Statistics in Biosciences (2023) 15:430–454
https://doi.org/10.1007/s12561-023-09372-y

Analysis of the Cox Model with Longitudinal Covariates with Measurement Errors and Partly Interval Censored Failure Times, with Application to an AIDS Clinical Trial

Yanqing Sun[1] · Qingning Zhou[1] · Peter B. Gilbert[2]

Received: 13 May 2022 / Revised: 18 April 2023 / Accepted: 27 April 2023 /
Published online: 20 May 2023

Abstract

Time-dependent covariates are often measured intermittently and with measurement errors. Motivated by the AIDS Clinical Trials Group (ACTG) 175 trial, this paper develops statistical inferences for the Cox model for partly interval censored failure times and longitudinal covariates with measurement errors. The conditional score methods developed for the Cox model with measurement errors and right censored data are no longer applicable to interval censored data. Assuming an additive measurement error model for a longitudinal covariate, we propose a nonparametric maximum likelihood estimation approach by deriving the measurement error induced hazard model that shows the attenuating effect of using the plug-in estimate for the true underlying longitudinal covariate. An EM algorithm is devised to facilitate maximum likelihood estimation that accounts for the partly interval censored failure times. The proposed methods can accommodate different numbers of replicates for different individuals and at different times. Simulation studies show that the proposed methods perform well with satisfactory finite-sample performances and that the naive methods ignoring measurement error or using the plug-in estimate can yield large biases. A hypothesis testing procedure for the measurement error model is proposed. The proposed methods are applied to the ACTG 175 trial to assess the associations of treatment arm and time-dependent CD4 cell count on the composite clinical endpoint of AIDS or death.

Keywords AIDS clinical trial · Cox model · Longitudinal covariates · Measurement errors · Partly interval censored data

✉ Qingning Zhou
qzhou8@uncc.edu

[1] Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC, USA

[2] Vaccine and Infectious Disease and Public Health Sciences Divisions, Fred Hutchinson Cancer Center, Seattle, WA, USA

1 Introduction

This work is motivated by the AIDS Clinical Trials Group (ACTG) 175 trial that compared four antiretroviral regimens in study participants living with HIV-1 [1]. CD4 cell count has long been considered as an important prognostic biomarker for disease progression. The participants had CD4 cell count measured every 12 weeks and were followed for occurrence of the composite clinical endpoint of AIDS or death. One of the study objectives was to assess the associations of treatments and time-dependent CD4 cell count with the clinical endpoint of AIDS or death. One complication is that CD4 cell count is measured intermittently and with measurement errors. Naive approaches that ignore the measurement errors or replace them with their estimated values can lead to biased estimation [2, 3]. The other challenge is that the time to the clinical endpoint is partly interval censored in which the time to death is subject to right censoring while the time to AIDS is interval censored between two visit dates. Statistical modeling with interval censored data has been well studied. There has also been extensive study of measurement errors in covariates for right censored data. However, to the best of our knowledge, no methodology exists for achieving valid statistical inference for the proportional hazards model with both partly interval censored failure times and longitudinal covariates subject to measurement error.

Many authors have studied regression analysis of interval censored failure time data under the Cox proportional hazards model [4–8]. Most of the existing work focused on time-independent covariates. Recently, Zeng et al. [9] considered maximum likelihood estimation for a class of semiparametric transformation models that includes the proportional hazards model and allows for time-dependent covariates. On the other hand, research on partly interval censored failure time data is fairly limited. Due to the presence of exact failure times, partly interval censored data requires a different treatment than interval censored data. To our knowledge, only two papers considered the proportional hazards model for partly interval censored data. In particular, Kim [10] studied maximum likelihood estimation for the proportional hazards model with time-independent covariates, while Zhou et al. [11] developed an EM algorithm for nonparametric maximum likelihood estimation for a class of semiparametric transformation models as in Zeng et al. [9] that allows for time-dependent covariates for partly interval censored data. However, these two papers did not consider covariate measurement errors.

