Annals of the Institute of Statistical Mathematics (2024) 76:617–648
https://doi.org/10.1007/s10463-024-00899-5

Empirical likelihood MLE for joint modeling right censored survival data with longitudinal covariates

Jian-Jian Ren[1] · **Yuyin Shi**[2]

Received: 5 November 2023 / Accepted: 9 January 2024 / Published online: 29 April 2024

Abstract
Up to now, almost all existing methods for joint modeling survival data and longitudinal data rely on parametric/semiparametric assumptions on longitudinal covariate process, and the resulting inferences critically depend on the validity of these assumptions that are difficult to verify in practice. The kernel method-based procedures rely on choices of kernel function and bandwidth, and none of the existing methods provides estimate for the baseline distribution in proportional hazards model. This article proposes a proportional hazards model for joint modeling right censored survival data and intensive longitudinal data taking into account of within-subject historic change patterns. Without any parametric/semiparametric assumptions or use of kernel method, we derive empirical likelihood-based maximum likelihood estimators and partial likelihood estimators for the regression parameter and the baseline distribution function. We develop stable computing algorithms and present some simulation results. Analyses of real dataset are conducted for smoking cessation data and liver disease data.

Keywords Empirical likelihood · Intensive longitudinal data · Maximum likelihood estimator · Proportional hazards model · Right censored data

1 Introduction

In medical research, epidemiology, social and behavior science studies, etc., the interrelationships between time-to-event variable, i.e., survival time, and longitudinal covariates are often the primary research interest. Due to the challenges encountered in some important clinical trials on AIDS and cancer research, statisticians

✉ Jian-Jian Ren
jjren@umd.edu

[1] Statistics Program, Department of Mathematics, University of Maryland, College Park, MD 20742, USA

[2] U.S. Food and Drug Administration, Center for Biologics Evaluation and Research (CBER), 10903 New Hampshire Avenue, Silver Spring, MD 20993-0002, USA

started modeling survival data and longitudinal data jointly, viz. Tsiatis et al. (1995), Henderson et al. (2000), Hsieh et al. (2006), Ding and Wang (2008), Su and Wang (2012), and Cao et al. (2015); see review articles on this subject by Tsiatis and Davidian (2004), Li and Ren (2011). Such joint modeling procedure or methodology has broad applications in many scientific research fields, but it is a considerably difficult problem due to censoring on the survival time and the fact that the longitudinal covariate is an unknown and not completely observed stochastic process. Up to now, almost all existing methods rely on parametric or semiparametric assumptions on the longitudinal covariate process, and the resulting inferences critically depend on the validity of these assumptions that are difficult to verify in practice. The kernel method-based procedures (Cao et al. 2015) rely on choices of kernel function and bandwidth, and such methods are not really necessary for very intensive longitudinal data. Moreover, none of the existing methods considers model checking, nor provides estimate for the baseline distribution function (d.f.) in proportional hazards model, which can play a role in constructing goodness of fit test for model checking (Ren and He 2011).

In statistical literature, other important problems such as joint modeling censored survival time and *intensive longitudinal covariates* and joint modeling censored survival time and *multi-phase intensive longitudinal covariates* have not been carefully studied. Different from the conventional longitudinal covariates, the intensive longitudinal covariates and the multi-phase intensive longitudinal covariates refer to situations where considerably more repeated measurements are taken within a relative short period of time. The intent of collecting these intensive or multi-phase intensive longitudinal data is to conduct micro-analytic assessment of the patterns of change; thus, in these situations, it is important to consider the within-subject *historic* patterns of change in the modeling procedure. Here, *historic patterns of change* refer to the change patterns of the longitudinal covariates over certain time interval of interest. However, in the literature, there have not been any modeling procedures that directly study the relationship between survival time and within-subject historic patterns of change, while intensive longitudinal covariates and multi-phase intensive longitudinal covariates have been encountered in various studies, such as smoking cessation data (Shiffman et al. 1996; Piper et al. 2009) and child development data (Cole et al. 2011).

This article considers proportional hazards (PH) model for joint modeling right censored survival data and longitudinal covariates with historic change patterns taken into account via the use of modeling function. Without any parametric or semiparametric assumptions made on longitudinal covariate process and without any kernel or bandwidth choices, we use empirical likelihood method (Owen 1988, 2001; DiCiccio et al. 1991; Chen and Qin 1993; Qin and Lawless 1994; Mykland 1995; Ren 2001, 2008; Zhou 2005, 2016; Ren and Zhou 2011; Ren and Riddlesworth 2014; Cai and Chen 2018; Wang et al. 2019; Muller et al. 2019; Chang and McKeague 2022) to derive the maximum likelihood estimator (MLE) for our proposed PH model.

In particular, Ren and Zhou (2011) used empirical likelihood (Owen 1988) to derive the full likelihood function for regression parameter and baseline d.f. in the Cox model with only time-independent covariate, and explicitly derived the relation

between full likelihood function and Cox's partial likelihood function. They showed that the leading term of Taylor's expansion of the full-profile likelihood estimating function is in fact the Cox's partial likelihood estimating function. In this article, the version of partial likelihood estimating function under our PH model with unknown time-dependent covariate process is also derived via empirical likelihood method.

The rest of this article is organized as follows. Section 2 gives our proposed PH model setting with comparison to existing models and presents some real data examples; Section 3 derives the empirical likelihood-based MLE for regression parameter β_0 and baseline d.f. F_0 in our PH model for right censored survival time with longitudinal covariates and also derives the corresponding partial likelihood estimating function; Section 4 develops computing algorithm and presents some simulation results, where careful description of our simulation study setting is given and the factor of intensity level for the intensive longitudinal data is considered, then provides some discussions and concluding remarks; and Section 5 conducts real data analysis for smoking cessation data and liver disease data. All proofs and technical derivations are given in Appendix which is included in Supplementary Materials.

2 Proportional hazards model with historic change patterns

Throughout this article, we use the following notations for observed survival time T with time-independent covariate $\boldsymbol{Z}$ and longitudinal covariate process $\boldsymbol{X}(t)$:

$$(T_i, \boldsymbol{Z}_i, \boldsymbol{\mathcal{X}}_i), \quad i = 1, 2, \cdots, n, \tag{1}$$

where for the ith individual, T_i is the survival time, $\boldsymbol{Z}_i$ is a vector of q explanatory variables observed on time-independent covariate $\boldsymbol{Z} = (Z_1, \cdots, Z_q)^\top \in \mathbb{R}^q$ which does not change over the course of the studies, and the observed longitudinal data on unknown covariate process $\boldsymbol{X}(t) = (X_1(t), \cdots, X_p(t))^\top \in \mathbb{R}^p$ are denoted by

$$\boldsymbol{\mathcal{X}}_i = \big(\boldsymbol{X}_i(t_{i1}), \cdots, \boldsymbol{X}_i(t_{iJ_i})\big), \quad \text{where } \boldsymbol{X}_i(t_{ij}) = \big(X_{i1}(t_{ij}), \cdots, X_{ip}(t_{ij})\big)^\top \tag{2}$$

with $\boldsymbol{X}_i(t_{ij})$ as a vector of p explanatory variables observed on covariate $\boldsymbol{X} = (X_1, \cdots, X_p)^\top \in \mathbb{R}^p$ at time t_{ij}.

When T_i is subject to right censoring, the actually observed data on sample (1) are

$$(V_i, \delta_i, \boldsymbol{Z}_i, \boldsymbol{\mathcal{X}}_i), \quad i = 1, 2, \cdots, n, \tag{3}$$

where (V_i, δ_i) represents the right censored observation on T_i given by

$$V_i = \begin{cases} T_i, & \text{if } T_i \le C_i, \quad \delta_i = 1 \\ C_i, & \text{if } T_i > C_i, \quad \delta_i = 0 \end{cases} \tag{4}$$

with C_i as the right censoring variable which is independent of $(T_i, \boldsymbol{Z}_i, \boldsymbol{\mathcal{X}}_i)$. In practical situations, right censoring C_i occurs due to drop-outs or the end time of the study or any other possible factors. In statistical literature, this type of censored data without longitudinal covariates $\boldsymbol{\mathcal{X}}_i$ in (3) has been extensively studied under the Cox model in the past few decades; see Cox (1972), Efron (1977), Oakes (1977), Tsiatis

(1981), Andersen and Gill (1982), Fleming and Harrington (1991), Andersen et al. (1993), Ren and Zhou (2011), Ren and He (2011), among others. Our interest of the study here is the effects of covariate $(\boldsymbol{Z}, \boldsymbol{X}(\cdot))$ on survival time T based on the actually observed data (3).

2.1 Existing methods

The existing works for joint modeling right censored survival time and longitudinal covariates on the line of Tsiatis and Davidian (2004), Ding and Wang (2008), among others consider the following Cox's proportional hazards (PH) model for data (3):

$$\lambda(t \mid \boldsymbol{Z}_i, \boldsymbol{X}_i^t) = \lambda_0(t) \exp\{\boldsymbol{\beta}_Z^\top \boldsymbol{Z}_i + \boldsymbol{\beta}_X^\top \boldsymbol{X}_i(t)\}, \tag{5}$$

where the history of the longitudinal process up to time t is denoted by

$$\boldsymbol{X}_i^t = \{\boldsymbol{X}_i(s) \mid 0 \le s < t\}, \tag{6}$$

$\boldsymbol{\beta}_Z \in \mathbb{R}^q$ and $\boldsymbol{\beta}_X \in \mathbb{R}^p$ are regression parameters, $\lambda(t \mid \boldsymbol{Z}_i, \boldsymbol{X}_i^t)$ is the conditional hazard function of T given $\boldsymbol{Z} = \boldsymbol{Z}_i, \boldsymbol{X}^t = \boldsymbol{X}_i^t$, and $\lambda_0(t)$ is an arbitrary baseline hazard function with corresponding distribution function (d.f.) F_0.

Since the covariate process $\boldsymbol{X}_i(t)$ is only observable at time points $t_{i1} < \cdots < t_{iJ_i}$ as given in (2), almost all existing works make various parametric or semiparametric assumptions on $\boldsymbol{X}(t)$ in order to estimate the covariate process; see review articles by Tsiatis and Davidian (2004), Li and Ren (2011). Cao et al. (2015) used the kernel method-based procedures which rely on choices of kernel function and bandwidth. In practice, the parametric or semiparametric assumptions on covariate process $\boldsymbol{X}(t)$ are not really verifiable, and the bandwidth selection needs to be studied. Thus, it is of great interest and importance to develop methods without these assumptions/selections.

It should be noticed that above model (5) is different from the usual Cox model with time-dependent covariates. In PH model (5), $\boldsymbol{X}_i(t)$ is only observed at time t_{ij}, while for the Cox model with time-dependent covariates, $\boldsymbol{X}_i(t)$ is assumed to be known and observed fully at any time t; see Kalbfleisch and Prentice (2002), Tsiatis and Davidian (2004), Zeng and Cai (2005), among others.

As for the issue of modeling within-subject historic patterns of change mentioned in Section 1 of this article, one should notice that in above PH model (5) the regression covariates used on observed $\boldsymbol{\mathcal{X}}_i$ in (3) is $\boldsymbol{X}_i(t)$ which does not reflect the *history* of the observed longitudinal covariates $\boldsymbol{X}_i^t$ given by (6), i.e., does not reflect the *historic change patterns* in observed $\boldsymbol{\mathcal{X}}_i$. To address this point, we take a closer look at some real data examples in the next subsection.

2.2 Data examples

Example 1 AIDS Data. In the HIV clinical trials considered by Tsiatis et al. (1995), T_i is time to progression to AIDS or death of a patient, which is subject to right censoring (4). For each patient, time-independent covariates $\boldsymbol{Z}_i$ include treatment

assignment, demographic information, etc., while the longitudinal covariates $\boldsymbol{X}_i(t)$ include measurements of immunologic and virologic status such as CD4 count and viral RNA copy number which are taken at subsequent clinic visits. This is an example of ***right censored survival data with longitudinal covariates***, for which above PH model (5) is applicable. The *Liver Disease Data* considered in Section 5 of this paper are another example of the same type.

Example 2 Smoking Cessation Data. In a smoking cessation study (Shiffman et al. 1996), 151 smokers who had quit for at least 24 h were considered, and T_i is time to the first lapse, which is subject to right censoring (4) due to drop-outs and the end time of study. During the study, each participant was randomly sampled 5 times per day to record the measurements on covariate variables such as X_1 =[Urge to Smoke], X_2 =[Mood], etc. The resulting $\boldsymbol{\mathcal{X}}_i$ as given in (2) are *intensive longitudinal covariates*, because compared to Example 1, considerably more information was recorded on covariate $\boldsymbol{X}$ within a relatively short period of time. This is an example of ***right censored survival data with intensive longitudinal covariates***.

