum likelihood estimation approach will be derived in Sect. 3, and in particular, for the determination of the proposed estimators, an EM algorithm is developed. Section 4 establishes the asymptotic properties of the proposed estimators of regression parameters. Some results obtained from a simulation study are presented in Sect. 5 and suggest that the proposed approach works well in practical situations. Section 6 provides an application and some discussion and concluding remarks are given in Sect. 7.

2 Models, assumptions and likelihood functions

Consider a failure time study that involves n independent subjects and let T_i and $\mathbf{X_i}$ denote the failure time of interest and a p-dimensional vector of covariates associated with subject i. In the following, suppose that for each subject, there exist two monitoring variables or observation times U_i and V_i with $U_i < V_i$, and instead of observing T_i, one observes only U_i and V_i and the indicator variables $\delta_{1i} = I(T_i < U_i)$,

$\delta_{2i} = I(U_i \leq T_i < V_i)$ and $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$. That is, we only know if the failure for subject i has occurred before U_i, during the examination interval $[U_i, V_i)$ or after V and observe case II interval-censored data (Sun 2006).

To describe the covariate effect on T_i, we will assume that given the covariates $\mathbf{X_i}$, the cumulative hazard function of T_i has the form

$$\Lambda_i(t|\mathbf{X_i}) = \Lambda_0(t)\exp\{\beta'\mathbf{X_i}\}, \tag{1}$$

where $\Lambda_0(t)$ denotes an unspecified baseline cumulative hazard function and $\boldsymbol{\beta}$ a p-dimensional vector of regression parameters. That is, T_i follows the proportional hazards model. In the following, we will assume that given the covariate $\mathbf{X_i}$, the failure time T_i is independent of the observation times U_i and V_i or we have the independent interval censoring.

Under the assumptions above, if there was no missing covariate, the likelihood function would have the form

$$L_c(\beta, \gamma, \Lambda_0) = \prod_{i=1}^{n} f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}; \beta, \Lambda(t))\, f(\mathbf{X_i}; \gamma).$$

In the above, $f(\mathbf{X_i}; \gamma)$ denotes the density function of the covariate with the unknown parameter γ and

$$\begin{aligned} f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}) &\propto [1 - \exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{1i}} \\ &\times [\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\} - \exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{2i}} \\ &\times [\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{3i}}, \quad i = 1, \ldots, n. \end{aligned} \tag{2}$$

It follows that we would have the log likelihood function

$$\begin{aligned} l_n(\beta, \gamma, \Lambda_0) &= \log[L_c(\theta, \beta, \gamma, \Lambda(t))] \\ &= \sum_{i=1}^{n} \log[f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}; \beta, \Lambda(t))] + \sum_{i=1}^{n} \log[f(\mathbf{X_i}; \gamma)] \\ &= l_1(\beta, \Lambda_0) + l_2(\gamma). \end{aligned} \tag{3}$$

It is easy to see that one can maximize $l_1(\beta, \Lambda_0)$ and $l_2(\gamma)$ separately if the goal is to estimate β, γ and Λ_0, or can ignore $l_2(\gamma)$ since γ is usually not of interest. As will be seen below, we have to estimate β, γ and Λ_0 together when there are missing covariates. Now suppose that some covariates may be missing and the covariate can be written as $\mathbf{X_i}' = (\mathbf{X_i^{obs}}', \mathbf{X_i^{mis}}')$, where $\mathbf{X_i^{obs}}$ denotes the components of the covariates that are known or can be observed and $\mathbf{X_i^{mis}}$ the components of the covariates that are missing. Also suppose that we can write the density function of the covariates as

$$f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma) \propto f(\mathbf{X_i}^{obs}) f(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}}; \gamma).$$

Let $R_i = (R_{i1}, ..., R_{ip})'$ denote the missing indicator with $R_{ij} = 1$ if the jth component of the covariate associated with subject i is observed and 0 otherwise. In the following, we will assume that the covariate is missing at random, meaning that

$$f(R_i|U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{mis}}, \mathbf{X_i^{obs}}) = f(R_i|U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{obs}})$$

for the conditional density function of R_i. Then the observed likelihood function has the form

$$L_o(\theta) = \prod_{i=1}^{n} \int f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i^{mis}}, \mathbf{X_i^{obs}}; \beta, \Lambda(t)) f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma)\mathbf{dX_i^{mis}},$$

where $\theta = (\beta, \gamma, \Lambda_0)$. In the next section, we will discuss estimation of θ by maximizing $L_o(\theta)$.

3 Sieve maximum likelihood estimation

In this section, we will discuss estimation of θ by maximizing $L_o(\theta)$ with the focus on making inference about β. For this, it is apparent that it would be difficult directly to maximize it and thus we will develop an EM algorithm. Before presenting the algorithm, we will first discuss the use of the sieve approach and then the data augmentation.

It is well-known that the sieve approach can be used to approximate an unknown function in order to reduce the number of unknown parameters and the computational burden (Ma et al. 2015; Zhao et al. 2015; Li et al. 2017). More specifically, for the estimation here, we suggest to approximate the baseline cumulative hazard function $\Lambda_0(t)$ by monotone spline functions such as

$$\Lambda_n(t) = \sum_{l=1}^{s+k_n} \alpha_l I_l(t)$$

(Ramsay 1988). In the above, $\{I_l(t), l = 1, \ldots, s + k_n\}$ are integrated spline basis functions with the order s and the number of knots k_n, and the α_l's are nonnegative coefficients that ensure monotonicity of $\Lambda_n(t)$. The degree s determines the smoothness of the true baseline cumulative hazard function and is often taken to be 1, 2, or 3, which corresponds to linear, quadratic, or cubic basis functions, respectively. In practice, for the choice of s and k_n, one commonly used method is to try different values of them and compare the obtained results. As an alternative, one could also use the AIC to choose the values of s and k_n that give the smallest AIC, and more discussion on this is given below.

Now we discuss the data augmentation and for this, we will first assume that all covariates have been observed. Then the log likelihood function $l_1(\beta, \Lambda(t))$ would have the form

$$\begin{aligned} l_1(\beta, \Lambda_0) = \sum_{i=1}^{n} \log \Big\{ & \big[1 - \exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}\big]^{\delta_{1i}} \\ & \times \big[\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\} - \exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}\big]^{\delta_{2i}} \\ & \big[\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\}\big]^{\delta_{3i}} \Big\}. \end{aligned} \tag{4}$$

By replacing Λ_0 with Λ_n, we have that

$$\begin{aligned} l_1^*(\beta, \alpha_l) = \sum_{i=1}^{n} \log \Bigg\{ & [1 - \exp \left\{ - \left(\sum_{l=1}^{s+k_n} \alpha_l I_l(V_i) \right) \exp(\beta'\mathbf{X_i}) \right\}]^{\delta_{1i}} \\ & \times \left[\exp \left\{ - \left(\sum_{l=1}^{s+k_n} \alpha_l I_l(U_i) \right) \exp(\beta'\mathbf{X_i}) \right\} \right. \\ & \left. - \exp \left\{ - \left(\sum_{l=1}^{s+k_n} \alpha_l I_l(V_i) \right) \exp(\beta'\mathbf{X_i}) \right\} \right]^{\delta_{2i}} \\ & \times \left[\exp \left\{ - \left(\sum_{l=1}^{s+k_n} \alpha_l I_l(U_i) \right) \exp(\beta'\mathbf{X_i}) \right\} \right]^{\delta_{3i}} \Bigg\}. \end{aligned} \tag{5}$$

Note that as pointed out by McMahan et al. (2013), the direct maximization of the function above with the traditional algorithm would suffer numerical instability. Also one may often get local maximizers and have other issues like convergence. In the following, we will augment the observed data.