There is extensive literature on statistical methods for right censored failure time data when some covariates are subject to measurement error. It is well known that standard estimation procedures yield biased estimation if measurement error is not taken into account. For the proportional hazards model, Prentice [2] showed that the naive approach using the observed covariate values with measurement errors in place of the underlying covariate values in the partial likelihood may result in substantial estimation bias, and proposed a modified partial likelihood method that required estimating the conditional expectation of the hazard of each individual at each failure time. Nakamura [12] proposed the

corrected partial score method that yielded approximately unbiased estimates. Hughes [13] investigated regression dilution bias in the presence of covariate measurement errors. Buzas [14] removed the condition of normal errors while assuming that the moment generating function of the error distribution is known. Huang and Wang [15] proposed a nonparametric-correction approach for the Cox proportional hazards model. Hu and Lin [16] extended the work of Nakamura [12] and Buzas [14] to obtain a class of consistent estimators when the true covariate is ascertained on a randomly selected validation set. Tsiatis and Davidian [3] proposed the conditional score estimator, which was further studied by Song et al. [17, 18]. Yi and Lawless [19] employed the corrected score methods of Nakamura [12] assuming a piecewise constant form of the baseline hazard function. Fu and Gilbert [20] extended the conditional score approach to accommodate missing values of the longitudinal covariates following a two-phase sampling design. Tsiatis and Davidian [21] overviewed joint modeling of longitudinal covariates and time-to-event data.

All of the aforementioned works considered right censored data. Although both measurement error problems and interval censored data have been well studied, the literature on statistical methods for interval censored failure time data with covariate measurement error is rather limited. Song and Ma [22] proposed a multiple imputation method for the Cox model with time-independent covariates to impute the time-to-event that falls within an interval and then analyzed the imputed data sets by the conditional score approach for right censored data. Mandal et al. [23] applied multiple imputation to handle both covariates with measurement error and interval censored failure time data under the linear transformation model. The imputed data were then analyzed using the method of Chen et al. [24]. Wen and Chen [25] proposed a conditional score approach for the proportional odds model with interval censoring and covariate measurement error using the working independence strategy. All of these approaches were developed assuming time-independent covariates with measurement errors. The multiple imputation approach depends on the imputation models and can only be approximate. The conditional score methods proposed by Tsiatis and Davidian [3] for right censored data and studied by many others [17, 18, 20] are no longer applicable for interval censored data.

In this article, we develop an estimation method for the Cox proportional hazards model for partly interval censored failure times and longitudinal covariates measured with error. Assuming an additive measurement error model for a longitudinal covariate, we propose a nonparametric maximum likelihood estimation approach by deriving the measurement error induced hazard model that shows the attenuating effect of ignoring measurement errors. An EM algorithm is devised to facilitate maximum likelihood estimation that accounts for the partly interval censored failure times. Simulation studies show that the proposed methods perform well with satisfactory finite-sample performances and that the naive methods ignoring measurement error or using the plug-in estimate can yield large biases. The simulation studies also show the attenuating bias of using the plug-in estimate for the true underlying longitudinal covariate. While the commonly used additive measurement error model for a time-independent covariate can be checked and often holds well in practice, use of a measurement error model for time-varying covariates requires

more care. Additive random effects models with known time-dependent basis functions are commonly used, but misspecification may lead to bias. Although statistical models of longitudinal covariates measured with error have been studied by many authors (e.g., papers noted above), few methods are available to evaluate their goodness-of-fit. In this article, we also propose a diagnostic testing procedure for the measurement error model of longitudinal covariates.

The rest of this article is organized as follows. Section 2 introduces the data structure, models and model assumptions. Section 3.1 derives the measurement error induced hazard model. Section 3.2 presents a nonparametric maximum likelihood estimation approach. An EM algorithm is devised to facilitate maximum likelihood estimation that accounts for the partly interval censored failure times. Section 3.3 derives the variance estimator based on the profile likelihood that accounts for variation in the parameter estimation for the measurement error model. A test procedure for the measurement error model of longitudinal covariates is given in Section 4. The finite-sample performance of the proposed methods is examined through simulation studies in Section 5. The proposed methods are applied to the ACTG 175 trial data in Section 6. Some concluding remarks are given in Section 7.