In an empirical analysis of smoking cessation data by Li et al. (2006), one of the research interests was the relationship between T_i and the changes in the level of urge to smoke. Since the longitudinal covariates of the participants are momentary feelings, they fluctuate greatly over time, thus using the average of $\boldsymbol{X}_i(t_{ij})$ in a particular time interval as the modeling function provides more reliable information which can be used to characterize the historic change patterns of a participant's level of urge to smoke.

Example 3 Child Development Data. In the study conducted by Cole et al. (2011), children of 18 months of age were invited for four lab visits at age 18, 24, 36 and 48 months, and T_i is the age at which a specified development goal is achieved for a child. As African child development data (Turnbull 1974), here T_i is subject to double censoring, because some children already achieved the development goal at the first visit which is left censored; and some children did not achieve it at the end of the study or dropped out which is right censored. To conduct micro-analytic assessment of the change patterns in child's emotion self-control, at each lab visit an 8-minute test was performed on the child, and the emotion variables such as anger and calm were recorded every 15 s. The resulting $\boldsymbol{\mathcal{X}}_i$ from 4 lab visits are *multi-phase intensive longitudinal covariates*, because there is rather long time between the visits, and at each visit considerably more information was recorded on the covariates $\boldsymbol{X}$ within a very short period of time. For each child, time-independent covariates $\boldsymbol{Z}_i$ include gender, regions, etc. This is an example of ***doubly censored survival data with multi-phase intensive longitudinal covariates***.

To characterize the historic change patterns of a child's anger expression during the 8-minute test, the modeling functions used by Cole et al. (2011) are the total amount of anger and the average duration of anger for the child in a particular time interval.

For doubly censored data, we refer to Chang and Yang (1987), Gu and Zhang (1993), Mykland and Ren (1996), among others. Our focus of this article is right censored survival time with longitudinal covariates (3)-(4), not doubly censored

survival time. But the intensive longitudinal covariates and the modeling functions considered in Cole et al. (2011) are of special interests to our proposed PH model with historic patterns of change under consideration; see description in the next subsection. As for the problem of PH model with multi-phase intensive longitudinal covariates, it is not considered in this article because it needs a separate and careful study.

2.3 Proposed proportional hazards model

As pointed out at the end of Sect. 2.1, in model (5), the regression covariates used on observed $\mathcal{X}_i$ in (3) are $\boldsymbol{X}_i(t)$, rather than $\boldsymbol{X}_i^t$ which, by (6), is the history of the longitudinal process up to time t. Motivated by data examples described in Section 2.2, we are interested in the effects of $(\boldsymbol{Z}_i, \boldsymbol{X}_i^t)$ on survival time T_i, thus we propose the following PH model with historic patterns of change taken into account:

$$\lambda(t \mid \boldsymbol{Z}_i, \boldsymbol{X}_i^t) = \lambda_0(t) \exp\{\boldsymbol{\beta}_Z^\top \boldsymbol{Z}_i + \boldsymbol{\beta}_X^\top \varphi(\boldsymbol{X}_i^t)\} = \lambda_0(t) \exp\{\boldsymbol{\beta}_Z^\top \boldsymbol{Z}_i + \boldsymbol{\beta}_X^\top \boldsymbol{Z}_i(t)\}, \quad (7)$$

where $\boldsymbol{Z}_i(t) = \varphi(\boldsymbol{X}_i^t) \in \mathbb{R}^r$, $\varphi(\cdot)$ is a modeling function which characterizes the historic change patterns in longitudinal covariates prior to time t, and the rest of the notations represent the same as in (5), but the dimension of parameter $\boldsymbol{\beta}_X \in \mathbb{R}^r$ relies on the choice of modeling function $\varphi(\cdot)$, thus we do not necessarily have $r = p$ here.

Note that our consideration of $\boldsymbol{Z}_i(t) = \varphi(\boldsymbol{X}_i^t)$ in (7) is motivated by the modeling function $\varphi(\cdot)$ in *Smoking Cessation Data* and *Child Development Data* which are described in Section 2.2. Also note that our above proposed PH model (7) includes PH model (5) as a special case when $\boldsymbol{Z}_i(t) = \boldsymbol{X}_i(t)$.

Due to the complexity of model (7) and that our main interest here is $\boldsymbol{Z}_i(t) = \varphi(\boldsymbol{X}_i^t)$, for simplicity of presentation, the rest of this article considers the case that PH model (7) does not contain time-independent covariate $\boldsymbol{Z}_i$, and the longitudinal covariate $\boldsymbol{X}_i(t)$ and $\boldsymbol{Z}_i(t)$ are scalar rather than vector, i.e., $p = r = 1$ in (7):

$$\lambda(t \mid X^t) = \lambda_0(t) \exp\{\beta_0 Z(t)\}, \quad (8)$$

where $Z(t) = \varphi(X^t)$ with $X^t = \{X(s) \mid 0 \le s < t\}$, β_0 is the regression parameter, $\lambda_0(t)$ is an arbitrary baseline hazard function with corresponding d.f. F_0, and $\lambda(t \mid X^t)$ is the conditional hazard function of T given X^t in the following sense:

$$\lambda(t \mid X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le T < t + \Delta \mid T \ge t, X^t\}, \quad (9)$$

which is the conditional probability of failure at time t given that the failure does not occur prior to time t and that the covariate process $X(\cdot)$ has the history of the process up to time t, denoted by X^t.

Following the notations in data (1)-(3), for PH model (8) the observed right censored data on survival time T with longitudinal covariate process $X(t)$ are given by:

$$(V_i, \delta_i, \mathcal{X}_i), \quad i = 1, 2, \cdots, n, \quad (10)$$

where $\mathcal{X}_i = \big(X_i(t_{i1}), \cdots, X_i(t_{iJ_i})\big)$ with (V_i, δ_i) given by (4). Notice that in practice we do not observe anything on $X_i(t)$ for $t > V_i$ in (10), thus the time points observed on $X_i(t)$, i.e., $X_i(t_{ij})$, satisfy $t_{iJ_i} \le V_i$. In turn, the actually observed data in (10) on $(T_i, X_i(t))$'s for PH model (8) are given by

$$(V_i, \delta_i, \mathcal{X}_i^{V_i}), \quad i = 1, 2, \cdots, n, \tag{11}$$

where $\mathcal{X}_i^t = \{X_i(t_{ij}) \mid t_{ij} \le t\}$ represents the vector of all observed $X_i(s)$ satisfying $s \le t$.

In this article, we derive the empirical likelihood-based MLE for (β_0, F_0) in our proposed PH model (8) based on right censored survival data with longitudinal covariates (11). Under model assumption (8), our approach does not make any parametric or semiparametric assumptions or use kernel method on the longitudinal covariate process $X(t)$. It should be noted that the generalization of our results here to PH model (7) with time-independent covariate $\boldsymbol{Z}_i$ and multivariate dimension $\boldsymbol{Z}_i(t)$ for $p > 1$ or $r > 1$ is straightforward, but the computing algorithms are quite complicated and require further studies, which are developed in upcoming paper by Ren and Shi (2023).

3 Maximum likelihood estimator

To derive the empirical likelihood function for (β_0, F_0) in our proposed PH model (8) based on right censored survival data with longitudinal covariates (11), we let

$$\bar{F}(t \mid X^t) = P\{T \ge t \mid X^t\}, \tag{12}$$

then from (9) and Kalbfleisch and Prentice (2002; Page 96), we have that for any fixed $t > 0$,

$$\begin{aligned} f(t \mid X^t) &= \lambda(t \mid X^t)\bar{F}(t \mid X^t) = \lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \mid T \ge t, X^t\}}{\Delta} \bar{F}(t \mid X^t) \\ &= \lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \mid X^t\}}{\Delta P\{T \ge t \mid X^t\}} \bar{F}(t \mid X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le T < t + \Delta \mid X^t\}. \end{aligned} \tag{13}$$

Assuming that the covariate process $X(t)$ is external, that is for any $s < t$, $X(t)$ is not affected by the occurrence of failure at time s, and assuming the uniform convergence in (13), in Appendix we show that for any $t > 0$:

$$\int_0^t f(s \mid X^s)\, ds = P\{0 \le T < t \mid X^t\} \equiv F(t \mid X^t) \quad \Rightarrow \quad \frac{d}{dt}\big(F(t \mid X^t)\big) = f(t \mid X^t). \tag{14}$$

Thus, from (12)-(14) and $\bar{F}(t \mid X^t) + F(t \mid X^t) \equiv 1$, we have

$$\Lambda(t \mid X^t) = \int_0^t \lambda(s \mid X^s)\, ds = \int_0^t \frac{f(s \mid X^s)}{\bar{F}(s \mid X^s)}\, ds = -\ln \bar{F}(t \mid X^t)$$

which gives the version of Lehmann family property under model assumption (8):

$$\bar{F}(t \mid X^t) = \exp\{-\Lambda(t \mid X^t)\} = \exp\Big\{ -\int_0^t \lambda_0(s)\, e^{\beta_0 Z(s)}\, ds \Big\}. \tag{15}$$

Consider the probability density function (p.d.f.) of (V, δ) in (11) given X^t in the following sense: $g(t, \delta \mid X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le V < t + \Delta,\ \delta = \delta \mid X^t\}$, then we have

Lemma 1

$$g(t, 1 \mid X^t) = f(t \mid X^t)\bar{F}_C(t) \quad \textit{and} \quad g(t, 0 \mid X^t) = \bar{F}(t \mid X^t) f_C(t), \tag{16}$$

where $f_C(t)$ and $F_C(t)$ are the p.d.f. and d.f. of right censoring variable C, respectively.

Thus, the likelihood for (V_i, δ_i) in (11) given $X_i^{V_i}, 1 \le i \le n$, is given by

$$\prod_{i=1}^n g(V_i, \delta_i \mid X_i^{V_i}) = \prod_{i=1}^n \big[f(V_i \mid X_i^{V_i})\bar{F}_C(V_i)\big]^{\delta_i} \big[\bar{F}(V_i \mid X_i^{V_i}) f_C(V_i)\big]^{1-\delta_i}$$

which is proportional to

$$L(V, \delta, X^V) = \prod_{i=1}^n \big[f(V_i \mid X_i^{V_i})\big]^{\delta_i} \big[\bar{F}(V_i \mid X_i^{V_i})\big]^{1-\delta_i}. \tag{17}$$

However, above $L(V, \delta, X^V)$ is not the likelihood for observed data (11) because from (8) and (11) we know $X_i^{V_i} \ne \mathcal{X}_i^{V_i}$, and in practice we only observe $X_i(t)$ at time points $t_{i1} < \cdots < t_{iJ_i}$, thus under model assumption (8), term $\bar{F}(V_i \mid X_i^{V_i})$ is not computable by equation (15). A natural way to deal with this is to use a step function $\hat{Z}_i(t)$ as an estimate for $Z_i(t) = \varphi(X_i^t)$:

$$\hat{Z}_i(t) = \tilde{Z}_i(t_{iJ_i}) I\{t \ge t_{iJ_i}\} + \sum_{j=0}^{J_i - 1} \tilde{Z}_i(t_{ij}) I\{t_{ij} \le t < t_{i,j+1}\}, \quad 0 \le t \le V_i \tag{18}$$

where $\tilde{Z}_i(t) = \varphi(\mathcal{X}_i^t)$ is estimate of $Z_i(t) = \varphi(X_i^t)$. Note that we clearly have $\hat{Z}_i(t) \approx Z_i(t)$ if points t_{ij}'s are very dense in interval $(0, V_i)$ and if $\hat{Z}_i(t)$ in (18) is calculated by replacing $\tilde{Z}_i(t)$ with true value $Z_i(t)$. Under model assumption (8), this $\hat{Z}_i(t)$ in (18) gives

Lemma 2 *Let $f_0(t)$ be the p.d.f. of $F_0(t)$. For any $t \in [t_{iJ_i}, V_i]$,*

$$\bar{F}(t \mid \mathcal{X}_i^t) = \exp\Big\{ -\int_0^t \lambda_0(s)\, e^{\beta_0 \hat{Z}_i(s)}\, ds \Big\} = \big[\bar{F}_0(t)\big]^{c_i} \prod_{j=1}^{J_i} \big[\bar{F}_0(t_{ij})\big]^{c_{i,j-1} - c_{ij}} \tag{19}$$

$$f(t \mid \mathcal{X}_i^t) = c_i f_0(t) \big[\bar{F}_0(t)\big]^{c_i - 1} \prod_{j=1}^{J_i} \big[\bar{F}_0(t_{ij})\big]^{c_{i,j-1} - c_{ij}} \tag{20}$$

where $c_{ij} = e^{\beta \tilde{Z}_i(t_{ij})}$ *and* $c_i = c_{iJ_i}$ *with* $\beta = \beta_0$ *and* $c_{i0} = 1$.