Let $N_i(t)$ denote the latent Poisson process with the mean function $\Lambda_n(t)\exp\{\beta'\mathbf{X_i}\}$, $i = 1, \ldots, n$, and define $Z_i = N_i(t_{1i})$ and $W_i = N_i(t_{2i}) - N_i(t_{1i})$ for $\delta_{1i} = 0$, where $t_{1i} = V_i I(\delta_{1i} = 1) + U_i I(\delta_{1i} = 0)$, and $t_{2i} = V_i I(\delta_{2i} = 1) + U_i I(\delta_{3i} = 1)$. Then Z_i and W_i are Poisson random variables with means $\Lambda_n(t_{1i})\exp\{\beta'\mathbf{X_i}\}$ and $\{\Lambda_n(t_{2i}) - \Lambda_n(t_{1i})\}\exp\{\beta'\mathbf{X_i}\}$, respectively, and they are independent given $\delta_{1i} = 0$. Furthermore, note that if T_i is left-censored or interval-censored, we have that

$$P(T_i \leq t_{1i}) = P(N_i(t_{1i}) > 0) = P(Z_i > 0) = 1 - \exp\{-\Lambda_n(V_i)\exp(\beta'\mathbf{X_i})\},$$

or

$$\begin{aligned} P(t_{1i} < T_i \leq t_{2i}) &= P\{N_i(t_{1i}) = 0, N_i(t_{2i}) > 0\} = P(Z_i = 0, W_i > 0) \\ &= \exp\{-\Lambda_n(U_i)\exp(\beta'\mathbf{X_i})\} - \exp\{-\Lambda_n(V_i)\exp(\beta'\mathbf{X_i})\}, \end{aligned}$$

and for right-censored T_i, we have that

$$P(T_i \geq t_{2i}) = P\{N_i(t_{2i}) = 0\} = P(Z_i = 0, W_i = 0) = \exp\{-\Lambda_n(U_i)\exp(\beta'\mathbf{X_i})\}.$$

Thus if the Z_i's and W_i's were observed, the log likelihood function corresponding to $l_1^*(\beta, \alpha_l)$ would have the form

$$\begin{aligned} l_1^{**}(\beta, \alpha_l) = \sum_{i=1}^{n} & \log\{P_{Z_i}(Z_i) P_{W_i}(W_i)^{\delta_{2i}+\delta_{3i}} \{\delta_{1i} I(Z_i > 0) \\ & + \delta_{2i} I(Z_i = 0, W_i > 0) + \delta_{3i} I(Z_i = 0, W_i = 0)\}\}. \end{aligned}$$

In the above, $P_A(.)$ denotes the probability function associated with the random variable A.

In addition, note that one can decompose or write Z_i and W_i as

$$Z_i = \sum_{l=1}^{k} Z_{il}, \; W_i = \sum_{l=1}^{k} W_{il},$$

the summation of k independent Poisson random variables Z_{il}'s and W_{il}'s with means $\alpha_l I_l(t_{1i}) \exp(\beta' \mathbf{X_i})$ and $\alpha_l \{I_l(t_{2i}) - I_l(t_{1i})\} \exp(\beta' \mathbf{X_i})$, respectively. Then by treating $\{ (Z_i, W_i, Z_{il}, W_{il}, \mathbf{X_i^{mis}}) \}$ to be known, we would have the complete log likelihood function

$$\begin{aligned} l_1^{***}(\beta, \alpha_l) = \sum_{i=1}^{n} \sum_{l=1}^{k} & \log \left\{ P_{Z_{il}}(Z_{il}) P_{W_{il}}(W_{il})^{\delta_{2i}+\delta_{3i}} \right. \\ & \times \{\delta_{1i} I(Z_i > 0) + \delta_{2i} I(Z_i = 0, W_i > 0) + \delta_{3i} I(Z_i = 0, W_i = 0)\}\} \end{aligned}$$

corresponding to $l_1^*(\beta, \alpha_l)$.

Now we are ready to discuss the two steps of the proposed EM algorithm. Let $\mathbf{O_i} = (U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{obs}}, \mathbf{R_i})$ denote the observed data on subject i and $\theta^{\mathbf{(d)}} = (\beta^{(d)'}, \alpha_l^{(d)'}, \gamma^{(d)'})'$ the estimator of the parameters given after the d iterations. In the E-step of the $(d+1)$th iteration, we need to determine the expectation $Q(\theta | \theta^{\mathbf{(d)}}) = E[l_1^{***}(\beta, \alpha_l) + l_2(\gamma) | \mathbf{O_i}, \theta^{\mathbf{(d)}}]$ or

$$\begin{aligned} Q(\theta|\theta^{\mathbf{(d)}}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{l=1}}^{\mathbf{k}} & [\{\mathbf{E}(\mathbf{Z_{il}}|\mathbf{O_i}, \theta^{\mathbf{(d)}}) + (\delta_{\mathbf{2i}} + \delta_{\mathbf{3i}})\mathbf{E}(\mathbf{W_{il}}|\mathbf{O_i}, \theta^{\mathbf{(d)}})\} \times \{\mathbf{log}(\alpha_l) + \beta_1' \mathbf{X_i^{obs}}\} \\ & + \{\beta_2' E(Z_{il} \mathbf{X_i^{mis}} | \mathbf{O_i}, \theta^{\mathbf{(d)}}) + \beta_2' (\delta_{i2} + \delta_{i3}) E(W_{il} \mathbf{X_i^{mis}} | \mathbf{O_i}, \theta^{\mathbf{(d)}})\} \\ & - \alpha_l \exp(\beta_1' \mathbf{X_i^{obs}}) E(\exp(\beta_2' \mathbf{X_i}^{mis} | \mathbf{O_i}, \theta^{\mathbf{(d)}})) \{(\delta_{1i} + \delta_{2i}) I_l(V_i) + \delta_{3i} I_l(U_i)\}] \\ & + \sum_{i=1}^{n} \int \log\{f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma)\} \mathbf{f}(\mathbf{X_i^{mis}} | \mathbf{O_i}, \theta^{\mathbf{(d)}}) \mathbf{dX_i^{mis}} + \mathbf{l}(\theta^{\mathbf{(d)}}). \end{aligned}$$

In the above, β_1 and β_2 denote the components of β corresponding to the observed and missing covariates, respectively, and $l(\theta^{\mathbf{(d)}})$ is a function of $\theta^{\mathbf{(d)}}$ free of θ.