2 Preliminaries

Suppose T_i is the failure time of interest with the end of follow-up time τ. Let Z_i be the $d \times 1$ vector of time-independent covariates that includes baseline covariates and treatment assignment, and $X_i(t)$ the time-dependent covariate of interest. Let $\bar{X}_i(t) = \{X_i(u), 0 \le u \le t\}$ denote the history of $X_i(\cdot)$ up to time t. We assume that the conditional hazard function of T_i given $\bar{X}_i(t)$ and Z_i only depends on Z_i and the current value $X_i(t)$. Let $\lambda(t|\bar{X}_i(t), Z_i)$ be the conditional hazard function of T_i given $\bar{X}_i(t)$ and Z_i. We consider the proportional hazards model

$$\lambda(t|\bar{X}_i(t), Z_i) = \lambda(t) \exp\{\beta X_i(t) + \gamma^T Z_i\}, \tag{1}$$

for $0 \le t \le \tau$, where $\lambda(t)$ is an unspecified baseline function, and β and γ are 1- and d-dimensional vectors of parameters, respectively. We investigate model (1) under partly interval censored failure time data and when the time-dependent covariate $X_i(t)$ is subject to measurement error.

Partly interval censored failure time data include observations of failure times that are precisely observed, and failure times that are left, interval and/or right censored. Let Δ_i indicate whether the failure time T_i is exactly observed, i.e., $\Delta_i = 1$ if T_i is exactly observed and 0 otherwise. If $\Delta_i = 0$, let $(L_i, R_i]$ denote the smallest observed interval that brackets T_i, where $L_i \ge 0$ is the last monitoring time at which failure has not occurred and $R_i \ge 0$ is the first monitoring time at which failure has occurred. Let $R_i = \infty$ if failure has not occurred by the last monitoring time. Thus, if $L_i = 0$, T_i is left censored; if $R_i = \infty$, T_i is right censored; if $0 < L_i < R_i < \infty$, T_i is interval censored. The partly interval censored failure time data for individual i can be represented by $\{(\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i\}$. The notations $\Delta_i T_i$, $(1-\Delta_i)L_i$ and $(1-\Delta_i)R_i$ mean that we observe T_i if $\Delta_i = 1$ and observe $(L_i, R_i]$ if $\Delta_i = 0$.

In the ACTG 175 study, the failure time of interest is the time to composite endpoint of AIDS or death, whichever occurs first. For individual i, if death has occurred before AIDS, then we observe the exact death time T_i and $\Delta_i = 1$; if AIDS has occurred prior to death, then we observe a time interval $(L_i, R_i]$ that brackets the AIDS onset time T_i and $\Delta_i = 0$.

Linear mixed effects models are commonly used to model longitudinal covariates measured with errors [3, 17, 20]. Suppose that $X_i(t)$ is measured at times $v_{i1} < \cdots < v_{i,M_i}$ before τ with errors and there are B_{ij} repeated measurements or replicates of $X_i(v_{ij})$, where we let $B_{ij} = 1$ if there are no replicates. Let $W_{i,b}(v_{ij})$ denote the bth measurement of $X_i(\cdot)$ at time v_{ij}, $j = 1, \ldots, M_i$, $b = 1, \ldots, B_{ij}$. We consider the linear mixed effects model for longitudinal covariates with measurement errors:

$$W_{i,b}(v_{ij}) = X_i(v_{ij}) + e_{ij,b} = \theta_i^T f(v_{ij}) + e_{ij,b}, \tag{2}$$