Thus, based on the idea of (17) and some algebra work shown in Appendix, the likelihood for observed data (11) under model assumption (8) is proportional to

$$\begin{aligned} &\prod_{i=1}^{n} \left[f(V_i \mid \mathcal{X}_i^{V_i})\right]^{\delta_i} \left[\bar{F}(V_i \mid \mathcal{X}_i^{V_i})\right]^{1-\delta_i} \\ &\quad = \prod_{i=1}^{n} \left[c_i f_0(V_i)\right]^{\delta_i} \left[\bar{F}_0(V_i)\right]^{c_i - \delta_i} \Big(\prod_{j=1}^{J_i} \left[\bar{F}_0(t_{ij})\right]^{c_{i,j-1} - c_{ij}} \Big). \end{aligned} \tag{21}$$

Hence, the likelihood function for (β_0, F_0) in our proposed PH model (8) based on right censored survival data with longitudinal covariates (11) is given by

$$L(\beta, F) = \prod_{i=1}^{n} \left[c_i \, dF(V_i)\right]^{\delta_i} \left[\bar{F}(V_i)\right]^{c_i - \delta_i} \Big(\prod_{j=1}^{J_i} \left[\bar{F}(t_{ij})\right]^{c_{i,j-1} - c_{ij}} \Big), \tag{22}$$

where F is an arbitrary d.f., $dF(t) = F(t) - F(t-)$ and

$$c_{ij} = e^{\beta \tilde{Z}_i(t_{ij})}, \, j = 1, \cdots, J_i; \quad \text{with } \; c_{i0} = 1 \; \text{ and } \; c_i = c_{iJ_i}. \tag{23}$$

The maximum likelihood estimator (MLE) $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) is the solution that maximizes above likelihood function $L(\beta, F)$ given in (22).

To find the MLE $\hat{F}_n$ for F_0, while the general formulas are presented later in Section 5 with proofs given in Appendix, for simplicity of presentation here we assume that there are no ties among V_i's in observed data (11) satisfying:

$$V_1 < V_2 < \cdots < V_n \tag{24}$$

and we restrict all possible candidates to those d.f.'s that assign all their probability masses to points V_i's and interval (V_n, ∞), i.e., d.f. F in $L(\beta, F)$ of (22) is given by

$$F(t) = \sum_{i=1}^{n} p_i I\{V_i \le t\} \tag{25}$$

where $p_i = F(V_i) - F(V_i-)$ for $1 \le i \le n$, and $\sum_{i=1}^{n+1} p_i = 1$ with $p_{n+1} \in [0, 1]$ as the probability mass of F on interval (V_n, ∞). Thus, for $\boldsymbol{p} = (p_1, \cdots, p_n, p_{n+1})$ we show that likelihood function in (22) has the following expression for F given by (25):

Lemma 3 *For* $c_i = e^{\beta \tilde{Z}_i(t_{iJ_i})}$,

$$L(\beta, F) = \prod_{i=1}^{n} \left[c_i \, p_i\right]^{\delta_i} \Big(\sum_{j=i+1}^{n+1} p_j \Big)^{d_i - \delta_i} \equiv L(\beta, \boldsymbol{p}) \tag{26}$$

where $d_i = e^{\beta \tilde{Z}_i(t_{iJ_i})} + \sum_{k=1}^{n} \sum_{j=1}^{J_k} \left(e^{\beta \tilde{Z}_k(t_{k,j-1})} - e^{\beta \tilde{Z}_k(t_{kj})} \right) I\{V_i \le t_{kj} < V_{i+1}\}$.

3.1 Estimating function with restriction

Theorem 1 *For any fixed β satisfying assumption (AS1): $d_i \geq 1$, $i = 1, \cdots, n$, likelihood function $L(\beta, F)$ in (26) is maximized by*:

$$1 - \hat{F}_n(t;\beta) = \prod_{V_i \leq t} \Big(1 - \frac{\delta_i}{e_i}\Big), \quad \text{where} \quad e_i = d_i + \cdots + d_n. \tag{27}$$

Replacing F in (26) by $\hat{F}_n(t;\beta)$, from the proof of Theorem 1 we obtain the profile likelihood function for β_0 in PH model (8) under assumption (AS1):

$$\ell(\beta) = L\big(\beta, \hat{F}_n(\cdot;\beta)\big) = \prod_{i=1}^{n} \Big(\frac{c_i}{e_i}\Big)^{\delta_i} \Big(1 - \frac{\delta_i}{e_i}\Big)^{e_i - \delta_i}. \tag{28}$$

From differentiation, we obtain the *profile estimating function*: $\Psi_n(\beta) = n^{-1} \frac{d}{d\beta} \big(\log \ell(\beta)\big)$, then under assumption (AS1), the MLE $\hat{\beta}_n$ for β_0 in PH model (8) with observed right censored survival data with longitudinal covariates (11) is given by the solution of equation: $\Psi_n(\beta) = 0$, and the MLE $\hat{F}_n$ for F_0 is given by:

$$\hat{F}_n(t) = \hat{F}_n(t;\hat{\beta}_n), \qquad t > 0 \tag{29}$$

where $\hat{F}_n(t;\beta)$ is given by (27), and algebraic simplification gives:

$$\Psi_n(\beta) = n^{-1} \sum_{i=1}^{n} \delta_i \Big(\tilde{Z}_i(t_{iJ_i}) + e_i' \log\Big(1 - \frac{1}{e_i}\Big)\Big), \quad \text{with} \quad e_i' = \frac{de_i}{d\beta}. \tag{30}$$

As mentioned at the end of Section 1, Ren and Zhou (2011) showed that the leading term of Taylor's expansion of the empirical likelihood-based full-profile likelihood estimating function for regression parameter in the Cox model with time-independent covariate is in fact the Cox's partial likelihood estimating function. Thus, the leading term of Taylor's expansion of above profile estimating function $\Psi_n(\beta)$, i.e., using linear approximation $\log(1 - x) \approx -x$ on term $\log\big(1 - \frac{1}{e_i}\big)$ in (30), gives the following version of *partial likelihood estimating function* for β_0 in PH model (8):

$$\psi_n(\beta) = n^{-1} \sum_{i=1}^{n} \delta_i \Big(\tilde{Z}_i(t_{iJ_i}) - \frac{e_i'}{e_i}\Big), \tag{31}$$

which we also call *approximated profile estimating function*. Then, the *partial likelihood estimator* or *approximated MLE* $\tilde{\beta}_n$ for β_0 in PH model (8) based on data (11) is given by the solution of equation: $\psi_n(\beta) = 0$.

Noticing the simplified form of e_i given in equation (32) below, it is easy to see that above $\psi_n(\beta)$ has the same form, though not exactly the same, as the partial likelihood estimating function (4.15) on Page 102 in Kalbfleisch and Prentice

(2002), where without t_{ij}'s, our $\hat{Z}_i(t)$ in equation (18) is just their assumed known $Z_i(t)$.

With proofs given in Appendix, the following lemma gives some properties of estimating functions $\Psi_n(\beta)$ and $\psi_n(\beta)$.

Lemma 4 (a) *For* $\mathcal{E}_{ik} = \{t_{kj} \mid t_{kj} \geq V_i,\ 1 \leq j \leq J_k\}$ *and* $J_{ik} = \max\{j \mid t_{kj} < V_i,\ 1 \leq j \leq J_k\}$,

$$e_i = \sum_{k=i}^{n} \left(e^{\beta \tilde{Z}_k(t_{kJ_k})} I\{\mathcal{E}_{ik} = \emptyset\} + e^{\beta \tilde{Z}_k(t_{kJ_{ik}})} I\{\mathcal{E}_{ik} \neq \emptyset\} \right); \tag{32}$$

(*b*) $\psi_n(\beta)$ *is non-increasing in* $\beta \in \mathbb{R}$.

Remark 1 The results in this article are under the assumption that covariate process $X(t)$ is external, because the proof of (14) requires equation (6.8) on Page 196 in Kalbfleisch and Prentice (2002) to hold. As pointed out by Kalbfleisch and Prentice (2002; Page 198), this equation (6.8) does not hold for internal process $X(t)$. Moreover, it should be noted that for fixed β under assumption (AS1) in Theorem 1, likelihood function $L(\beta, F)$ in (26) has finite maximum value, and $\hat{F}_n(t;\beta)$ in (27) is a proper d.f. because $e_i \geq d_i \geq 1$. Here is an example that (AS1) does not hold: $n = 2, \delta_1 = \delta_2 = 1,\ d_1 \geq 1,\ d_2 < 1$, for which in (26) we have $\sup_{\boldsymbol{p}} L(\beta, \boldsymbol{p}) = \infty$. The next subsection provides a data-based algorithm to compute MLE $\hat{\beta}_n$ without assumption (AS1).

3.2 Generalized estimating function

To remove assumption (AS1) for estimation function $\Psi_n(\beta)$ given in (30), we notice from the proof of Theorem 1 that in fact (AS1) can be replaced by assumption:

$$e_i \geq 1,\ i = 1, \cdots, n \qquad (AS2)$$

which ensures that $L(\beta, \boldsymbol{p})$ in (26) has finite maximum value for fixed β, and that $\hat{F}_n(t;\beta)$ in (27) is a proper d.f. From (32) of Lemma 4, here we use the "shifting" method based on Remarks 1-2 in Ren and Zhou (2011) to achieve (AS2) for β being restricted on an arbitrary compact set.

Consider any constant $M > 0$ and rewrite PH model (8) as follows:

$$\lambda(t \mid X^t) = \lambda_1(t) \exp\{\beta_0 Z(t) + M\}, \tag{33}$$

where $\lambda_1(t) = \lambda_0(t)e^{-M}$ is a hazard function with corresponding d.f. $F_1(t)$ and p.d.f. $f_1(t)$. From the derivation of (15), we have the version of Lehmann family property under the "shifting" PH model (33):

$$\bar{F}(t \mid X^t) = \exp\left\{ -\int_0^t \lambda_1(s)\, e^{\beta_0 Z(s) + M}\, ds \right\}. \tag{34}$$

Following the derivation of (26), we obtain the likelihood function for (β_0, F_1) in PH model (33) with observed data (11) given by:

$$L_M(\beta, F) = \prod_{i=1}^{n} \left[C_i\, p_i\right]^{\delta_i} \Big(\sum_{j=i+1}^{n+1} p_j \Big)^{D_i - \delta_i} \equiv L_M(\beta, \boldsymbol{p}) \tag{35}$$

where $C_i = e^M c_i$, $D_i = e^M d_i$, and F is any d.f. given by (25). Letting $E_i = e^M e_i$, notice that for large enough $M > 0$, (AS1) in Theorem 1 holds for D_i's and above assumption (AS2) holds for E_i's with any β in the neighborhood of β_0, thus from Theorem 1 we know that above likelihood function $L_M(\beta, F)$ is maximized for fixed β by

$$1 - \hat{G}_n(t; \beta) = \prod_{V_i \le t} \Big(1 - \frac{\delta_i}{E_i}\Big), \quad \text{where} \quad E_i = e^M e_i. \tag{36}$$

Hence, the derivation of (28) and (30) gives the *generalized profile estimating function*:

$$\Psi_{M,n}(\beta) = n^{-1} \sum_{i=1}^{n} \delta_i \Big(\tilde{Z}_i(t_{iJ_i}) + E_i' \log\big(1 - \frac{1}{E_i}\big)\Big), \quad \text{with} \quad E_i' = \frac{dE_i}{d\beta} \tag{37}$$

and the MLE $\hat{\beta}_n$ for β_0 is given by the solution of equation: $\Psi_{M,n}(\beta) = 0$, with the MLE $\hat{F}_n$ for F_0 given by $\hat{\bar{F}}_n(t) = \left[\hat{\bar{G}}_n(t; \hat{\beta}_n)\right]^{e^M}$.

In practice, we use estimating function $\Psi_n(\beta)$ in (30) to find the MLE $\hat{\beta}_n$ whenever possible. If constant M is needed for the use of generalized estimating function $\Psi_{M,n}(\beta)$, we choose the one with smallest possible M based on the data; see Section 4 for a brief description and see Appendix for more detailed explanation on data-based choice of M.