For the determination of the expectations above, we need to calculate

$$E(Z_{il}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\alpha_l^{(d)} I_l(V_i) E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})})}{\Lambda^{(d)}(V_i)},$$

and

$$E(W_{il}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\alpha_l^{(d)}\{I_l(V_i) - I_l(U_i)\} \times E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})})}{\Lambda^{(d)}(V_i) - \Lambda^{(d)}(U_i)},$$

where $\Lambda^{(d)}(\cdot) = \sum_{l=1}^{k} \alpha_l^{(d)} I_l(\cdot)$. Note that if there are no missing covariates, by following Wang et al. (2016), we have that

$$E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}\mathbf{X_i}^{\mathbf{mis}})\delta_{1i}}{1 - \exp\{-\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}\mathbf{X_i}^{\mathbf{mis}})\}},$$

and

$$E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}\mathbf{X_i}^{\mathbf{mis}})\delta_{2i}}{1 - \exp[-\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}\mathbf{X_i}^{\mathbf{mis}})]}.$$

When there exist missing categorical covariates, by following Herring and Ibrahim (2001), we have that

$$E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{\mathbf{mis}}(j)} \frac{\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))\delta_{1i}p_{ij}}{1 - \exp\{-\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))\}},$$

and

$$E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{\mathbf{mis}}(j)} \frac{\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))\delta_{2i}p_{ij}}{1 - \exp[-\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))]}.$$

Here $\mathbf{X_i^{mis}(j)}$ denotes the jth possible missing data pattern for subject i and p_{ij} the conditional probability of a given missing data pattern, which can be estimated in the dth iteration of the EM algorithm by

$$\begin{aligned} p_{ij} &= P(\mathbf{X_i}^{\mathbf{mis}} = x_i^{\mathbf{mis}}(j)|\mathbf{O_i}, \text{`}^{(\mathbf{d})}) \\ &= \frac{f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}^{\mathbf{obs}}, x_i^{\mathbf{mis}}(j)) f(\mathbf{X_i}^{\mathbf{obs}}, x_i^{\mathbf{mis}}(j); \gamma^{(d)})}{\sum_{x_i^{\mathbf{mis}}(j)} f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}^{\mathbf{obs}}, x_i^{\mathbf{mis}}(j)) f(\mathbf{X_i}^{\mathbf{obs}}, x_i^{\mathbf{mis}}(j); \gamma^{(d)})}. \end{aligned}$$

For the situation where missing covariates are continuous, the calculation, which will be described at “Appendix I”, will involve integrations and do not have the closed forms.

In the M-step of the $(d+1)$th iteration, we need to maximize $Q(\theta, \theta^{(d)})$. For this, one can solve the following score equations

$$\begin{aligned}
\frac{\partial Q}{\partial \beta_1} &= \sum_{i=1}^{n}[\{E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) + \delta_{2i}E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})\} - \{(\delta_{2i}+\delta_{1i})\Lambda(V_i) + \delta_{3i}\Lambda(U_i)\} \\
&\quad \exp(\beta_1'\mathbf{X_i}^{\mathbf{obs}})E(\exp(\beta_2'\mathbf{X_i}^{\mathbf{mis}}))|\mathbf{O_i}, \theta^{(\mathbf{d})})]\mathbf{X_i}^{\mathbf{obs}} = 0, \qquad (6) \\
\frac{\partial Q}{\partial \beta_2} &= \sum_{i=1}^{n}\Bigg[\{E(Z_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) + \delta_{2i}E(W_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})\} \\
&\quad - \{(\delta_{2i}+\delta_{1i})\Lambda(V_i) + \delta_{3i}\Lambda(U_i)\} \\
&\quad \times \exp(\beta_1'\mathbf{X_i}^{\mathbf{obs}})\frac{\partial E(\exp(\beta_2'\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}))}{\partial \beta_2}\Bigg] = 0, \qquad (7) \\
\frac{\partial Q}{\partial \alpha_l} &= \sum_{i=1}^{n}[\alpha_l^{-1}\{E(Z_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})}) + \delta_{2i}E(W_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})\} \\
&\quad - \{(\delta_{2i}+\delta_{1i})I_l(V_i) + \delta_{3i}I_l(U_i)\} \\
&\quad \times \exp(\beta_1'\mathbf{X_i}^{\mathbf{obs}})E(\exp(\beta_2'\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}))] = 0, \qquad (8) \\
\frac{\partial Q}{\partial \gamma} &= \sum_{i=1}^{n}\frac{\partial[\int \log\{f(\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}}; \gamma)\}f(\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})d\mathbf{X_i}^{\mathbf{mis}}]}{\partial \gamma^{(d)}} = 0. \qquad (9)
\end{aligned}$$

In the above,

$$\begin{aligned}
&\frac{\partial E(\exp(\beta_2'\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}))}{\partial \beta_2} = \sum_{x_i^{\mathbf{mis}}(j)} \exp\left(\beta_2' x_i^{\mathbf{mis}}(j)\right) x_i^{\mathbf{mis}}(j)p_{ij}, \\
&E(\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \sum_{x_i^{\mathbf{mis}}(j)} x_i^{\mathbf{mis}}(j)p_{ij}, \\
&E(Z_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \sum_{x_i^{\mathbf{mis}}(j)} \frac{\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))x_i^{\mathbf{mis}}(j)\delta_{1i}p_{ij}}{1 - \exp\{-\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))\}},
\end{aligned}$$

and

$$\begin{aligned}
&E(W_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) \\
&\quad = \sum_{x_i^{\mathbf{mis}}(j)} \frac{\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))x_i^{\mathbf{mis}}(j)\delta_{2i}p_{ij}}{1 - \exp[-\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i}^{\mathbf{obs}} + \beta_2^{(d)'}x_i^{\mathbf{mis}}(j))]}.
\end{aligned}$$

The proposed EM algorithm can be summarized as follows.

Step 1. Select the initial estimates $\beta_1^{(0)}$, $\beta_2^{(0)}$, $\alpha_l^{(0)}$ and $\gamma^{(0)}$.

Step 2. At the $(d+1)$th iteration, compute the conditional expectations $E(Z_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(W_i\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(Z_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(W_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$, and $E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$.

Step 3. Obtain $\hat{\beta}_{\mathbf{1}}^{(d+1)}$ and $\hat{\beta}_{\mathbf{2}}^{(d+1)}$ by solving the Eqs. (6) and (7) with

$$\alpha_l^{*(d)}(\beta) = \frac{\sum_{i=1}^n \{E(Z_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})}) + \delta_{2i}E(W_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})\}}{\sum_{i=1}^n [\{(\delta_{2i}+\delta_{1i})I_l(V_i) + \delta_{3i}I_l(U_i)\}\exp(\beta_1'\mathbf{X_i}^{\mathbf{obs}})E(\exp(\beta_2'\mathbf{X_i}^{\mathbf{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}))]} \tag{10}$$

Step 4. Obtain $\hat{\alpha}_l^{(d+1)}(\beta)$ by solving the Eq. (8) and applying the Quasi-Newton method or the Eq. (10) given $\hat{\beta}_{\mathbf{1}}^{(d+1)}$, $\hat{\beta}_{\mathbf{2}}^{(d+1)}$.

Step 5. Obtain $\hat{\gamma}^{(d+1)}$ by solving the Eq. (9).

Step 6. Repeat Steps 2–5 until a pre-specified converge criterion is satisfied.

Note that for the application of the method proposed above, one needs to specify $f(\mathbf{X};\gamma)$, the density function of the covariates. For this, based on Herring and Ibrahim (2001), the standard option is joint normal distribution, Bernoulli distribution or the logistic regression model if missing covariates are continuous, binary, or categorical, respectively. More discussion on this is given below.