where $f(v_{ij})$ is an $r \times 1$ vector of known design functions, θ_i is an $r \times 1$ vector of unobserved random effects, and $e_{ij,b}$ is the measurement error at time v_{ij}. We assume $\theta_i = \vartheta + v_i$, where ϑ is a vector of fixed parameters and v_i $(i = 1, \ldots, n)$ are independent and identically distributed (iid) $N(0, G)$ with G being a $r \times r$ nonnegative definite matrix. We also assume that $e_{ij,b}$ $(j = 1, \ldots, M_i, b = 1, \ldots, B_{ij})$ are iid $N(0, \sigma^2)$ independent of v_i. Thus, the unknown parameters for the measurement error model are $\theta_W = (\vartheta, G, \sigma^2)$. Also, note that the design function $f(\cdot)$ is usually chosen as a vector of basis functions, such as polynomials. In our simulation study and real data analysis below, we consider $f(t) = (1, t)$ or $(1, t, t^2)$.

Define $W_{ij} = (W_{i,1}(v_{ij}), \ldots, W_{i,B_{ij}}(v_{ij}))$ and $e_{ij} = (e_{ij,1}, \ldots, e_{ij,B_{ij}})$. Let $\tilde{v}_i = (v_{i1}, \ldots, v_{i,M_i})^T$, $\tilde{W}_i = (W_{i1}, \ldots, W_{i,M_i})^T$ and $\tilde{e}_i = (e_{i1}, \ldots, e_{i,M_i})^T$. The observed data consist of a random sample of n iid observations

$$\left\{\Delta_i,\ \Delta_i T_i,\ (1-\Delta_i)L_i,\ (1-\Delta_i)R_i,\ Z_i,\ \tilde{v}_i,\ \tilde{W}_i\right\}, \quad i = 1, \ldots, n.$$

We will employ individual-specific estimation of the longitudinal covariate $X_i(t)$ via model (2). It does not require repeated measurements at each measurement time v_{ij}, as long as the number of longitudinal measurements over time is sufficient to estimate θ_i, i.e., $M_i \geq r$. The proposed estimation method allows $B_{ij} = 1$ for all i, j. However, the repeated measurements reduce the standard error in estimating θ_i and thus in estimating $X_i(t)$, which results in increased efficiency in estimating β for model (1).

3 Estimation of the Cox Model with Partly Interval Censored Failure Times and Longitudinal Covariates with Measurement Errors

In this section, we propose a method for estimation of the Cox model (1). In Sect. 3.1, we derive the measurement error induced hazard model under the additive measurement error model for longitudinal covariates. In Sect. 3.2, we design an EM algorithm for the nonparametric maximum likelihood estimation of the

measurement error induced hazard model based on partly interval censored failure times. A variance estimation procedure is proposed in Sect. 3.3.

3.1 Measurement Error Induced Hazard Model

The true longitudinal covariate $X_i(t)$ is not observed. We obtain an individual-specific estimate $\hat{X}_i(t)$ of $X_i(t)$ using the ordinary least squares method based on the observed data $(\tilde{v}_i, \tilde{W}_i)$ and propose an approach by deriving the conditional hazard function of T_i at time t conditional on Z_i and $\hat{X}_i(t)$. Only the longitudinal covariates in the past can be meaningfully used to model current or future risk of failure. For example, in assessing the association of time-dependent CD4 cell count with the composite clinical endpoint of AIDS or death in the ACTG 175 trial, only the CD4 count measurements before AIDS or death are meaningfully associated with the endpoint. Therefore, we estimate $X_i(t)$ based on the data before t to preserve the predictability [3, 20].

Let $M_i(t)$ denote the index of the last measurement time before t such that $v_{i,M_i(t)} < t \le v_{i,M_i(t)+1}$. Since θ_i is r-dimensional, at least r longitudinal measurements from individual i before t are required, i.e., $M_i(t) \ge r$. Let $\tilde{v}_i(t) = (v_{i1}, \ldots, v_{i,M_i(t)})^T$, $\tilde{W}_i(t) = (W_{i1}, \ldots, W_{i,M_i(t)})^T$ and $\tilde{e}_i(t) = (e_{i1}, \ldots, e_{i,M_i(t)})^T$. Under model (2), $\tilde{W}_i(t) = \tilde{F}_i(t)\theta_i + \tilde{e}_i(t)$, where $\tilde{F}_i(t) = \mathbb{B}_i(t)\tilde{f}_i(t)^T$ with $\mathbb{B}_i(t) = \text{diag}(\mathbf{1}_{B_{i1}}, \ldots, \mathbf{1}_{B_{i,M_i(t)}})$, $\mathbf{1}_m$ is a $m \times 1$-vector of ones, and $\tilde{f}_i(t) = (f(v_{i1}), \ldots, f(v_{i,M_i(t)}))$. Hence the ordinary least squares estimator of θ_i based on $(\tilde{v}_i(t), \tilde{W}_i(t))$ for individual i equals