It should be noted that using Taylor's expansion $\log(1 - x) \approx -x$ on term $\log\left(1 - \frac{1}{E_i}\right)$ in above (37), the resulting approximation of $\Psi_{M,n}(\beta)$ is the same as $\psi_n(\beta)$ given by (31) for any constant M. Since $\psi_n(\beta)$ is well defined for any β and since $\psi_n(\beta)$ is monotone based on Lemma 4, the computation of partial likelihood estimator or approximated MLE $\tilde{\beta}_n$ is much simpler than the MLE $\hat{\beta}_n$, although $\hat{\beta}_n$ often has better performance.

4 Simulation studies

This section presents some simulation results on the partial likelihood estimator or approximated MLE $\tilde{\beta}_n$ given by $\psi_n(\beta) = 0$ in (31), and the MLE $\hat{\beta}_n$ given by $\Psi_n(\beta) = 0$ in (30) or $\Psi_{M,n}(\beta) = 0$ in (37) with a data-based choice of M. We consider two cases of modeling function $\varphi(\cdot)$ in our proposed PH model (8) which are motivated by the *Smoking Cessation Data* and *Child Development Data* described in Section 2.2. Routines in C++ for computation here are available from the authors.

Data-based choice of M

The main idea of data-based choice of M is as follows. Since the computation of $\tilde{\beta}_n$ is simple and does not depend on M, and since asymptotically $\tilde{\beta}_n$ and $\hat{\beta}_n$ are very close, we first compute $\tilde{\beta}_n$, then search for $\hat{\beta}_n$ in the neighborhood of $\tilde{\beta}_n$: $(\tilde{\beta}_n - \rho,\ \tilde{\beta}_n + \rho)$ with small $\rho > 0$. If all $e_i \geq 1$ in such neighborhood, then no need to use M and compute the MLE $\hat{\beta}_n$ by $\Psi_n(\beta) = 0$; If not all $e_i \geq 1$ in the neighborhood, choose the smallest $M > 0$ such that all $E_i \geq 1$ in the neighborhood, then compute $\hat{\beta}_n$ by $\Psi_{M,n}(\beta) = 0$.

More detailed description on data-based choice of M can be found in Appendix. But it should be noted that such procedure is mainly for the sake of Monte Carlo simulation studies, because it is not possible to do sample by sample checking for 1000 times. For the analysis of real data, our experiences show that it is quite easy to determine how to compute $\hat{\beta}_n$ because we are only dealing with one data sample at a time.

4.1 Case 1 simulation studies

Motivated by *Smoking Cessation Data*, here we consider the average of longitudinal covariate process $X(t)$ in time interval $(0,\ t)$ as the modeling function in PH model (8):

$$\begin{cases} Z(t) = \varphi(X^t) = \frac{1}{t}\int_0^t X(s)ds \\ \tilde{Z}_i(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}) = \frac{\sum_{k=1}^{j}(t_{ik}-t_{i,k-1})X_i(t_{ik})}{t_{ij}} \end{cases}. \tag{38}$$

Let $\text{Exp}(\mu)$ represent the exponential distribution with mean μ and $U(a,\ b)$ represent the uniform distribution on interval $(a,\ b)$. For each case of $\beta_0 = 0, \pm 1$ under PH model (8), we generate 1000 samples (11) with sample size $n = 100, 200, 500, 1000$, respectively; see detailed description given below. For $\hat{Z}_i(t)$ in Eq. (18) calculated by using estimated value $\tilde{Z}_i(t)$ as given and by replacing $\tilde{Z}_i(t)$ with the true value $Z_i(t)$, respectively, Tables 1, 2, 3, 4, 5 6 include the simulation average of MLE $\hat{\beta}_n$ and partial likelihood estimator $\tilde{\beta}_n$ with the simulation standard deviation (s.d.) given in the parenthesis next to the average, respectively. The censoring percentages are reported in the tables.

We consider $F_0 = \text{Exp}(1)$ with $\lambda_0(t) = 1, F_C = \text{Exp}(3)$, and $X(t) = X(t;\omega) = \omega\, t$ with ω as a random variable (r.v.) from $U(0,\ 1)$, which by the 1st equation in above (38) gives $Z(t) = \frac{\omega t}{2}$, and by equation (15) gives

$$G(t;\omega) = F(t \mid X^t) = \begin{cases} 1 - \exp\left\{\frac{2(1-e^{\beta_0 \omega t/2})}{\beta_0 \omega}\right\} & \text{if } \beta_0 \neq 0 \\ 1 - e^{-t} & \text{if } \beta_0 = 0 \end{cases}. \tag{39}$$

Thus, generating ω_i from $U(0,\ 1)$, then generating T_i from above proper/improper d.f. $G(t;\omega_i)$ and C_i from F_C give (V_i, δ_i) in (11) with longitudinal covariate $X_i(t) = \omega_i\, t$ that satisfies PH model assumption (8). To generate time points t_{ij}'s, we consider m time points for each unit of V_i; that is we let $J_i = m([V_i] + 1)$ be the total number of time points for ith observation in (11), which gives $\mathcal{X}_i^{V_i} = \{X_i(t_{ij}) = \omega_i t_{ij} \mid j = 1, \cdots, J_i\}$,

Table 1 Case 1: $\boldsymbol{\beta_0 = 1}$, t_{ij}'s generated from $U(\mathbf{0}, V_i)$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.500 (1.24)	1.822 (1.27)	2.223 (1.18)	2.484 (1.13)	1.494 (0.85)	1.707 (0.83)	2.113 (0.76)	2.263 (0.74)
$m = 10$	1.218 (1.22)	1.531 (1.26)	1.525 (1.13)	1.819 (1.14)	1.218 (0.81)	1.441 (0.80)	1.496 (0.76)	1.683 (0.74)
$m = 20$	1.076 (1.08)	1.373 (1.14)	1.188 (1.03)	1.475 (1.08)	0.957 (0.76)	1.173 (0.77)	1.059 (0.73)	1.266 (0.73)
$m = 30$	0.902 (1.10)	1.189 (1.16)	0.961 (1.05)	1.246 (1.11)	0.987 (0.77)	1.210 (0.78)	1.039 (0.75)	1.254 (0.75)
$m = 40$	0.936 (1.11)	1.230 (1.17)	0.979 (1.07)	1.260 (1.13)	0.919 (0.77)	1.134 (0.77)	0.950 (0.75)	1.160 (0.75)
$m = 50$	0.902 (1.13)	1.188 (1.19)	0.927 (1.09)	1.209 (1.16)	0.935 (0.73)	1.156 (0.73)	0.959 (0.71)	1.171 (0.71)
Censoring %	22.4%				22.3%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.489 (0.52)	1.620 (0.52)	2.023 (0.46)	2.123 (0.46)	1.496 (0.34)	1.603 (0.33)	2.019 (0.30)	2.084 (0.29)
$m = 10$	1.236 (0.46)	1.368 (0.45)	1.466 (0.42)	1.582 (0.41)	1.246 (0.33)	1.353 (0.33)	1.472 (0.31)	1.557 (0.30)
$m = 20$	1.050 (0.45)	1.182 (0.44)	1.134 (0.43)	1.260 (0.42)	1.087 (0.31)	1.195 (0.30)	1.169 (0.30)	1.266 (0.29)
$m = 30$	1.003 (0.44)	1.138 (0.43)	1.048 (0.43)	1.176 (0.42)	1.028 (0.30)	1.135 (0.30)	1.072 (0.29)	1.172 (0.29)
$m = 40$	0.979 (0.45)	1.111 (0.45)	1.003 (0.44)	1.132 (0.43)	1.005 (0.31)	1.114 (0.31)	1.030 (0.30)	1.133 (0.30)
$m = 50$	0.956 (0.46)	1.089 (0.45)	0.974 (0.44)	1.103 (0.44)	0.974 (0.31)	1.081 (0.30)	0.991 (0.30)	1.096 (0.30)
Censoring %	22.3%				22.3%			

Table 2 Case 1: $\boldsymbol{\beta_0 = 1}$, $\boldsymbol{t_{ij}}$'s evenly chosen in $(\mathbf{0}, \boldsymbol{V_i})$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	2.802 (1.60)	3.065 (1.57)	3.264 (1.74)	3.533 (1.69)	2.557 (1.02)	2.727 (0.99)	2.941 (1.08)	3.102 (1.06)
$m = 10$	1.691 (1.27)	1.989 (1.29)	1.792 (1.29)	2.098 (1.30)	1.620 (0.85)	1.822 (0.83)	1.712 (0.86)	1.911 (0.84)
$m = 20$	1.218 (1.16)	1.525 (1.21)	1.244 (1.16)	1.553 (1.20)	1.213 (0.78)	1.433 (0.77)	1.233 (0.78)	1.455 (0.77)
$m = 30$	1.078 (1.12)	1.382 (1.18)	1.091 (1.12)	1.396 (1.17)	1.095 (0.76)	1.305 (0.76)	1.102 (0.76)	1.313 (0.76)
$m = 40$	1.008 (1.10)	1.310 (1.16)	1.017 (1.10)	1.321 (1.16)	1.033 (0.74)	1.246 (0.75)	1.039 (0.74)	1.253 (0.75)
$m = 50$	0.970 (1.10)	1.264 (1.16)	0.975 (1.09)	1.271 (1.16)	0.996 (0.74)	1.208 (0.75)	1.000 (0.74)	1.212 (0.75)
Censoring %	22.4%				22.3%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	2.351 (0.58)	2.469 (0.58)	2.672 (0.61)	2.786 (0.61)	2.286 (0.40)	2.384 (0.41)	2.586 (0.42)	2.680 (0.43)
$m = 10$	1.572 (0.50)	1.697 (0.49)	1.643 (0.51)	1.767 (0.50)	1.557 (0.35)	1.658 (0.35)	1.622 (0.35)	1.722 (0.35)
$m = 20$	1.220 (0.47)	1.348 (0.46)	1.238 (0.47)	1.365 (0.46)	1.223 (0.33)	1.326 (0.32)	1.241 (0.33)	1.343 (0.33)
$m = 30$	1.110 (0.46)	1.240 (0.45)	1.117 (0.46)	1.246 (0.45)	1.119 (0.32)	1.223 (0.32)	1.127 (0.32)	1.231 (0.32)
$m = 40$	1.053 (0.45)	1.185 (0.45)	1.059 (0.45)	1.190 (0.45)	1.066 (0.32)	1.170 (0.32)	1.070 (0.32)	1.173 (0.32)
$m = 50$	1.021 (0.45)	1.150 (0.45)	1.024 (0.45)	1.154 (0.45)	1.035 (0.32)	1.141 (0.32)	1.037 (0.32)	1.143 (0.32)
Censoring %	22.3%				22.3%			

Table 3 Case 1: $\boldsymbol{\beta_0 = 0}$, $\boldsymbol{t_{ij}}$'s generated from $\boldsymbol{U(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	0.407 (1.06)	0.614 (1.14)	1.129 (0.98)	1.333 (1.01)	0.349 (0.67)	0.515 (0.71)	0.968 (0.63)	1.131 (0.63)
$m = 10$	0.154 (0.99)	0.335 (1.06)	0.471 (0.87)	0.662 (0.96)	0.125 (0.66)	0.268 (0.69)	0.404 (0.59)	0.554 (0.63)
$m = 20$	−0.039 (0.97)	0.134 (1.05)	0.100 (0.90)	0.269 (0.98)	0.058 (0.63)	0.195 (0.67)	0.169 (0.60)	0.303 (0.64)
$m = 30$	−0.030 (0.97)	0.140 (1.03)	0.050 (0.92)	0.216 (0.99)	−0.000 (0.63)	0.138 (0.67)	0.066 (0.60)	0.199 (0.64)
$m = 40$	−0.064 (0.98)	0.107 (1.05)	−0.009 (0.94)	0.156 (1.01)	−0.050 (0.62)	0.082 (0.65)	−0.003 (0.60)	0.125 (0.63)
$m = 50$	−0.108 (1.01)	0.050 (1.07)	−0.062 (0.97)	0.091 (1.04)	−0.092 (0.62)	0.039 (0.65)	−0.057 (0.60)	0.072 (0.64)
Censoring %	25.2%				25.0%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	0.336 (0.39)	0.456 (0.41)	0.895 (0.37)	1.006 (0.36)	0.316 (0.37)	0.415 (0.28)	0.872 (0.26)	0.953 (0.25)
$m = 10$	0.152 (0.36)	0.257 (0.39)	0.395 (0.34)	0.504 (0.36)	0.158 (0.25)	0.239 (0.27)	0.379 (0.24)	0.475 (0.25)
$m = 20$	0.020 (0.37)	0.117 (0.39)	0.122 (0.35)	0.215 (0.37)	0.044 (0.24)	0.114 (0.26)	0.134 (0.22)	0.207 (0.25)
$m = 30$	−0.002 (0.35)	0.094 (0.37)	0.060 (0.34)	0.148 (0.36)	0.030 (0.24)	0.100 (0.26)	0.081 (0.23)	0.152 (0.25)
$m = 40$	−0.012 (0.37)	0.080 (0.39)	0.032 (0.35)	0.120 (0.37)	−0.022 (0.24)	0.045 (0.26)	0.016 (0.23)	0.081 (0.25)
$m = 50$	−0.042 (0.38)	0.048 (0.40)	−0.011 (0.37)	0.074 (0.39)	−0.018 (0.23)	0.048 (0.25)	0.009 (0.23)	0.074 (0.25)
Censoring %	25.1%				25.0%			