4 Asymptotic properties

Let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n)$ denote the sieve maximum likelihood estimator of θ defined in the previous section and $\hat{\theta}_n^* = (\hat{\beta}_n, \hat{\Lambda}_n)$. Now we will establish the asymptotic properties of $\hat{\theta}_n^*$. For this, let $\theta_0^* = (\beta_0, \Lambda_0)$ denote the true value of θ^* and define the distance between $\theta^1 = ({\beta_1}^1, {\beta_2}^1, \Lambda^1)$ and $\theta^2 = ({\beta_1}^2, {\beta_2}^2, \Lambda^2)$ as

$$d(\theta^1, \theta^2) = \{\,||{\beta_1}^1 - {\beta_1}^2||^2 + ||{\beta_2}^1 - {\beta_2}^2||^2 + ||\Lambda^1 - \Lambda^2||_2^2\,\}^{1/2}.$$

In the above, $||v||$ denotes the Euclidean norm of a vector v and $||\Lambda^1 - \Lambda^2||_2^2 = \int[\{\Lambda^1(u) - \Lambda^2(u)\}^2 + \{\Lambda^1(v) - \Lambda^2(v)\}^2]df(u, v)$, where $f(u, v)$ represents the joint density function of U and V. Then we have the following consistency and asymptotic normality results.

Theorem 1 *Assume that the regularity conditions given in "Appendix A.2" hold. Then as $n \to \infty$, we have that $d(\hat{\theta}_n, \theta_0) \to 0$ almost surely and*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma)$$

in distribution with Σ given in "Appendix A.2".

The proof of the results above is sketched in "Appendix A.2". For inference about β, it is apparent that one needs to estimate Σ and one common approach would be to employ the Louis's Formula. However, it can be seen below that this would be computationally intensive for the situation considered here and thus by following

Wen and Lin (2011) and others, we propose to employ the nonparametric bootstrap method (Efron 1981; Su and Wang 2016). Specifically, let Q be an integer and for each $1 \leq q \leq Q$, draw a new data set, denoted by $O^{(q)}$, of the sample size n with replacement from the original observed data $\{ O_i;\ i = 1, \ldots, n \}$. Let $\hat{\beta}_n^{(q)}$ denote the estimator of β defined above based on the bootstrap samples $O^{(q)}$, $q = 1, ..., Q$. respectively. Then one can estimate the covariance matrix of $\hat{\beta}_n$ by using the sample covariance matrix of the $\hat{\beta}_n^{(q)}$'s and the numerical results below suggest that this approach seems to work well.

5 A simulation study

In this section, we present some results obtained from a simulation study to evaluate the finite sample performance of the sieve maximum likelihood estimation procedure proposed in the previous sections. In the study, it was assumed that there exist two covariates $\mathbf{X_i^{obs}}$ and $\mathbf{X_i^{miss}}$. Note that as discussed above, we can write $f(\mathbf{X^{obs}}, \mathbf{X^{miss}}; \gamma)$ as

$$f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma) = f(\mathbf{X_i^{obs}}) f(\mathbf{X_i^{mis}} | \mathbf{X_i^{obs}}; \gamma).$$

For the generation of the covariates, we assumed that $f(\mathbf{X_i^{obs}})$ is $Bernoulli(0.6)$ or $normal(1, 0.25)$, and set $f(\mathbf{X_i^{miss}} \mid \mathbf{X_i^{obs}})$ to be

$$Bernoulli\left(0.5\mathbf{X_i^{obs}} + 0.73(1 - \mathbf{X_i^{obs}})\right) \text{ or } normal(1.5 + \mathbf{X_i^{obs}}, 0.25).$$

Given the covariates, the failure times of interest T_i's were generated based on model (1) with $\Lambda_0(t) = t^3$, t or $\Lambda_0(t) = log(1 + t)$. For the missing mechanism, we considered the following two situations

$$P(R_i = 1 | O_i) = \frac{\exp\{U + V + \mathbf{X_i^{obs}}\}}{1 + \exp\{U + V + \mathbf{X_i^{obs}}\}},$$

and

$$P(R_i = 1 | O_i) = \frac{\exp\{0.22U + 0.22V + 0.22\mathbf{X_i^{obs}}\}}{1 + \exp\{0.22U + 0.22V + 0.22\mathbf{X_i^{obs}}\}},$$

which will be referred to as set-ups I and II and correspond to the missing rates of 30% and 40%, respectively. For the generation of the observation times or censoring intervals, it was assumed that the U_i's and V_i's follow the uniform distribution over the region $\{(u, v) : 0 \leq u \leq 0.28, u + 0.8 \leq v \leq 1.2\}$. The results given below are based on the sample size $n = 200$ with 1000 replications.

Table 1 gives the obtained results on estimation of the regression parameters β_1 and β_2 with their true values being $\{0.2, 0.5\}$ and $\{0.5, 1\}$, respectively, and under the set up I for the missing mechanism. Here for the I-spline approximation to the

Table 1 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$, 30% missing and censoring rates

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	– 0.021	0.199	0.196	95.2	– 0.018	0.225	0.227	94.7
		CC	0.064	0.254	0.261	93.7	0.041	0.248	0.240	94.6
		MI	– 0.063	0.195	0.195	93.5	– 0.220	0.152	0.151	66.8
		Full	– 0.026	0.196	0.205	94.2	0.000	0.197	0.200	95.8
0.2	0.5	Proposed	– 0.024	0.198	0.198	94.7	– 0.010	0.226	0.232	94.4
		CC	0.041	0.246	0.246	94.2	0.038	0.242	0.248	94.6
		MI	– 0.049	0.192	0.191	94.3	– 0.204	0.154	0.153	72.3
		Full	– 0.017	0.196	0.195	95.4	0.014	0.193	0.196	95.4
0.5	1	Proposed	– 0.093	0.211	0.212	92.8	– 0.052	0.229	0.228	95.2
		CC	0.091	0.306	0.306	93.4	0.125	0.307	0.313	92.2
		MI	– 0.180	0.185	0.185	83.3	– 0.453	0.142	0.141	9.30
		Full	– 0.032	0.212	0.212	94.8	– 0.009	0.211	0.212	95.2

cumulative baseline hazards function, we took $s = 3$, the degree or order of the spline basis functions, and $k_n = 5$, the number of knots, and chose the knots equally spaced between the smallest and largest observation times by following Wang et al. (2016). In the table, we calculated the estimated bias given by the average of the estimates minus the true value (Bias), the sample standard error (SE), the average of the estimated standard error (ESE), and the 95% empirical coverage probability (CP). For comparison, we also applied the naive or complete data approach, which deletes the subjects with missing covariates, and the full data approach, assuming no missing covariates, which are denoted by CC and Full in the table, respectively. In addition, we also considered the most commonly used multiple imputation method (Horton and Kleinman 2007; Schomaker and Heumann 2018), denoted by MI in the table.

One can see from Table 1 that the proposed method seems to give good performance, which is similar to that of the Full approach. Both are better than the CC method and in particular, the CC methods gave larger biases. The results also suggest that the proposed method gave much better performance than the multiple imputation method, which clearly should not be used for estimation of β_2. In addition, the results on the coverage probabilities indicate that the normal approximation to the distribution of the proposed estimator appears to be reasonable. To further see this, we investigated the quantile plots of the standardized estimator against the standard normal distribution and present them in Fig. 1, which again suggest the normal approximation is appropriate.