$$\hat{\theta}_i(t) = \{\tilde{F}_i^T(t)\tilde{F}_i(t)\}^{-1}\tilde{F}_i^T(t)\tilde{W}_i(t). \tag{3}$$

It is easy to see that $\tilde{F}_i^T(t)\tilde{F}_i(t) = \sum_{j=1}^{M_i(t)} B_{ij} f_i(v_{ij}) f_i^T(v_{ij})$ and $\tilde{F}_i^T(t)\tilde{W}_i(t) = \sum_{j=1}^{M_i(t)} f_i(v_{ij}) \sum_{b=1}^{B_{ij}} W_{i,b}(v_{ij})$.

We estimate θ_i based on the observations from subject i without pulling information from other individuals because only the past history of subject i can forecast his/her risk of failure. The longitudinal covariate $X_i(t)$ is estimated by $\hat{X}_i(t) = f^T(t)\hat{\theta}_i(t)$ based on the observed error-prone covariate information for individual i up to time t. This allows us to derive the measurement error induced hazard model conditional on the observed information from individual i's past.

Since $\hat{\theta}_i(t) = \theta_i + \{\tilde{F}_i^T(t)\tilde{F}_i(t)\}^{-1}\tilde{F}_i^T(t)\tilde{e}_i(t)$, we have

$$\hat{X}_i(t) = X_i(t) + f^T(t)\{\tilde{F}_i^T(t)\tilde{F}_i(t)\}^{-1}\tilde{F}_i^T(t)\tilde{e}_i(t).$$

The two terms $X_i(t)$ and $\tilde{e}_i(t)$ are independent under model (2). Then conditional on $(\theta_i, \tilde{v}_i(t))$, $\hat{X}_i(t)$ is normally distributed with mean $X_i(t)$ and variance $d_i(t, \sigma^2) = \sigma^2 f^T(t)\{\tilde{F}_i^T(t)\tilde{F}_i(t)\}^{-1}f(t)$. An estimator of σ^2 can be constructed using the residuals:

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} M_i^{-1}\sum_{j=1}^{M_i} B_{ij}^{-1}\sum_{b=1}^{B_{ij}}(W_{i,b}(v_{ij}) - \hat{X}_i(v_{ij}))^2. \tag{4}$$

Next, we derive the induced hazard model of T_i conditional on Z_i and $\hat{X}_i(t)$ under the measurement error model (2). The estimator $\hat{X}_i(t)$ is based on the observed information before t, and thus is predictable for the risk of failure at t. Define the counting process increment $dN_i(t) = I(t \le T_i < t + dt, v_{ir} \le t)$ and the at-risk process $Y_i(t) = I(T_i \ge t, v_{ir} \le t)$. That is, $dN_i(t) = 1$ if the failure time occurs at time t and after the r^{th} longitudinal measurement.