Table 4 Case 1: $\boldsymbol{\beta_0 = 0}$, $\boldsymbol{t_{ij}}$'s evenly chosen in $(\mathbf{0}, \boldsymbol{V_i})$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.222 (1.22)	1.437 (1.26)	1.576 (1.32)	1.807 (1.34)	1.014 (0.79)	1.183 (0.80)	1.307 (0.84)	1.480 (0.84)
$m = 10$	0.485 (1.02)	0.676 (1.09)	0.566 (1.02)	0.770 (1.10)	0.403 (0.67)	0.561 (0.70)	0.470 (0.67)	0.631 (0.71)
$m = 20$	0.166 (0.97)	0.341 (1.05)	0.190 (0.97)	0.366 (1.05)	0.137 (0.63)	0.283 (0.67)	0.157 (0.63)	0.302 (0.67)
$m = 30$	0.063 (0.96)	0.239 (1.04)	0.079 (0.96)	0.251 (1.04)	0.056 (0.63)	0.193 (0.67)	0.068 (0.63)	0.204 (0.67)
$m = 40$	0.013 (0.96)	0.184 (1.03)	0.023 (0.96)	0.191 (1.03)	0.017 (0.63)	0.151 (0.66)	0.023 (0.62)	0.156 (0.66)
$m = 50$	−0.015 (0.96)	0.154 (1.02)	−0.005 (0.95)	0.159 (1.03)	−0.007 (0.63)	0.126 (0.66)	−0.001 (0.62)	0.130 (0.66)
Censoring %	25.2%				25.0%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	0.847 (0.45)	0.999 (0.44)	1.126 (0.48)	1.248 (0.47)	0.839 (0.32)	0.944 (0.32)	1.077 (0.33)	1.176 (0.33)
$m = 10$	0.354 (0.38)	0.466 (0.40)	0.411 (0.38)	0.527 (0.40)	0.350 (0.27)	0.445 (0.28)	0.401 (0.27)	0.501 (0.28)
$m = 20$	0.129 (0.36)	0.229 (0.39)	0.145 (0.36)	0.244 (0.38)	0.143 (0.25)	0.224 (0.27)	0.156 (0.25)	0.234 (0.27)
$m = 30$	0.056 (0.36)	0.152 (0.38)	0.065 (0.36)	0.159 (0.38)	0.077 (0.25)	0.148 (0.27)	0.084 (0.25)	0.155 (0.27)
$m = 40$	0.021 (0.36)	0.114 (0.38)	0.028 (0.36)	0.119 (0.38)	0.044 (0.25)	0.114 (0.27)	0.049 (0.25)	0.117 (0.27)
$m = 50$	0.002 (0.36)	0.095 (0.38)	0.007 (0.36)	0.097 (0.38)	0.024 (0.25)	0.093 (0.27)	0.029 (0.25)	0.096 (0.27)
Censoring %	25.1%				25.0%			

Table 5 Case 1: $\boldsymbol{\beta_0 = -1}$, t_{ij}'s generated from $\boldsymbol{U(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.647 (0.96)	−0.559 (1.02)	0.092 (0.84)	0.212 (0.89)	−0.649 (0.59)	−0.594 (0.62)	−0.012 (0.50)	0.071 (0.53)
$m = 10$	−0.930 (0.97)	−0.848 (1.02)	−0.576 (0.86)	−0.482 (0.92)	−0.883 (0.59)	−0.844 (0.61)	−0.591 (0.52)	−0.536 (0.54)
$m = 20$	−0.988 (0.94)	−0.919 (0.98)	−0.854 (0.88)	−0.771 (0.92)	−0.989 (0.60)	−0.953 (0.62)	−0.873 (0.56)	−0.824 (0.58)
$m = 30$	−1.031 (0.95)	−0.962 (0.99)	−0.957 (0.92)	−0.876 (0.95)	−1.040 (0.59)	−1.005 (0.61)	−0.980 (0.57)	−0.930 (0.59)
$m = 40$	−1.085 (0.98)	−1.014 (1.02)	−1.038 (0.96)	−0.955 (0.99)	−1.030 (0.61)	−0.995 (0.63)	−0.993 (0.60)	−0.944 (0.61)
$m = 50$	−1.065 (0.96)	−0.996 (0.99)	−1.032 (0.94)	−0.953 (0.97)	−1.053 (0.59)	−1.017 (0.61)	−1.028 (0.59)	−0.980 (0.60)
Censoring %	30.1%				30.0%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.659 (0.37)	−0.648 (0.38)	−0.075 (0.30)	−0.027 (0.32)	−0.667 (0.25)	−0.666 (0.25)	−0.091 (0.20)	−0.062 (0.21)
$m = 10$	−0.856 (0.37)	−0.846 (0.38)	−0.587 (0.33)	−0.562 (0.34)	−0.836 (0.25)	−0.834 (0.25)	−0.578 (0.22)	−0.562 (0.23)
$m = 20$	−0.915 (0.36)	−0.906 (0.36)	−0.815 (0.34)	−0.791 (0.34)	−0.917 (0.24)	−0.917 (0.24)	−0.821 (0.23)	−0.807 (0.23)
$m = 30$	−0.963 (0.35)	−0.956 (0.35)	−0.911 (0.33)	−0.889 (0.34)	−0.946 (0.24)	−0.948 (0.24)	−0.897 (0.23)	−0.879 (0.23)
$m = 40$	−0.998 (0.35)	−0.990 (0.36)	−0.967 (0.34)	−0.946 (0.35)	−0.965 (0.23)	−0.967 (0.24)	−0.937 (0.23)	−0.922 (0.23)
$m = 50$	−0.963 (0.35)	−0.957 (0.36)	−0.946 (0.35)	−0.923 (0.35)	−0.954 (0.24)	−0.953 (0.24)	−0.936 (0.24)	−0.919 (0.24)
Censoring %	30.1%				30.0%			

Table 6 Case 1: $\boldsymbol{\beta_0 = -1}$, t_{ij}'s evenly chosen in $(\mathbf{0}, V_i)$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.099 (0.99)	0.031 (1.05)	0.179 (1.04)	0.322 (1.10)	−0.208 (0.61)	−0.138 (0.66)	0.015 (0.63)	0.118 (0.68)
$m = 10$	−0.640 (0.92)	−0.558 (0.97)	−0.580 (0.92)	−0.479 (0.97)	−0.653 (0.59)	−0.608 (0.63)	−0.607 (0.59)	−0.542 (0.62)
$m = 20$	−0.889 (0.91)	−0.817 (0.96)	−0.882 (0.92)	−0.795 (0.96)	−0.860 (0.60)	−0.823 (0.63)	−0.864 (0.61)	−0.808 (0.63)
$m = 30$	−0.969 (0.92)	−0.899 (0.96)	−0.974 (0.92)	−0.890 (0.96)	−0.927 (0.60)	−0.889 (0.62)	−0.937 (0.61)	−0.883 (0.63)
$m = 40$	−1.008 (0.92)	−0.940 (0.96)	−1.017 (0.93)	−0.934 (0.96)	−0.961 (0.60)	−0.926 (0.62)	−0.974 (0.61)	−0.921 (0.63)
$m = 50$	−1.033 (0.92)	−0.968 (0.96)	−1.043 (0.93)	−0.963 (0.96)	−0.981 (0.61)	−0.946 (0.63)	−0.996 (0.61)	−0.943 (0.63)
Censoring %	30.1%				30.0%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.272 (0.33)	−0.244 (0.35)	−0.087 (0.33)	−0.025 (0.36)	−0.279 (0.24)	−0.273 (0.25)	−0.098 (0.23)	−0.059 (0.25)
$m = 10$	−0.662 (0.33)	−0.651 (0.34)	−0.629 (0.33)	−0.599 (0.34)	−0.659 (0.24)	−0.658 (0.24)	−0.627 (0.24)	−0.607 (0.24)
$m = 20$	−0.845 (0.34)	−0.839 (0.34)	−0.850 (0.34)	−0.825 (0.34)	−0.836 (0.24)	−0.836 (0.24)	−0.841 (0.24)	−0.824 (0.24)
$m = 30$	−0.906 (0.34)	−0.898 (0.35)	−0.918 (0.34)	−0.893 (0.35)	−0.893 (0.24)	−0.895 (0.24)	−0.906 (0.24)	−0.889 (0.24)
$m = 40$	−0.935 (0.34)	−0.928 (0.35)	−0.949 (0.34)	−0.925 (0.35)	−0.921 (0.24)	−0.922 (0.24)	−0.935 (0.24)	−0.919 (0.24)
$m = 50$	−0.954 (0.34)	−0.946 (0.35)	−0.968 (0.34)	−0.944 (0.35)	−0.938 (0.24)	−0.938 (0.25)	−0.952 (0.24)	−0.936 (0.25)
Censoring %	30.1%				30.0%			

then $\hat{Z}_i(t)$ in (18) calculated by using estimated value $\tilde{Z}_i(t)$ as given by the 2nd equation in above (38) and calculated by replacing $\tilde{Z}_i(t)$ with the true value $Z_i(t) = \frac{\omega_i t}{2}$, respectively, give the estimating functions $\Psi_n(\beta)$ in (30), $\psi_n(\beta)$ in (31) and $\Psi_{M,n}(\beta)$ in (37).

We use the following two different methods to generate time points t_{ij}'s:

$$\begin{cases} \text{Method 1:} & \text{Generate } t_{ij}\text{'s from } U(0, V_i), \text{ then arrange } t_{i1} < \cdots < t_{iJ_i} \\ \text{Method 2:} & \text{Evenly choose } t_{ij} = \frac{j}{J_i} V_i, 1 \le j \le J_i, \text{ in interval } (0, V_i) \end{cases}, \tag{40}$$

and in Tables 1, 2, 3, 4, 5 6 we consider cases with $m = 5, 10, 20, 30, 40, 50$, respectively, which indicate the intensity of intensive longitudinal covariate in observed data (11).

4.2 Case 2 simulation studies

Motivated by *Child Development Data*, here we consider the total of longitudinal covariate process $X(t)$ in time interval $(0, t)$ as the modeling function in PH model (8):

$$\begin{cases} Z(t) = \varphi(X^t) = \int_0^t X(s)ds \\ \tilde{Z}_i(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}) = \sum_{k=1}^{j} (t_{ik} - t_{i,k-1}) X_i(t_{ik}) \end{cases}. \tag{41}$$

For each case of $\beta_0 = 0, \pm 1$ under PH model (8), we generate 1000 samples (11) with sample size $n = 100, 200, 500, 1000$, respectively, for $F_0(t) = 1 - e^{-t^2/2}$ with $\lambda_0(t) = t$, $F_C = \text{Exp}(3)$, and $X(t) = X(t;\omega) = \frac{\omega t}{2}$ with ω as a r.v. from $U(0, 1)$, which by the 1st equation in (41) gives $Z(t) = \frac{\omega t^2}{4}$, and by equation (15) gives

$$G(t;\omega) = F(t \mid X^t) = \begin{cases} 1 - \exp\left\{ \dfrac{1 - e^{\beta_0 \omega t^2/4}}{\beta_0 \omega/2} \right\} & \text{if } \beta_0 \neq 0 \\ 1 - e^{-t^2/2} & \text{if } \beta_0 = 0 \end{cases}. \tag{42}$$

The observed sample $(V_i, \delta_i, \mathcal{X}_i^{V_i})$ in (11), the time points t_{ij}'s and $\tilde{Z}_i(t_{ij})$'s are generated following the description for Case 1 simulation studies via Eqs. (40)–(42) with $m = 5, 10, 20, 30, 40, 50$, respectively. For $\hat{Z}_i(t)$ in (18) calculated by using estimated value $\tilde{Z}_i(t)$ as given by the 2nd equation in above (41) and calculated by replacing $\tilde{Z}_i(t)$ with the true value $Z_i(t) = \frac{\omega_i t^2}{4}$, respectively, Tables 7, 8, 9, 10, 11, 12 include censoring percentages and the simulation average of $\hat{\beta}_n$ and $\tilde{\beta}_n$ with the simulation s.d. given in the parenthesis.