Tables 2, 3, 4, 5 and 6 present the estimation results obtained similarly as Table 1. Specifically, in Tables 2 and 3, the same set-up as Table 1 was used except that Table 2 considered the set up II for the missing mechanism and Table 3 used $\Lambda_0(t) = t$. In Table 4, instead of discrete covariates, both covariates were generated from the normal distribution mentioned above and Table 5 investigated the situation where $\Lambda_0(t) = log(1 + t)$ with all other set-ups being the same as in Table 1. Instead of 30% censoring rate, we considered 50% censoring rate in Table 6 also with the other

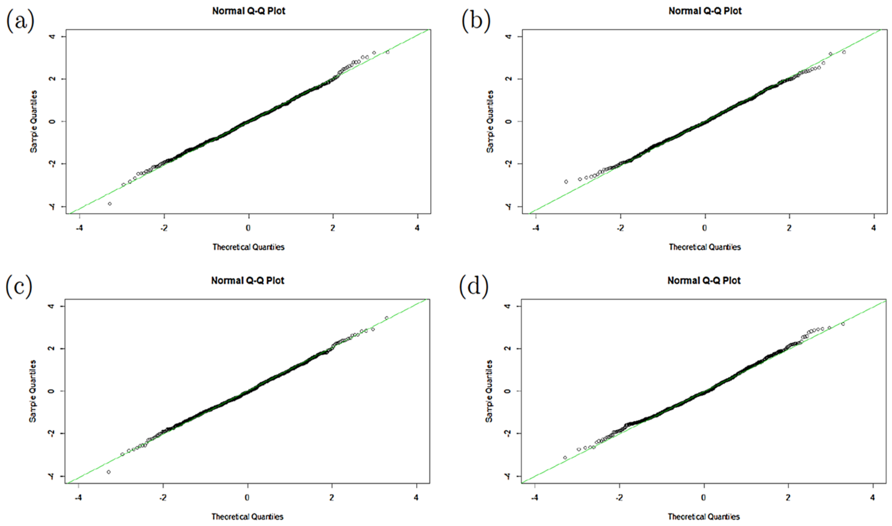

Fig. 1 Quantile plots of the standardized estimates of **a** β_1 with $\beta_1 = \beta_2 = 0.5$, (b) β_2 with $\beta_1 = \beta_2 = 0.5$, **c** β_1 with $\beta_1 = 0.5$ and $\beta_2 = 1$, and **d** β_2 with $\beta_1 = 0.5$ and $\beta_2 = 1$

Table 2 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$ and 40% missing rate and 30% censoring rate

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	– 0.031	0.200	0.198	94.9	– 0.021	0.245	0.243	94.6
		CC	0.036	0.279	0.285	94.8	0.062	0.275	0.285	94.5
		MI	– 0.078	0.193	0.192	92.7	– 0.281	0.145	0.143	49.6
		Full	– 0.023	0.196	0.196	95.4	– 0.024	0.197	0.198	94.6
0.2	0.5	Proposed	– 0.035	0.199	0.199	94.5	– 0.010	0.246	0.251	95.1
		CC	0.014	0.269	0.265	94.6	0.026	0.267	0.274	93.4
		MI	– 0.063	0.191	0.190	94.0	– 0.267	0.146	0.145	53.2
		Full	– 0.017	0.195	0.196	95.4	– 0.014	0.196	0.195	95.4
0.5	1	Proposed	– 0.117	0.213	0.212	92.8	– 0.070	0.247	0.250	93.0
		CC	0.074	0.349	0.339	94.5	0.124	0.398	0.350	92.9
		MI	– 0.210	0.184	0.185	79.7	– 0.564	0.135	0.134	1.00
		Full	– 0.055	0.212	0.210	94.1	– 0.037	0.211	0.204	95.4

set-ups being the same as in Table 1. One can see that in all situations, the proposed method and the Full approach gave similar performance and both seem to do well. In contrast, both the CC method and the multiple imputation method may yield biased estimates and low coverage probabilities.

As pointed out above, the focus here has been on the situation with missing at random and a natural question is how the proposed method would perform if one faces non-ignorable missing. To see this, we repeated the study giving Table 1 except that

Table 3 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t$ and 30% missing rate and 30% censoring rate

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	0.021	0.203	0.213	94.7	0.049	0.243	0.247	94.3
		CC	0.101	0.250	0.248	93.3	0.082	0.245	0.237	93.0
		MI	−0.041	0.185	0.185	94.6	−0.200	0.149	0.150	74.0
		Full	0.016	0.194	0.196	94.4	0.005	0.194	0.189	95.3
0.2	0.5	Proposed	−0.015	0.196	0.196	95.4	0.004	0.235	0.230	94.8
		CC	0.056	0.242	0.238	94.5	0.064	0.239	0.227	93.3
		MI	−0.043	0.185	0.185	94.1	−0.197	0.151	0.150	74.1
		Full	0.007	0.191	0.189	95.2	−0.000	0.191	0.188	95.1
0.5	1	Proposed	0.013	0.250	0.247	95.2	0.148	0.290	0.278	92.5
		CC	0.150	0.297	0.308	92.0	0.198	0.303	0.312	89.9
		MI	−0.109	0.191	0.190	90.9	−0.371	0.155	0.155	31.7
		Full	0.071	0.223	0.221	94.0	0.108	0.222	0.227	92.6

Table 4 Estimation of regression parameters β_1 and β_2 with continuous covariates, $\Lambda_0(t) = t^3$, and 30% missing and censoring rates

Scenario		Estimator	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0	−0.1	Proposed	0.032	0.232	0.230	96.0	−0.068	0.096	0.112	93.0
		CC	0.196	0.404	0.308	81.0	−0.156	0.231	0.179	77.0
		MI	−0.069	0.321	0.324	94.3	−0.109	0.210	0.209	92.7
		Full	−0.058	0.349	0.325	95.0	0.034	0.247	0.241	97.0
0.5	−0.5	Proposed	−0.078	0.267	0.271	95.0	0.011	0.142	0.139	96.0
		CC	0.364	0.369	0.370	86.0	−0.294	0.256	0.286	81.0
		MI	0.051	0.246	0.244	94.6	−0.088	0.172	0.176	92.3
		Full	−0.041	0.382	0.363	96.0	−0.023	0.270	0.264	95.0
0.25	−0.25	Proposed	−0.013	0.215	0.180	94.0	−0.017	0.129	0.138	95.0
		CC	0.103	0.202	0.224	87.0	−0.246	0.109	0.102	38.0
		MI	0.049	0.261	0.255	93.3	−0.083	0.185	0.180	91.1
		Full	−0.007	0.257	0.242	92.0	−0.023	0.186	0.180	93.0

instead of set up I for the missing mechanism, we considered the following mechanism

$$P(R_i = 1|O_i) = \frac{\exp\{0.5U + 0.22V + 0.22\mathbf{X}_\mathbf{i}^{\mathbf{obs}} + 0.22\mathbf{X}_\mathbf{i}^{\mathbf{miss}}\}}{1 + \exp\{0.5U + 0.22V + 0.22\mathbf{X}_\mathbf{i}^{\mathbf{obs}} + 0.22\mathbf{X}_\mathbf{i}^{\mathbf{miss}}\}}.$$

Table 7 gives the results on estimation of the regression parameters β_1 and β_2 given by the four methods discussed above and they suggest that again both the proposed

Table 5 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = log(1+t)$ and 30% missing and censoring rates