Our approach is motivated by the conditional score method of [3]. They first derived the conditional likelihood of $\{dN_i^*(t), \hat{X}_i(t)\}$ given $(\theta_i, Z_i, \tilde{v}_i(t), Y_i(t) = 1)$, where $N_i^*(t)$ is the counting process for the right censored data. They then noted that the conditional likelihood given $Y_i(t) = 1$, $Q_i(t, \beta, \sigma^2) = \hat{X}_i(t) + d_i(t, \sigma^2)\beta dN_i^*(t)$, is a complete sufficient statistic for θ_i, and thus, conditional on $Q_i(t, \beta, \sigma^2)$, removes the dependence of the conditional distribution on the random effects θ_i. [3, 20] derived the conditional intensity process by conditioning on $Q_i(t, \beta, \sigma^2)$, which turns out to be a Cox model with Z_i and $Q_i(t, \beta, \sigma^2)$ as the independent variables. We note that their papers did not derive the intensity model because of $dN_i^*(t)$ involved in $Q_i(t, \beta, \sigma^2)$. Further, their approaches do not work in the current setting because the counting process framework can not be utilized for interval censored or partly interval censored data.

We pursue a different approach by deriving the hazard model for T_i conditional on $Z_i, \tilde{v}_i(t)$, $Y_i(t) = 1$ and $\hat{X}_i(t)$. Let $\mathcal{F}_t$ be the filtration generated by $\{N_i(s), Y_i(s), Z_i, X_i(s), \hat{X}_i(s), \tilde{v}_i(s), \tilde{W}_i(s)\}$, $0 \le s \le t$. Then $\hat{X}_i(t)$ and $\sigma^2_{i,rel}(t)$ are both predictable with respect to $\mathcal{F}_t$. The following proposition presents the conditional hazard function of T_i at time t given $(\hat{X}_i(t), Z_i, \tilde{v}_i(t), v_{ir} \le t)$.

Proposition 1 *Under Conditions* (A1)-(A3) *given in the Appendix*,

$$\lambda^*(t|\hat{X}_i(t), Z_i, \tilde{v}_i(t)) = \lambda(t) \exp\left\{\beta\omega_i(t)\hat{X}_i(t) + \gamma^T Z_i + O_i(\beta, t, \theta_W)\right\}, \text{ for } t \ge v_{ir}, \tag{5}$$

where $\omega_i(t) = 1 - \sigma^2_{i,rel}(t)$, $O_i(\beta, t, \theta_W) = \beta\sigma^2_{i,rel}(t)[\vartheta^T f(t) + \frac{1}{2} f^T(t) G f(t)\beta]$, $\sigma^2_{i,rel}(t) = d_i(t, \sigma^2)/(f^T(t)Gf(t) + d_i(t, \sigma^2))$, *and* $\theta_W = (\vartheta, G, \sigma^2)$.

The proof of Proposition 1 *is given in Web Appendix A. We refer to model* (5) *as the measurement error induced hazard model. This approach based on the induced hazard model can be easily extended to handle multivariate* $X_i(t)$. *The parameter* $\sigma^2_{i,rel}(t)$ *measures the percentage of the measurement error variation over the total variation in* $W_{i,b}(v_{ij})$ *under model* (2). *The factor* $\omega_i(t) = 1 - \sigma^2_{i,rel}(t)$ *is termed as the reliability ratio* [26] *representing the attenuating effect of use of the estimated covariate* $\hat{X}_i(t)$. *If there is no measurement error, i.e.,* $\sigma^2 = 0$, *then* $\omega_i(t) = 1$, $O_i(\beta, t, \theta_W) = 0$ *and* $\hat{X}_i(t) = W_i(t) = X_i(t)$. *If the measurement times* $\tilde{v}_i$ *do not vary with i, then* $\sigma^2_{i,rel}(t)$ *does not depend on i.*

Since only the longitudinal measurements in the past can be meaningfully used to model current or future risk of failure, the subject-specific estimates, $\hat{X}_i(t)$, *are based on the measurements* W_{ij} *before* τ *if the failure event has not occurred by the end*

of study time, before T_i if the failure time is observed ($\Delta_i = 1$), and before L_i that is right before the failure time T_i if $\Delta_i = 0$.