Table 7 Case 2: $\boldsymbol{\beta_0 = 1}$, $\boldsymbol{t_{ij}}$'s generated from $\boldsymbol{U(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.164 (1.31)	1.509 (1.38)	0.848 (1.05)	1.135 (1.13)	1.044 (0.86)	1.289 (0.87)	0.759 (0.70)	0.974 (0.72)
$m = 10$	0.935 (1.28)	1.268 (1.34)	0.762 (1.13)	1.064 (1.21)	0.973 (0.84)	1.212 (0.86)	0.806 (0.75)	1.028 (0.78)
$m = 20$	0.868 (1.25)	1.191 (1.33)	0.778 (1.17)	1.078 (1.25)	0.894 (0.79)	1.132 (0.82)	0.807 (0.74)	1.036 (0.77)
$m = 30$	0.829 (1.18)	1.140 (1.26)	0.775 (1.13)	1.067 (1.21)	0.873 (0.77)	1.105 (0.79)	0.816 (0.74)	1.044 (0.76)
$m = 40$	0.791 (1.08)	1.104 (1.17)	0.750 (1.05)	1.049 (1.14)	0.860 (0.74)	1.093 (0.76)	0.820 (0.72)	1.042 (0.74)
$m = 50$	0.827 (1.17)	1.131 (1.25)	0.796 (1.14)	1.083 (1.22)	0.878 (0.79)	1.112 (0.81)	0.843 (0.77)	1.073 (0.79)
Censoring %	31.2%				31.1%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.165 (0.52)	1.312 (0.51)	0.873 (0.43)	1.007 (0.42)	1.112 (0.35)	1.231 (0.35)	0.840 (0.29)	0.947 (0.29)
$m = 10$	1.052 (0.49)	1.202 (0.48)	0.889 (0.44)	1.029 (0.43)	1.044 (0.33)	1.161 (0.32)	0.887 (0.29)	0.999 (0.29)
$m = 20$	0.951 (0.47)	1.096 (0.45)	0.871 (0.44)	1.012 (0.43)	0.993 (0.32)	1.109 (0.32)	0.909 (0.30)	1.022 (0.30)
$m = 30$	0.952 (0.47)	1.095 (0.46)	0.890 (0.45)	1.031 (0.44)	0.940 (0.32)	1.054 (0.32)	0.885 (0.31)	0.999 (0.30)
$m = 40$	0.915 (0.48)	1.058 (0.47)	0.871 (0.47)	1.011 (0.46)	0.930 (0.31)	1.045 (0.30)	0.885 (0.30)	0.999 (0.29)
$m = 50$	0.917 (0.45)	1.060 (0.44)	0.882 (0.44)	1.020 (0.43)	0.939 (0.31)	1.054 (0.31)	0.904 (0.30)	1.015 (0.30)
Censoring %	31.1%				31.0%			

Table 8 Case 2: $\boldsymbol{\beta_0 = 1}$, $\boldsymbol{t_{ij}}$'s evenly chosen in $\boldsymbol{(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	3.424 (1.91)	3.736 (1.86)	2.823 (1.62)	3.127 (1.57)	3.009 (1.17)	3.197 (1.14)	2.531 (1.01)	2.710 (0.98)
$m = 10$	1.882 (1.40)	2.232 (1.42)	1.696 (1.30)	2.034 (1.32)	1.768 (0.92)	1.996 (0.90)	1.607 (0.86)	1.833 (0.84)
$m = 20$	1.284 (1.24)	1.619 (1.29)	1.222 (1.19)	1.544 (1.25)	1.261 (0.82)	1.503 (0.82)	1.198 (0.79)	1.435 (0.79)
$m = 30$	1.120 (1.19)	1.444 (1.25)	1.079 (1.17)	1.396 (1.23)	1.109 (0.80)	1.349 (0.80)	1.073 (0.77)	1.311 (0.78)
$m = 40$	1.031 (1.17)	1.345 (1.25)	1.006 (1.15)	1.311 (1.22)	1.039 (0.78)	1.275 (0.79)	1.014 (0.76)	1.244 (0.78)
$m = 50$	0.987 (1.16)	1.292 (1.24)	0.973 (1.14)	1.271 (1.22)	0.999 (0.77)	1.230 (0.78)	0.980 (0.76)	1.208 (0.77)
Censoring %	31.2%				31.1%			
	31.2%				31.1%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	2.700 (0.65)	2.821 (0.65)	2.300 (0.56)	2.418 (0.56)	2.589 (0.45)	2.693 (0.45)	2.227 (0.39)	2.320 (0.39)
$m = 10$	1.694 (0.53)	1.823 (0.52)	1.553 (0.50)	1.681 (0.49)	1.661 (0.37)	1.768 (0.37)	1.533 (0.35)	1.635 (0.35)
$m = 20$	1.268 (0.49)	1.405 (0.47)	1.207 (0.47)	1.342 (0.46)	1.266 (0.34)	1.377 (0.34)	1.211 (0.33)	1.319 (0.33)
$m = 30$	1.137 (0.48)	1.273 (0.47)	1.100 (0.46)	1.235 (0.45)	1.145 (0.33)	1.255 (0.33)	1.110 (0.33)	1.220 (0.32)
$m = 40$	1.070 (0.47)	1.209 (0.46)	1.044 (0.46)	1.180 (0.46)	1.080 (0.33)	1.191 (0.33)	1.059 (0.32)	1.170 (0.32)
$m = 50$	1.029 (0.47)	1.169 (0.46)	1.009 (0.46)	1.147 (0.45)	1.048 (0.33)	1.160 (0.32)	1.028 (0.32)	1.139 (0.32)
Censoring %	31.1%				31.0%			

Table 9 Case 2: $\boldsymbol{\beta_0 = 0}$, $\boldsymbol{t_{ij}}$'s generated from $\boldsymbol{U(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.026 (1.15)	0.162 (1.23)	−0.111 (0.96)	0.037 (1.03)	0.043 (0.65)	0.198 (0.69)	−0.041 (0.55)	0.083 (0.59)
$m = 10$	−0.060 (1.09)	0.125 (1.17)	−0.112 (0.99)	0.046 (1.07)	−0.063 (0.67)	0.083 (0.71)	−0.106 (0.61)	0.019 (0.65)
$m = 20$	−0.104 (1.09)	0.074 (1.17)	−0.128 (1.03)	0.032 (1.12)	−0.100 (0.68)	0.036 (0.72)	−0.123 (0.65)	0.003 (0.69)
$m = 30$	−0.094 (1.05)	0.080 (1.13)	−0.105 (1.01)	0.053 (1.09)	−0.116 (0.64)	0.020 (0.68)	−0.127 (0.61)	−0.004 (0.66)
$m = 40$	−0.164 (1.03)	−0.002 (1.10)	−0.171 (1.00)	−0.021 (1.07)	−0.159 (0.65)	−0.027 (0.68)	−0.166 (0.63)	−0.044 (0.67)
$m = 50$	−0.188 (1.06)	−0.022 (1.14)	−0.192 (1.03)	−0.036 (1.12)	−0.106 (0.62)	0.032 (0.66)	−0.109 (0.61)	0.018 (0.65)
Censoring %	32.8%				32.7%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	0.034 (0.38)	0.144 (0.41)	−0.030 (0.33)	0.052 (0.35)	0.043 (0.27)	0.125 (0.28)	−0.026 (0.23)	0.039 (0.25)
$m = 10$	−0.021 (0.37)	0.080 (0.39)	−0.058 (0.34)	0.024 (0.36)	−0.014 (0.25)	0.060 (0.27)	−0.056 (0.23)	0.007 (0.25)
$m = 20$	−0.059 (0.35)	0.035 (0.37)	−0.079 (0.33)	0.005 (0.36)	−0.038 (0.24)	0.030 (0.26)	−0.060 (0.23)	0.003 (0.25)
$m = 30$	−0.085 (0.37)	0.009 (0.39)	−0.095 (0.36)	−0.012 (0.38)	−0.045 (0.23)	0.022 (0.25)	−0.059 (0.23)	0.002 (0.24)
$m = 40$	−0.089 (0.36)	0.002 (0.38)	−0.097 (0.35)	−0.013 (0.37)	−0.057 (0.24)	0.011 (0.25)	−0.068 (0.24)	−0.003 (0.25)
$m = 50$	−0.067 (0.35)	0.025 (0.37)	−0.072 (0.34)	0.012 (0.36)	−0.055 (0.24)	0.011 (0.26)	−0.063 (0.24)	0.000 (0.25)
Censoring %	32.7%				32.6%			

Table 10 Case 2: $\boldsymbol{\beta_0 = 0}$, $\boldsymbol{t_{ij}}$'s evenly chosen in $\boldsymbol{(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.647 (1.40)	1.902 (1.44)	1.330 (1.19)	1.561 (1.24)	1.266 (0.86)	1.465 (0.87)	1.031 (0.74)	1.216 (0.76)
$m = 10$	0.630 (1.06)	0.843 (1.15)	0.545 (0.99)	0.747 (1.08)	0.490 (0.67)	0.659 (0.71)	0.429 (0.63)	0.584 (0.68)
$m = 20$	0.230 (1.00)	0.411 (1.08)	0.204 (0.96)	0.374 (1.05)	0.169 (0.62)	0.320 (0.67)	0.151 (0.61)	0.294 (0.66)
$m = 30$	0.104 (0.99)	0.278 (1.07)	0.093 (0.96)	0.259 (1.05)	0.072 (0.62)	0.220 (0.66)	0.061 (0.60)	0.202 (0.65)
$m = 40$	0.038 (0.98)	0.212 (1.06)	0.034 (0.96)	0.196 (1.05)	0.023 (0.61)	0.165 (0.66)	0.019 (0.60)	0.156 (0.65)
$m = 50$	0.004 (0.98)	0.172 (1.06)	0.002 (0.97)	0.161 (1.05)	−0.004 (0.61)	0.136 (0.66)	−0.007 (0.60)	0.127 (0.65)
Censoring %	32.8%				32.7%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	1.035 (0.48)	1.166 (0.47)	0.849 (0.42)	0.977 (0.42)	0.968 (0.33)	1.075 (0.32)	0.801 (0.29)	0.904 (0.28)
$m = 10$	0.402 (0.38)	0.524 (0.40)	0.352 (0.36)	0.466 (0.38)	0.386 (0.26)	0.487 (0.28)	0.337 (0.24)	0.433 (0.26)
$m = 20$	0.145 (0.36)	0.247 (0.38)	0.130 (0.35)	0.226 (0.37)	0.151 (0.24)	0.234 (0.26)	0.135 (0.23)	0.213 (0.26)
$m = 30$	0.066 (0.36)	0.166 (0.38)	0.061 (0.34)	0.153 (0.37)	0.081 (0.24)	0.160 (0.26)	0.071 (0.23)	0.146 (0.25)
$m = 40$	0.026 (0.35)	0.121 (0.38)	0.022 (0.35)	0.111 (0.37)	0.045 (0.24)	0.117 (0.26)	0.039 (0.24)	0.109 (0.25)
$m = 50$	0.004 (0.35)	0.096 (0.38)	0.002 (0.35)	0.089 (0.37)	0.024 (0.24)	0.085 (0.26)	0.019 (0.24)	0.088 (0.25)
Censoring %	32.7%				32.6%			