Scenario		Estimator	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	0.018	0.210	0.207	94.7	0.041	0.247	0.247	94.3
		CC	– 0.242	0.209	0.212	80.1	– 0.289	0.211	0.207	71.3
		MI	– 0.028	0.194	0.190	94.3	– 0.191	0.161	0.160	77.3
		Full	0.016	0.197	0.200	94.7	0.017	0.203	0.200	95.4
0.2	0.5	Proposed	– 0.012	0.206	0.206	95.1	– 0.032	0.251	0.251	95.1
		CC	– 0.200	0.219	0.221	85.4	– 0.266	0.217	0.214	75.8
		MI	– 0.039	0.198	0.194	94.1	– 0.192	0.166	0.165	79.0
		Full	0.008	0.201	0.202	94.5	0.000	0.203	0.204	95.6
0.5	1	Proposed	0.020	0.244	0.235	95.5	0.080	0.266	0.262	94.8
		CC	– 0.101	0.231	0.231	92.2	– 0.121	0.226	0.221	92.3
		MI	– 0.085	0.196	0.194	92.6	– 0.345	0.155	0.55	38.3
		Full	0.016	0.197	0.200	94.7	0.017	0.203	0.200	95.4

Table 6 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$ and 30% missing rate and 50% censoring rate

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	– 0.013	0.218	0.216	95.0	0.018	0.269	0.264	94.8
		CC	– 0.268	0.237	0.233	80.0	– 0.300	0.231	0.228	74.5
		MI	– 0.052	0.208	0.203	94.3	– 0.226	0.170	0.170	72.7
		Full	0.005	0.209	0.212	94.8	– 0.004	0.202	0.211	95.6
0.2	0.5	Proposed	– 0.025	0.220	0.221	95.6	0.014	0.280	0.280	95.5
		CC	– 0.206	0.253	0.250	86.3	– 0.273	0.241	0.243	80.5
		MI	– 0.057	0.218	0.213	93.8	– 0.223	0.184	0.181	75.4
		Full	0.002	0.214	0.218	95.2	– 0.005	0.214	0.223	95.9
0.5	1	Proposed	– 0.051	0.214	0.216	94.9	0.030	0.264	0.259	94.0
		CC	– 0.116	0.241	0.234	90.6	– 0.187	0.242	0.243	88.2
		MI	– 0.112	0.192	0.190	90.2	– 0.424	0.164	0.162	24.7
		Full	0.041	0.207	0.213	96.0	– 0.055	0.212	0.212	94.5

method and the Full method gave good performance. However, the CC and multiple imputation methods did not seem to provide reasonable results. Of course, one may be careful about the proposed method since no theoretical justification can be provided yet.

Table 7 Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$ and 30% missing and censoring rates under the non-ignorable missing mechanism

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	– 0.035	0.193	0.198	95.1	– 0.024	0.236	0.234	95.0
		CC	0.022	0.427	0.419	94.1	– 0.091	0.408	0.401	94.3
		MI	– 0.071	0.194	0.194	92.7	– 0.246	0.147	0.146	58.5
		Full	– 0.026	0.196	0.205	94.2	0.000	0.197	0.200	95.8
0.2	0.5	Proposed	– 0.038	0.196	0.197	94.9	– 0.015	0.237	0.234	94.6
		CC	0.022	0.245	0.241	94.7	– 0.097	0.397	0.404	94.2
		MI	– 0.056	0.192	0.191	94.1	– 0.231	0.150	0.150	65.7
		Full	– 0.017	0.196	0.195	95.4	0.014	0.193	0.196	95.4
0.5	1	Proposed	– 0.061	0.208	0.213	94.2	– 0.010	0.228	0.228	95.1
		CC	0.038	0.256	0.252	94.3	– 0.141	0.404	0.398	93.6
		MI	– 0.197	0.185	0.186	81.8	– 0.503	0.137	0.138	3.90
		Full	– 0.032	0.212	0.212	94.8	– 0.009	0.211	0.212	95.2

6 An application

Now we apply the sieve maximum likelihood estimation procedure proposed in the previous sections to the set of data arising from Alzhehelmer's Disease Neuroimaging Initiative, discussed by Li et al. (2020) among others. It is a longitudinal study and among others, one variable of interest is the Alzheimers disease (AD) conversion. Due to the nature of the study, only interval-censored data are available on the occurrence time of the AD conversion, and the participants in the study are classified into three groups based on their cognitive conditions, cognitively normal, mild cognitive impairment and Alzheimer's disease. By following Li et al. (2020) and others, we will focus on the patients in the mild cognitive impairment group to determine significant baseline prognostic factors or covariates for the AD conversion.

The data consist of five baseline covariates. They are the Rey Auditory Verbal Learning Test (RAVLT), the Middle temporal gyrus (MidTemp) from Neuroimaging, the participants's Alzheimer's Disease Assessment Scale 13 items (ADAS13), the functional assessment questionnaire score (FAQ), and the participant's baseline age (Age). Among the 396 participants in the mild cognitive impairment group, around 20% of them have missing information on the MidTemp. Also there are 3 subjects with missing ADAS13 and 3 subjects with missing FAQ, and in the analysis below, we will exclude these six subjects for simplicity. As mentioned above, Li et al. (2020) discussed the same problem but they only considered the situation where there do not exist missing covariates.

Table 8 presents the analysis results given by the proposed sieve maximum likelihood estimation procedure, including the estimated covariate effect (Estimate), the estimated standard error (SE) and the p-value for testing the covariate effect being zero. For comparison, we also include in the table the results given by Li et al. (2020)

Table 8 Analysis results of Alzhehelmer's Disease data

Covariate	Method	Estimate	SE	p-value
RAVLT	Li et al. (2020)	– 0.679	0.324	0.018
	Proposed	– 0.305	0.096	0.001
Midtemp	Li et al. (2020)	– 0.434	0.290	0.072
	Proposed	– 0.291	0.075	0.000
ADAS13	Li et al. (2020)	0.380	0.690	0.291
	Proposed	0.410	0.094	0.000
FAQ	Li et al. (2020)	0.426	0.244	0.040
	Proposed	0.410	0.071	0.000
Age	Li et al. (2020)	– 0.364	0.274	0.092
	Proposed	– 0.087	0.083	0.147

based on the 316 subjects with complete information on the MidTemp. One can see that the proposed method suggests that except Age, all other four covariates, RAVLT, MidTemp, ADAS13 and FAQ, had significant effects on the AD conversion. In contrast, the approach that ignored the missing information indicates that MidTemp may only have some mild effect and ADAS13 had no effect on predicting the AD conversion. In addition, as expected, the proposed method gave more efficient estimates than Li et al. (2020) for all covariates.

7 Discussion and concluding remarks

In this paper, we discussed the inference about the proportional hazards model when one faces interval-censored failure time data with missing covariates, and for the problem, a sieve maximum likelihood estimation procedure was proposed. In the method, I-spline functions were employed to approximate the unknown baseline cumulative hazard function and a Poisson-based EM algorithm was developed. The proposed estimator of regression parameters were shown to be consistent and asymptotically normal, and the numerical studies indicated that the proposed method seems to work well for practical situations and should be used when covariates are missing at random.

As pointed out above, the focus here has been on the situation where covariates may be missing at random, meaning that the missingness depends only on the observed values. It is worth to note that sometimes one may face more complicated situations where the missingness could depend on both observed and missing values, which is often referred to as nonignorable missing (Du et al. 2021). For the latter case, a valid inference procedure usually requires one to make some assumptions on or model the missingness, and it is easy to see that for the situation, one could often have to deal with the model misspecification issue.