3.2 Estimation of Measurement Error Induced Hazard Model with Partly Interval Censored Data

Next, we derive an estimator of the induced hazard model based on partly interval censored data. The observed data from a random sample of n study participants consist of $\{(\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i I(R_i < \infty), Z_i, \tilde{v}_i, \tilde{W}_i\}$, $i = 1, \ldots, n$. Recently, [11] developed maximum likelihood estimation for semiparametric transformation models with partly interval censored data. The method extended the EM algorithm approach of [9] for interval censored data to partly interval censored data. We adopt this approach to estimate the measurement error induced hazard model (5) with partly interval censored data.

Under model (5), the conditional survival function of T_i given $T_i \geq v_{ir}$ equals $\exp\big(-\int_{v_{ir}}^{t} \lambda^*(x|\hat{X}_i(x), Z_i, \tilde{v}_i(x))\,dx\big)$. Let $\Lambda_0(t) = \int_0^t \lambda(s)\,ds$. Note that θ_W in model (5) can be estimated based on model (2) such that we treat it as known for now. Let $h_i(t, \beta, \gamma) = \beta\omega_i(t)\hat{X}_i(t) + \gamma^T Z_i + O_i(\beta, t, \theta_W)$. The observed data likelihood function for (β, γ, Λ) under model (5) is $L_n(\beta, \gamma, \Lambda; \theta_W) =$

$$= \prod_{i=1}^{n} \Big\{ \big[\Lambda'(T_i)\exp\{h_i(T_i, \beta, \gamma)\}\big]^{I(v_{ir} \leq T_i)} \exp\Big(-\int_{v_{ir}}^{T_i} \exp\big\{h_i(t, \beta, \gamma)\big\}\,d\Lambda(t)\Big)\Big\}^{\Delta_i}$$
$$\Big\{\exp\Big(-\int_{v_{ir}}^{L_i} \exp\big\{h_i(t, \beta, \gamma)\big\}\,d\Lambda(t)\Big) - \exp\Big(-\int_{v_{ir}}^{R_i} \exp\big\{h_i(t, \beta, \gamma)\big\}\,d\Lambda(t)\Big)\Big\}^{1-\Delta_i}. \tag{6}$$

Because the likelihood (6) can become arbitrarily large within the class of absolutely continuous functions $\Lambda(\cdot)$, the nonparametric maximum likelihood estimator (NPMLE) is often obtained on a restricted space. Following this typical approach, e.g., [9], we regard $\Lambda(t)$ as a step function with nonnegative jumps at observed T_i and at the endpoints of the intervals $(L_i, R_i]$, $i = 1, \ldots, n$. Let $0 = t_0 < t_1 < \cdots < t_m$ be the ordered unique values of the set $\{(\Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i I(R_i < \infty)) : i = 1, \ldots, n\}$.

Let λ_k be the jump size of the estimator for $\Lambda(t)$ at t_k for $k = 1, \ldots, m$ and let $\lambda_0 = 0$. Let $h_i(t_{ik}, \beta, \gamma) = \beta\omega_{ik}\hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)$, where $\hat{X}_{ik} = \hat{X}_i(t_k)$ and $\omega_{ik} = \omega_i(t_k)$. With $\Lambda(t)$ a step function with jumps λ_k at t_k, $k = 1, \ldots, m$, the likelihood (6) becomes $L_n(\beta, \gamma, \Lambda; \theta_W) =$

$$= \prod_{i=1}^{n} \Big\{ \big[\Lambda\{T_i\}\exp\{h_i(T_i, \beta, \gamma)\}\big]^{I(v_{ir} \leq T_i)} \exp\Big(-\sum_{t_k \leq T_i} I(v_{ir} \leq t_k)\lambda_k \exp\big\{h_i(t_{ik}, \beta, \gamma)\big\}\Big)\Big\}^{\Delta_i}$$
$$\Big\{\exp\Big(-\sum_{t_k \leq L_i} I(v_{ir} \leq t_k)\lambda_k \exp\big\{h_i(t_{ik}, \beta, \gamma)\big\}\Big)$$
$$\Big[1 - \exp\Big(-\sum_{L_i < t_k \leq R_i} I(v_{ir} \leq t_k)\lambda_k \exp\big\{h_i(t_{ik}, \beta, \gamma)\big\}\Big)\Big]^{I(R_i < \infty)}\Big\}^{1-\Delta_i}, \tag{7}$$

where $\Lambda\{T_i\}$ denotes the jump size of $\Lambda(t)$ at T_i.