Table 11 Case 2: $\boldsymbol{\beta_0 = -1}$, t_{ij}'s generated from $\boldsymbol{U(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−1.045 (0.95)	−0.971 (0.97)	−0.982 (0.82)	−0.931 (0.83)	−0.995 (0.59)	−0.952 (0.60)	−0.939 (0.52)	−0.909 (0.53)
$m = 10$	−1.144 (0.91)	−1.072 (0.92)	−1.095 (0.83)	−1.040 (0.84)	−1.043 (0.58)	−1.012 (0.58)	−1.022 (0.54)	−0.997 (0.55)
$m = 20$	−1.148 (0.93)	−1.083 (0.95)	−1.122 (0.89)	−1.070 (0.91)	−1.093 (0.55)	−1.062 (0.56)	−1.071 (0.52)	−1.044 (0.53)
$m = 30$	−1.158 (0.94)	−1.091 (0.96)	−1.144 (0.91)	−1.089 (0.94)	−1.044 (0.57)	−1.012 (0.57)	−1.031 (0.55)	−1.003 (0.56)
$m = 40$	−1.117 (0.88)	−1.053 (0.90)	−1.105 (0.86)	−1.052 (0.89)	−1.075 (0.57)	−1.042 (0.59)	−1.068 (0.57)	−1.040 (0.58)
$m = 50$	−1.236 (0.98)	−1.172 (0.99)	−1.222 (0.95)	−1.168 (0.97)	−1.099 (0.57)	−1.070 (0.57)	−1.091 (0.56)	−1.065 (0.57)
Censoring %	36.5%				36.4%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.974 (0.32)	−0.960 (0.32)	−0.912 (0.28)	−0.904 (0.28)	−0.939 (0.22)	−0.929 (0.22)	−0.895 (0.20)	−0.886 (0.20)
$m = 10$	−1.034 (0.34)	−1.021 (0.34)	−1.007 (0.32)	−0.997 (0.32)	−0.988 (0.22)	−0.978 (0.22)	−0.963 (0.21)	−0.953 (0.21)
$m = 20$	−1.001 (0.30)	−0.992 (0.30)	−0.987 (0.29)	−0.978 (0.29)	−1.008 (0.21)	−0.998 (0.21)	−0.996 (0.21)	−0.987 (0.21)
$m = 30$	−1.015 (0.31)	−1.005 (0.31)	−1.007 (0.30)	−0.997 (0.30)	−0.990 (0.22)	−0.979 (0.22)	−0.997 (0.21)	−0.968 (0.21)
$m = 40$	−1.010 (0.31)	−0.999 (0.31)	−1.005 (0.31)	−0.994 (0.31)	−0.996 (0.22)	−0.986 (0.22)	−0.993 (0.22)	−0.985 (0.21)
$m = 50$	−1.014 (0.30)	−1.003 (0.30)	−1.013 (0.30)	−1.003 (0.30)	−0.993 (0.22)	−0.983 (0.22)	−0.989 (0.22)	−0.979 (0.22)
Censoring %	36.4%				36.3%			

Table 12 Case 2: $\boldsymbol{\beta_0 = -1}$, $\boldsymbol{t_{ij}}$'s evenly chosen in $\boldsymbol{(0, V_i)}$

	$n = 100$				$n = 200$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.005 (0.99)	0.102 (1.06)	−0.099 (0.88)	−0.011 (0.94)	−0.187 (0.55)	−0.129 (0.60)	−0.249 (0.50)	−0.201 (0.54)
$m = 10$	−0.605 (0.87)	−0.532 (0.92)	−0.599 (1.02)	−0.559 (0.88)	−0.647 (0.53)	−0.611 (0.54)	−0.654 (0.51)	−0.624 (0.53)
$m = 20$	−0.878 (0.87)	−0.812 (0.90)	−0.860 (1.00)	−0.821 (0.89)	−0.848 (0.53)	−0.818 (0.54)	−0.843 (0.51)	−0.817 (0.53)
$m = 30$	−0.968 (0.88)	−0.906 (0.90)	−0.940 (0.98)	−0.900 (0.88)	−0.919 (0.53)	−0.890 (0.54)	−0.913 (0.52)	−0.886 (0.53)
$m = 40$	−1.012 (0.88)	−0.950 (0.91)	−0.981 (0.99)	−0.947 (0.89)	−0.950 (0.53)	−0.922 (0.54)	−0.945 (0.52)	−0.919 (0.53)
$m = 50$	−1.036 (0.88)	−0.971 (0.90)	−1.013 (0.99)	−0.974 (0.90)	−0.968 (0.53)	−0.940 (0.54)	−0.966 (0.53)	−0.939 (0.54)
Censoring %	36.5%				36.4%			
	$n = 500$				$n = 1000$			
	$Z_i(t)$		$\tilde{Z}_i(t)$		$Z_i(t)$		$\tilde{Z}_i(t)$	
	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$	Ave. $\hat{\beta}_n$	Ave. $\tilde{\beta}_n$
$m = 5$	−0.288 (0.30)	−0.268 (0.32)	−0.331 (0.28)	−0.318 (0.29)	−0.313 (0.21)	−0.304 (0.21)	−0.353 (0.20)	−0.346 (0.20)
$m = 10$	−0.672 (0.30)	−0.660 (0.30)	−0.676 (0.29)	−0.666 (0.29)	−0.678 (0.20)	−0.672 (0.20)	−0.681 (0.19)	−0.674 (0.19)
$m = 20$	−0.844 (0.30)	−0.833 (0.30)	−0.847 (0.30)	−0.837 (0.30)	−0.840 (0.21)	−0.831 (0.21)	−0.835 (0.20)	−0.826 (0.20)
$m = 30$	−0.902 (0.30)	−0.891 (0.30)	−0.897 (0.30)	−0.887 (0.30)	−0.889 (0.20)	−0.881 (0.20)	−0.886 (0.20)	−0.878 (0.20)
$m = 40$	−0.928 (0.30)	−0.918 (0.30)	−0.927 (0.30)	−0.919 (0.30)	−0.915 (0.21)	−0.907 (0.21)	−0.913 (0.21)	−0.904 (0.20)
$m = 50$	−0.946 (0.31)	−0.935 (0.31)	−0.945 (0.31)	−0.935 (0.31)	−0.933 (0.21)	−0.926 (0.21)	−0.931 (0.21)	−0.924 (0.21)
Censoring %	36.4%				36.3%			

4.3 Discussion and concluding remarks

From Tables 1–12 we see that the right censoring percentage is rather high, but our algorithm runs stably. The MLE $\hat{\beta}_n$ and the partial likelihood estimator $\tilde{\beta}_n$ have similar performance, while in most cases $\hat{\beta}_n$ performs better. It should be noted that for smaller sample size n and low intensity m of intensive longitudinal covariate, $\hat{\beta}_n$ and $\tilde{\beta}_n$ perform better when $\hat{Z}_i(t)$ in (18) is calculated using the true value $Z_i(t)$, which is expected. Overall, the estimation of both $\hat{\beta}_n$ and $\tilde{\beta}_n$ for β_0 improves as the sample size n and the intensity m of intensive longitudinal covariate increase. This is consistent with the asymptotic properties of $\hat{\beta}_n$ and $\tilde{\beta}_n$ shown in Ren (2023).

For both Case 1 and Case 2 simulation studies, we can see that for time points t_{ij}'s generated from $U(0, V_i)$, sometimes the performance of $\hat{\beta}_n$ and $\tilde{\beta}_n$ does not improve steadily as m increases, which is due to the randomness of t_{ij}'s. But for time points t_{ij}'s evenly chosen in interval $(0, V_i)$, the performance of $\hat{\beta}_n$ and $\tilde{\beta}_n$ clearly improves as m increases, which makes sense because the bigger the m is, the more time points we observe on the longitudinal covariates that evenly capture the covariate process $X(t)$ over time interval $(0, t)$. Thus, our simulation results here suggest better designed time points for observing the intensive longitudinal data in certain situations.

In summary, our simulation studies here show that without parametric/semiparametric assumptions or use of kernel method on the longitudinal covariate process $X(t)$ in PH model (8), our MLE $\hat{\beta}_n$ performs well for large sample size n and high intensity m.

Variance Estimate and Model Checking. In Ren (2023), it is shown that both $\hat{\beta}_n$ and $\tilde{\beta}_n$ are asymptotically normal and that for MLE $\hat{F}_n(t)$ given in (29), $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process on any compact set within the support of survival time T. Thus, using centered n out of n bootstrap (Bickel and Ren 2001; Bickel et al. 1997), we can obtain the variance estimate for $\hat{\beta}_n$ and $\tilde{\beta}_n$. As for model checking for PH model (5) or PH model (8) in real data analysis, it has not been studied in literature. Based on the asymptotic results on MLE $\hat{F}_n(t)$ in Ren (2023) above mentioned and based on the idea in Ren and He (2011), a goodness-of-fit test will be constructed in an upcoming paper.

5 Real data analysis

Here, we conduct analysis for *Smoking Cessation Data* and *Liver Disease Data*, respectively. Although we know that for observed data (11), assumption (24) on not having ties among V_i's holds true with probability one for continuous lifetime variable, in practice we often encounter cases with ties among V_i's in data (11) due to measurement error or round off error. Thus, with the proofs given in Appendix, the general versions of estimating functions $\Psi_n(\beta)$ in (30), $\psi_n(\beta)$ in (31) and $\Psi_{M,n}(\beta)$ in (37) with ties among V_i's are presented below before our real data analysis.

General versions of estimating functions

For observed data (11), assume that all distinct observations among $V_1, \cdots, V_n$ are:

$$U_1 < U_2 < \cdots < U_N, \tag{43}$$

where $N \le n$. Then, with algebraic work, we write the likelihood function in (22) as:

$$\begin{aligned} L(\beta, F) &= \prod_{k=1}^{N} \prod_{V_i = U_k} \left[c_i\, dF(V_i)\right]^{\delta_i} \left[\bar{F}(V_i)\right]^{c_i - \delta_i} \Big(\prod_{j=1}^{J_i} \left[\bar{F}(t_{ij})\right]^{c_{i,j-1} - c_{ij}} \Big) \\ &= \prod_{k=1}^{N} \left[\tilde{c}_k\, dF(U_k)\right]^{\tilde{\delta}_k} \left[\bar{F}(U_k)\right]^{\tilde{d}_k - \tilde{\delta}_k} = \prod_{k=1}^{N} \left[\tilde{c}_k\, p_k\right]^{\tilde{\delta}_k} \Big(\sum_{j=k+1}^{N+1} p_j \Big)^{\tilde{d}_k - \tilde{\delta}_k} \equiv L(\beta, \boldsymbol{p}) \end{aligned} \tag{44}$$

where $F(t) = \sum_{k=1}^{N} p_k I\{U_k \le t\}$, $\boldsymbol{p} = (p_1, \cdots, p_N, p_{N+1})$ and for c_{ij} given in (23),

$$\begin{cases} \tilde{\delta}_k = \sum_{i=1}^{n} \delta_i I\{V_i = U_k\} \\ \tilde{c}_k = I\{\tilde{\delta}_k = 0\} + e^{(\beta/\tilde{\delta}_k) \sum_{i=1}^{n} \tilde{Z}_i(t_{iJ_i}) \delta_i I\{V_i = U_k\}} I\{\tilde{\delta}_k > 0\} \\ \tilde{d}_k = \sum_{i=1}^{n} \big(c_{iJ_i} I\{V_i = U_k\} + \sum_{j=1}^{J_i} (c_{i,j-1} - c_{ij}) I\{U_k \le t_{ij} < U_{k+1}\} \big) \end{cases} . \tag{45}$$

Thus, from the derivations of (30)–(31) and (37), respectively, we obtain the following general versions of these estimating functions when V_i's in observed data (11) have ties as given by above (43):

$$\begin{cases} \Psi_n(\beta) = n^{-1} \sum_{k=1}^{N} [\tilde{Z}_k I\{\tilde{\delta}_k >\} + \tilde{e}'_k \log(1 - \frac{\tilde{\delta}_k}{\tilde{e}_k})], \quad \text{with} \quad \tilde{e}'_k = \frac{d\tilde{e}_k}{d\beta} \\ \qquad\qquad \psi_n(\beta) = n^{-1} \sum_{k=1}^{N} [\tilde{Z}_k I\{\tilde{\delta}_k > 0\} - \frac{\tilde{e}'_k \tilde{\delta}_k}{\tilde{e}_k})] \\ \Psi_{M,n}(\beta) = n^{-1} \sum_{k=1}^{N} [\tilde{Z}_k I\{\tilde{\delta}_k > 0\} + \tilde{E}'_k \log(1 - \frac{\tilde{\delta}_k}{\tilde{E}_k})], \quad \text{with} \quad \tilde{E}'_i = \frac{d\tilde{E}_i}{d\beta} \end{cases} \tag{46}$$

where $\tilde{Z}_k = \sum_{i=1}^{n} \delta_i \tilde{Z}_i(t_{iJ_i}) I\{V_i = U_k\}$ and $\tilde{E}_k = e^M \tilde{e}_k$ for $\tilde{e}_k = \tilde{d}_k + \cdots + \tilde{d}_N$.