There exist several directions for future research. One is that the focus here has been on the proportional hazards model (1) and it is apparent that the same type of missing data could happen to other commonly used regression models such as the additive hazards model or the linear transformation model. It would be useful to develop some estimation procedures similar to that proposed above for these latter models. In the preceding sections, we have assumed that covariates are time-independent and it is clear that sometimes there may exist time-dependent covariates. Thus it would be helpful to generalize the proposed approach to allow for time-dependent covariates. Also in the above, it has been assumed that we observe case II interval-censored data and as pointed out by some authors, in practice, one may face a more general type of interval-censored data, case K interval-censored data (Sun 2006; Wang et al. 2016). It is apparent that the method given above cannot be directly applied to this latter situation and in other words, some generalizations of it are needed.

Acknowledgements The authors wish to thank the Editor-in-Chief, Dr. Mei-Ling Ting Lee, the Associate Editor and two reviewers for their many helpful and insightful comments and suggestions that greatly improved the paper. The research was partially supported by grants from the Natural Science Foundation of China [Grant Number 11731011], a grant from key project of the Yunnan Province Foundation, China [Grant Number 202001BB050049]. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Appendix

A.1. E-step of the EM algorithm for continuous covariates

In the E-step of the EM algorithm developed in Sect. 3, we need to calculate the expectations $E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$ and $E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$. As described there, when missing covariates are categorical, they are some summations and can be expressed in the closed form. However, for continuous covariates, this will not be the case and instead we have to deal with the integrals that do not have a closed form. More specifically, we have that

$$E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \int_{\mathbf{X_{miss}}} \frac{\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\mathbf{exp}(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i}^{\mathbf{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\delta_{\mathbf{1i}}}{1 - \mathbf{exp}\{-\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\mathbf{exp}(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i}^{\mathbf{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\}} \\ \times p(\mathbf{X_i^{miss}}|\mathbf{O_i}, \theta^{(\mathbf{d})})\mathbf{dX_i^{miss}},$$

and

$$E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \int_{\mathbf{X_i^{miss}}} \frac{\{\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{U_i}) - \mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\}\mathbf{exp}(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i}^{\mathbf{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\delta_{\mathbf{2i}}}{1 - \mathbf{exp}[-\{\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{U_i}) - \mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\}\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i}^{\mathbf{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})]} \\ \times p(\mathbf{X_i^{miss}}|\mathbf{O_i}, \theta^{(\mathbf{d})})\mathbf{dX_i^{miss}}$$

by using the notation defined before.

To calculate the integrals above, by following Herring and Ibrahim (2001), one can employ the Monte-Carlo estimation approach, which draws the sample from

$$p_{ij} = P(\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \frac{\mathbf{f(U_i, V_i}, \delta_{\mathbf{1i}}, \delta_{\mathbf{2i}}, \delta_{\mathbf{3i}}|\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}})\mathbf{f(X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}}; \gamma^{(\mathbf{d})})}{\int_{\mathbf{X_i^{mis}}} \mathbf{f(U_i, V_i}, \delta_{\mathbf{1i}}, \delta_{\mathbf{2i}}, \delta_{\mathbf{3i}}|\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}})\mathbf{f(X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}}; \gamma^{(\mathbf{d})})}$$
$$\propto \mathbf{f(U_i, V_i}, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}})\mathbf{f(X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}}; \gamma^{(\mathbf{d})}).$$

Note that $f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}})$ is log-concave (Ibrahim et al. 1999) and if $f(\mathbf{X_i}^{\mathbf{obs}}, \mathbf{X_i}^{\mathbf{mis}}; \gamma^{(\mathbf{d})})$ belongs to the exponential family, the logrithm of $P(\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$ is concave. It follows that one can use the Gibbs sampler (Gilks and Wild 1992) and adaptive rejection algorithm (Gilks and Wild 1992) to sample from $P(\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$.

More specifically for the determination of $E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$, for each subject with missing covariate $\mathbf{X_i^{miss}}$, we first apply the Gibbs sampler and adaptive reject algorithm to draw the sample $s_{i,1}, ..., s_{i,n_i}$ of size n_i from $p(\mathbf{X_i^{miss}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$. Then the conditional expectation can be approximated by

$$E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \frac{1}{\mathbf{n_i}} \sum_{\mathbf{k=1}}^{\mathbf{n_i}} \frac{\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\mathbf{exp}(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{s_{i,k}})\delta_{\mathbf{1i}}}{1 - \mathbf{exp}\{-\mathbf{\Lambda}^{(\mathbf{d})}(\mathbf{V_i})\mathbf{exp}(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{s_{i,k}})\}}.$$

In comparison to the categorical covariate situation, the above operation can be regarded as replacing each x_i^{miss} by n_i sampled values with equal weight. It is apparent that $E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$ can be calculated similarly.

A.2. Proofs of the asymptotic properties

In this Appendix, we will sketch the proof for the consistency and asymptotic normality of the proposed estimators given in Theorem 1 by employing the empirical process theory and nonparametric techniques. Define $Pf = \int f(x)dP(x)$ and $P_n f = n^{-1}\sum_{i=1}^{n} f(X_i)$ for a function f, a probability function P and a sample $X_1, \ldots, X_n$. For the proof, we need the following regularity conditions.

(A1) Assume that $\Lambda(\tau_1) < \infty$, $\Lambda(\tau_2) < \infty$, and there exists a positive constant a such that $P(V - U > a) > 0$. Also the union of the supports of U and V is contained in the interval $[r_1, r_2]$ with $0 < r_1 < r_2 < +\infty$.

(A2) The function Λ_0 is continuously differentiable on $[r_1, r_2]$, and satisfies $M^{-1} < \Lambda_0(r_1) < \Lambda_0(r_2) < M$ for some positive constant M.

(A3) The set of covariates (X, Z) has bounded support.

(A4) The conditional distribution $f(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}}; \gamma)$ is identifiable and has continuous second-order derivatives with respect to γ, and $-E_0[\partial^2/\partial\gamma^2)\log f(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}}; \gamma_0)]$ is positive definite.

(A5) For any $(\theta, \mathbf{\Lambda})$ near $(\theta_\mathbf{0}, \mathbf{\Lambda_0})$, $P_0(\log L(\theta, \mathbf{\Lambda}) - \log L(\theta_\mathbf{0}, \mathbf{\Lambda_0}) \leq -K(||\theta - \theta_\mathbf{0}||^2 + ||\mathbf{\Lambda} - \mathbf{\Lambda_0}||^2)$ for a fixed constant $K > 0$.

First we will prove the consistency and for this, we will verify the conditions of Theorem 5.7 of Van der Vaart (1998). Let $BV_\omega[r_1, r_2]$ denote the functions whose total variation in $[r_1, r_2]$ are bounded by a given constant. Then the class of functions

$$F_\omega = \left\{ \int_0^{U_k} \exp\{\beta^T X_i\} d\Lambda(s) : \Lambda \in BV_\omega[r_1, r_2] \right\}$$

is a convex hull of functions $\{I(U_k \geq s)\exp\{\beta^T X_i\}$ and thus it is a Donsker class. Furthermore,

$$\exp\left(-\int_0^{U_k} \exp\{\beta^T X_i\} d\Lambda(s) \right) - \exp\left(-\int_0^{U_{k+1}} \exp\{\beta^T X_i\} d\Lambda(s) \right)$$

is bounded away from zero. Therefore, $l(\theta, \hat{\alpha}|\mathbf{O}) = \log L(\theta, \hat{\alpha}|\mathbf{O})$ belongs to some Donsker class due to the preservation property of the Donsker class under Lipschitz-continuous transformations. Then we can conclude that $\sup_{\theta \in \Theta_n} |P_n l(\theta, \hat{\alpha}|\mathbf{O}) - P_n l(\theta_0, \hat{\alpha}|\mathbf{O})|$ converges in probability to 0 as $n \to 0$.