From a similar proof of Lemma 4, we can simplify the expression of $\tilde{e}_k$'s and can show that above general version of partial likelihood estimating function $\psi_n(\beta)$ is non-increasing in β. Thus, the data-based choice of M for above general version of generalized estimating function $\Psi_{M,n}(\beta)$ follows the same idea as described in Section 4. For our real data analysis in this section, we use above general versions of estimating functions in (46) to compute MLE $\hat{\beta}_n$ and partial likelihood estimator $\tilde{\beta}_n$ because there are ties among V_i's in the datasets under consideration, but $\Psi_{M,n}(\beta)$ is never needed here.

5.1 Smoking cessation data analysis

Consider the Smoking Cessation Data from Piper et al. (2009), and we use the dataset processed by Liu (2014) which contains the record of participants for a period of two weeks after their quitting smoking.

Here, the first lapse time T is the same as described in Example 2 of Section 2.2 with 59.7% right censoring/truncation percentage, and during the study each participant was randomly sampled 4 times per day to record: urge for smoking, negative effect and smoking cessation fatigue. There are 794 participants included in this study, who were in the following three treatment groups: *Placebo* with $n_1 = 83$ participants, *Monotherapy* with $n_2 = 412$ participants, and *Combined Pharmacotherapy* with $n_3 = 299$ participants.

Under our proposed PH model (8) with modeling function given by (38), we compute estimators $\hat{\beta}_n$ and $\tilde{\beta}_n$ based on this dataset and display the results in Table 13.

Discussion: In Liu (2014), a form of PH model (5) was used assuming the longitudinal covariate process to be Gaussian, and the resulting estimates based on the whole dataset for regression parameters of urge, negative effect and fatigue are 0.13, 0.15 and 0.03, respectively. Under our PH model (8) using $Z(t) = X(t)$ without any assumptions on $X(t)$, our $\hat{\beta}_n$ based on the whole dataset for these parameters are 0.10, 0.11 and 0.03, respectively. Evidently, our estimates $\hat{\beta}_n$ for regression parameters of urge and negative effect are smaller than those by Liu (2014), which could be due to the use of Gaussian assumption on the longitudinal covariate process in Liu's procedure.

Our further analysis under PH model (8) using $Z(t) = X(t)$ shows that under therapies, negative effect has positive impact to lapse time T, while under placebo, negative effect has no impact to T. Table 13 displays our estimates under proposed PH model (8) using the "average" modeling function $Z(t) = \varphi(X^t)$ given in (38), which indicates that under therapies, urge has less impact to T than placebo, because $\hat{\beta}_n = 0.11, 0.09$ are smaller than $\hat{\beta}_n = 0.17$, while under therapies and placebo, negative effect has similar impact to T, because $\hat{\beta}_n = 0.11, 0.11, 0.14$ are quite similar. Thus, the results in Table 13 make much more sense and suggest that the use of modeling function makes the observed longitudinal data more informative. Interestingly, our $\hat{\beta}_n$ and $\tilde{\beta}_n$ in Table 13 based on the whole dataset for regression parameters of urge, negative effect and fatigue are quite close to those above mentioned by Liu (2014) who did not consider the use of "average" modeling function.

5.2 Liver disease data analysis

Consider the Liver Disease Data which was collected from a Primary Biliary Cirrhosis (PBC) clinical trial (Murtaugh et al. 1994) conducted by Mayo Clinic

Table 13 Estimation for Smoking Cessation Data Using Average Modeling Function

		Urge		Negative Effect		Fatigue	
	#	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$
Whole Data Set	794	0.12	0.13	0.13	0.15	0.03	0.04
Placebo	83	0.17	0.18	0.11	0.14	0.06	− 0.05
Monotherapy	412	0.11	0.12	0.11	0.13	0.05	0.05
Combined Therapy	299	0.09	0.10	0.14	0.16	0.03	0.04

Table 14 Estimation for Liver Disease Data

	#	log(bilirubin)		bilirubin	
		$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$
Whole Data Set	312	1.2871	1.2891	0.1506	0.1507
Placebo	154	1.4072	1.4111	0.1830	0.1833
Treatment	158	1.1582	1.1621	0.1299	0.1303

between 1974 and 1984. During PBC trial period, 312 patients participated in the randomized trial of the drug D-penicillamine: *Placebo* with $n_1 = 154$ patients; *D-penicillamine* with $n_2 = 158$ patients. In an earlier study, Shapiro et al. (1979) showed that after a relatively stable phase, serum bilirubin increased sharply in the months preceding death.

Here, the survival time T is the time from entering the clinical trial to death and is subject to right censoring due to drop-outs, the end of the study, liver transplantation and other factors. The right censoring percentage of the data set is 55.13%, and observed data (11) have V_i's ranging $0.11 - 14.32$ years (more follow-up visits occurred after 1984). For each patient, longitudinal covariate serum bilirubin was measured at follow-up lab visits at 6 months, 1 year and annually thereafter, which gives $\mathcal{X}_i^{V_i}$ in data (11).

This is a data example similar to Example 1 in Sect. 2.2, thus we conduct our analysis under PH model (5), i.e., PH model (8) using $Z(t) = \log X(t)$ and $Z(t) = X(t)$, respectively, without any assumptions on longitudinal covariate process $X(t)$. Here, the use of $\log X(t)$ is following the clinical literature on bilirubin. Table 14 displays our estimators $\hat{\beta}_n$ and $\tilde{\beta}_n$ based on this PBC dataset.

Discussion: Ding and Wang (2008) conducted analysis for the same data under PH model (5) using $\log X(t)$ as the longitudinal covariate with Gaussian assumption. Their resulting estimate for β_X is $\hat{\beta}_X = 1.07$, which is smaller than our $\hat{\beta}_n = 1.2871$ and $\tilde{\beta}_n = 1.2891$ in Table 14 for the whole data set. Thus, our results without any parametric or semi-parametric assumptions on $X_i(t)$ are in stronger agreement with the conclusion by Shapiro et al. (1979). Interestingly, our results in Table 14 show that $\hat{\beta}_n$ and $\tilde{\beta}_n$ for treatment group are smaller than those for placebo group, which indicates that treatment has some effect to the disease.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-024-00899-5.

Acknowledgements This project was initially conceived during Dr. Ren's sabbatical year spent in *The Methodology Center* at Pennsylvania State University, and she is grateful for many discussions with Dr. Runze Li. Dr. Ren's research in this article was supported in part by U.S. National Science Foundation grants DMS-0905772 & DMS-1232424. The authors thank the Editor, Associate Editor and two referees for their valuable comments and suggestions on the earlier draft of this article.

Funding Research supported in part by NSF grants DMS-0905772 & DMS-1232424

References

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

Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Bickel, P. J., Ren, J. (2001). The bootstrap in hypothesis testing. state of the art in statistics and probability theory, Festschrift for Willem R. van Zwet. *Lecture Notes in Mathematical Statistics, IMS, 36*, 91–112.

Bickel, P. J., Gotze, F., van Zwet, W.R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica, 7*, 1–31.

Cai, S., Chen, J. (2018). Empirical likelihood inference for multiple censored samples. *Canadian Journal of Statistics, 46*, 212–232.

Cao, H., Churpek, M. M., Zeng, D., Fine, J. P. (2015). Analysis of the proportional hazards model wtih sparse longitudinal covariates. *Journal of American Statistical Association, 110*, 1181–1196.

Chang, H.-W., McKeague, I. W. (2022). Empirical likelihood based inference for functional means with application to wearable device data. *Journal of the Royal Statistical Society, Series B, 84*, 1947–1968.

Chang, M. N., Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics, 15*, 1536–1547.

Chen, J., Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika, 80*, 107–116.

Cole, P. M., Tan, P. Z., Hall, S. E., Zhang, Y., Crnic, K. A., Blair, C. B., Li, R. (2011). Developmental changes in anger expression and attention focus: Learning to wait. *Developmental Psychology, 47*, 1078–1089.

Cox, D. R. (1972). Regression models and life-tables (with Discussion). *Journal of The Royal Statistical Society, Series B, 34*, 187–220.

DiCiccio, T., Hall, P., Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics, 19*, 1053–1061.

Ding, J., Wang, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics, 64*, 546–556.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of American Statistical Association, 72*, 557–565.

Fleming, T. R., Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gu, M. G., Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics, 21*, 611–624.

Henderson, R., Diggle, P., Dobson, A. (2000). Joint modeling longitudinal measurements and event time data. *Biostatistics, 4*, 465–480.

Hsieh, F., Tseng, Y. K., Wang, J. L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics, 62*, 1037–1043.

Kalbfleisch, J.D., Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd Edition, Hoboken, New Jersey: John Wiley & Sons, INC.

Li, R., Ren, J. (2011). An overview on joint modeling of censored survival time and longitudinal data. *Analysis of High Dimensional Data*, Edited by T.Cai and X.T. Shen, 195-222.

Li, R., Root, T., Shiffman, S. (2006). A local linear estimation procedure for functional multilevel modeling. *Models for Intensively Longitudinal Data* (T. Walls and J. Schafer, Eds.), 63-83. Oxford University Press.

Liu, X. (2014). *Joint Modeling of Longitudinal and Survival Data: New Models, Computing Techniques and Applications*. Doctoral dissertation, Pennsylvania State University.

Muller, U. U., Peng, H., Schick, A. (2019). Inference about the slope in linear regression: An empirical likelihood approach. *The Annals of Institute of Statistical Mathematics, 71*, 181–211.

Murtaugh, P. A., Dickson, E. R., Dam, G. M. V., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., Gips, C. H. (1994). Primary biliary cirrhosis: Prediction of shortterm survival based on repeated patient visits. *Hepatology, 20*, 126–134.

Mykland, P. A. (1995). Dual likelihood. *The Annals of Statistics, 23*, 396–421.

Mykland, P. A., Ren, J. (1996). Self-consistent and maximum likelihood estimation for doubly censored data. *The Annals of Statistics, 24*, 1740–1764.

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika, 64*, 441–448.
Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika, 75*, 237–249.
Owen, A. B. (2001). *Empirical Likelihood*. London: Chapman and Hall.
Piper, M. E., Smith, S. S., Schlam, T. R., Fiore, M. C., Jorenby, D. E., Fraser, D., Baker, T. B. (2009). A randomized placebo-controlled clinical trial of five smoking cessation pharmacotherapies. *Archives of General Psychiatry, 66*, 1253–1262.
Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics, 22*, 300–325.
Ren, J. (2001). Weighted empirical likelihood ratio confidence intervals for the mean with censored data. *The Annals of Institute of Statistical Mathematics, 53*, 498–516.
Ren, J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics, 36*, 147–166.
Ren, J. (2023). Asymptotic properties of empirical likelihood-based MLE for proportional hazards model with right censored survival time and intensive longitudinal covariates. (Preprint)
Ren, J., He, B. (2011). Estimation and goodness-of-fit for the Cox model with various types of censored data. *Journal of Statistical Planning and Inference, 141*, 961–971.
Ren, J., Riddlesworth, T. D. (2014). Empirical likelihood bivariate nonparametric maximum likelihood estimator with right censored data. *The Annals of Institute of Statistical Mathematics, 66*, 913–930.
Ren, J., Shi, Y. (2023). Computation and simulation studies on the MLE for joint modeling right censored survival data with longitudinal covariates. (In preparation)
Ren, J., Zhou, M. (2011). Full likelihood inferences in the Cox model: An empirical likelihood approach. *The Annals of Institute of Statistical Mathematics, 63*, 1005–1018.
Shapiro, J. M., Smith, H., Schaffner, F. (1979). Serum bilirubin: A prognostic factor in primary biliary cirrhosis. *Gut, 20*, 137–140.
Shiffman, S., Gnys, M., Richards, T. J., Paty, J. A., Hickcox, M. (1996). Temptations to smoke after quitting: A comparison of lapsers and maintainers. *Health Psychology, 15*, 455–461.
Su, Y.-R., Wang, J.-L. (2012). Modeling left-truncated and right-censored survival data with longitudinal covariates. *The Annals of Statistics, 40*, 1465–1488.
Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics, 9*, 93–108.
Tsiatis, A. A., Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica, 14*, 809–834.
Tsiatis, A. A., Degruttola, V., Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: applications to survival data and CD4 counts in patients with AIDS. *Journal of American Statistical Association, 90*, 27–37.
Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of American Statistical Association, 69*, 169–173.
Wang, D., Wu, T., Zhao, Y. (2019). Penalized empirical likelihood for the sparse Cox regression model. *Journal of Statistical Planning & Inference, 201*, 71–85.
Zeng, D., Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics, 33*, 2132–2163.
Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika, 92*, 492–498.
Zhou, M. (2016). *Empirical Likelihood Method in Survival Analysis*. New York: CRC Press.