Now we verify that another condition of Theorem 5.7 of Van der Vaart (1998) also holds. That is, for any $\varepsilon > 0$, we have

$$\sup_{d(\theta, \theta_0) > \varepsilon} Pl(\theta, \hat{\alpha}|\mathbf{O}) < Pl(\theta_0, \hat{\alpha}|\mathbf{O}).$$

Note that this condition is satisfied if we can prove the model is identifiable. According to condition (A4) and similar arguments to the proof of Theorem 2.1 of Chang et al. (2007), we can show the identifiability of the model parameters. Now, by Theorem 5.7 of Van der Vaart (1998), we have $d(\hat{\theta}_n, \theta_0) = o_p(1)$, which completes the proof of consistency.

Before proving the asymptotic normality, we will need to establish the convergence rate. For this, we will first define the covering number of the class $\mathcal{L} = \{l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta\}$ and establish a needed lemma.

Lemma 1 *Assume that Conditions (A1), (A3)–(A4) hold. Then the covering number of the class* $\mathcal{L} = \{l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta\}$ *satisfies*

$$N(\epsilon, \mathcal{L}, L_2(P)) = O(\epsilon^{-1}).$$

Proof of Lemma 1 The proof is similar to that of Zeng et al. (2016) and Hu et al. (2017) and thus omitted.

To establish the convergence rate, for any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0, \Lambda_{n0})$. Following the calculation of (Shen and Wong 1994, p. 597), we can establish that $\log N_{[]}(\epsilon, \mathcal{F}_\eta, \| \ . \ \|_2) \leq CN\log(\eta/\epsilon)$ with $N = m + 1$, where $N_{[]}(\epsilon, \mathcal{F}_\eta, d)$ denotes the bracketing number (see the Definition 2.1.6 in Van Der Vaart and Wellner 1996) with respect to the metric or semi-metric d of a function class $\mathcal{F}$. Moreover, some algebraic calculations lead to $\| \ l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \ \|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \in \mathcal{F}_\eta$. Therefore, by Lemma 3.4.2 of Van Der Vaart and

Wellner (1996), we obtain

$$E_p \parallel n^{1/2}(P_n - P) \parallel_{\mathcal{F}_\eta} \leq C J_\eta(\epsilon, \mathcal{F}_\eta, \parallel . \parallel_2) \left\{ 1 + \frac{J_\eta(\epsilon, \mathcal{F}_\eta, \parallel . \parallel_2)}{\eta^2 n^{1/2}} \right\}, \qquad (S)$$

where $J_{[]}(\eta, \mathcal{F}_\eta, \parallel . \parallel_2) = \int_0^\eta \{log N_{[]}(\epsilon, \mathcal{F}_\eta, \parallel . \parallel_2)\}^{1/2} d\epsilon$. The right-hand side of (S) yields $\phi_n(\eta) = C\eta^{1/2}(1 + \frac{\eta^{1/2}}{\eta^2 n^{1/2}} M_1)$, where M_1 is a positive constant. Then $\phi_n(\eta)/\eta$ is a decreasing function, and $n^{2/3}\phi_n(-1/3) = O(n^{1/2})$. According the theorem 3.4.1 of Van Der Vaart and Wellner (1996), we can conclude that $d(\hat{\theta}, \theta_0) = O_p(n^{-1/3})$.

Now we prove the asymptotic normality of $\hat{\beta}_n$. Following the proof of Theorem 2 in Zeng et al. (2016), one can obtain that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = (E[\{l_\beta - l_\Lambda(s^*)\}\{l_\beta - l_\Lambda(s^*)\}^T)^{-1} G_n\{l_\beta - l_\Lambda(s^*)\} + o_p(1),$$

where l_β is the score function for β, $l_\Lambda(s^*)$ is the score function along this submodel $d\Lambda_{\epsilon,s^*} = (1 + \epsilon s^*)d\Lambda$. This implies that the influence function for $\hat{\beta}_n$ is exactly the efficient influence function, so that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound. □

References

Chen K, Jin Z, Ying Z (2002) Semiparametric analysis of transformation models with censored data. Biometrika 89:659–668

Chen HY, Little RJ (1999) Proportional hazards regression with missing covariates. J Am Stat Assoc 94(447):896–908

Chang IS, Wen CC, Wu YJ (2007) A profile likelihood theory for the correlated gamma-frailty model with current status family data. Stat Sin 17:1023–1046

Du MY, Li HQ, Sun JG (2021) Regression analysis of censored data with nonignorable missing covariates and application to Alzheimer Disease. Comput Stat Data Anal 157:1–15

Efron B (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 68(3):589–599

Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. J R Stat Soc Ser C (Appl Stat) 41:337–348

Herring AH, Ibrahim JG (2001) Likelihood-based methods for missing covariates in the Cox proportional hazards model. J Am Stat Assoc 96:292–302

Horton NJ, Kleinman KP (2007) Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat 61(1):79–90

Hu T, Zhou Q, Sun J (2017) Regression analysis of bivariate current status data under the proportional hazards model. Can J Stat 45:410–424

Ibrahim JG, Lipsitz SR, Chen MH (1999) Missing covariates in generalized linear models when the missing data mechanism is nonignorable. J R Stat Soc Ser B (Stat Methodol) 61(1):173–190

Li S, Hu T, Wang P et al (2017) Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments. Comput Stat Data Anal 110:75–86

Lipsitz SR, Ibrahim JG, Zhao LP (1994) A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. J Am Stat Assoc 94:1147–1160

Little RJ, Rubin DB (2002) Statistical analysis with missing data. Wiley, New York

Li S, Wu Q, Sun J (2020) Penalized estimation of semiparametric transformation models with interval-censored data and application to Alzheimer disease. Stat Methods Med Res 29(8):2151–2166

Ma L, Hu T, Sun J (2015) Sieve maximum likelihood regression analysis of dependent current status data. Biom J 102:731–738

McMahan CS, Wang L, Tebbs JM (2013) Regression analysis for current status data using the EM algorithm. Stat Med 32:4452–4466

Qi L, Wang CY, Prentice RL (2005) Weighted estimators for proportional hazards regression with missing covariates. J Am Stat Assoc 100:1250–1263

Ramsay JO (1988) Monotone regression splines in action. Stat Sci 3(4):425–441

Schomaker M, Heumann C (2018) Bootstrap inference when using multiple imputations. Stat Med 37(14):2252–2266

Sun J (2006) The statistical analysis of interval-censored failure time data. Springer, New York

Shen X, Wong WH (1994) Convergence rate of sieve estimates. Ann Stat 22:580–615

Su YR, Wang JL (2016) Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data. Ann Stat 44(3):1298–1331

Van der Vaart AW (1998) Asymptotic statistic. Cambridge University Press, Cambridge

Van Der Vaart A, Wellner JA (1996) Weak convergence and empirical processes: with applications to statistics. Springer, New York

Wen CC, Lin CT (2011) Analysis of current status data with missing covariates. Biometrics 67:760–769

Wang L, McMahan CS, Hudgens MG et al (2016) A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. Biometrics 72:222–231

Zhao S, Hu T, Ma L et al (2015) Regression analysis of informative current status data with the additive hazards model. Lifetime Data Anal 21:241–258

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

Zhou H, Pepe MS (1995) Auxiliary covariate data in failure time regression. Biometrika 82(1):139–149