We consider an EM algorithm to maximize $L_n(\beta, \gamma, \Lambda; \theta_W)$. Let η_{ik} be independent Poisson random variables with means $\mu_{ik} = \lambda_k \exp\{h_i(t_{ik}, \beta, \gamma)\}$, $i = 1, \ldots, n$, $k = 1, \ldots, m$. Following [11], for $i = 1, \ldots, n$, we define

$$\begin{aligned} A_i &= \Delta_i \textstyle\sum_{t_k < T_i} I(v_{ir} \le t_k)\, \eta_{ik}, \\ B_i &= \Delta_i \textstyle\sum_{t_k = T_i} I(v_{ir} \le t_k)\, \eta_{ik}, \\ C_i &= (1-\Delta_i) \textstyle\sum_{t_k \le L_i} I(v_{ir} \le t_k)\, \eta_{ik}, \\ D_i &= (1-\Delta_i) I(R_i < \infty) \textstyle\sum_{L_i < t_k \le R_i} I(v_{ir} \le t_k)\, \eta_{ik}. \end{aligned}$$

Let $\hat{X}_{i\cdot} = \{\hat{X}_{ik}, k = 1, \ldots, m\}$. The likelihood of the observed data given by $(\tilde{v}_i, T_i, \hat{X}_{i\cdot}, Z_i, A_i = 0, B_i = 1)$ for $\Delta_i = 1$ and $(\tilde{v}_i, L_i, R_i, \hat{X}_{i\cdot}, Z_i, C_i = 0, D_i > 0)$ for $\Delta_i = 0$ under model (5) is

$$L_n^* = \prod_{i=1}^{n} \{P(A_i = 0, B_i = 1)\}^{\Delta_i} \{P(C_i = 0, D_i > 0)\}^{1-\Delta_i}.$$

Note that $P(A_i = 0, B_i = 1)$ equals the term in the likelihood (7) corresponding to $\Delta_i = 1$, and $P(C_i = 0, D_i > 0)$ equals the term in the likelihood (7) corresponding to $\Delta_i = 0$. Hence, L_n^* equals the observed likelihood (7), which takes the form

$$\begin{aligned} L_n(\beta, \gamma, \Lambda; \theta_W) = \prod_{i=1}^{n} &\left\{ \prod_{t_k < T_i} P(\eta_{ik} = 0)^{I(v_{ir} \le t_k)} \prod_{t_k = T_i} P(\eta_{ik} = 1)^{I(v_{ir} \le t_k)} \right\}^{\Delta_i} \\ &\left\{ \prod_{t_k \le L_i} P(\eta_{ik} = 0)^{I(v_{ir} \le t_k)} \left[1 - \prod_{L_i < t_k \le R_i} P(\eta_{ik} = 0)^{I(v_{ir} \le t_k)} \right]^{I(R_i < \infty)} \right\}^{1-\Delta_i}. \end{aligned} \tag{8}$$

We maximize the likelihood (8) through an EM algorithm by treating η_{ik} as missing data. Let $R_i^* = \Delta_i T_i + (1-\Delta_i)\{L_i I(R_i = \infty) + R_i I(R_i < \infty)\}$. Let $1_{ik}^* = I(v_{ir} \le t_k \le R_i^*)$. The complete-data log likelihood is given by

$$Cl_n(\beta, \gamma, \Lambda; \theta_W) = \sum_{i=1}^{n} \sum_{k=1}^{m} 1_{ik}^* \left[\eta_{ik} \log(\mu_{ik}) - \log(\eta_{ik}!) - \mu_{ik} \right]. \tag{9}$$

Taking derivatives of (9), we obtain the score functions

$$\frac{\partial Cl_n(\beta, \gamma, \Lambda; \theta_W)}{\partial(\beta, \gamma)} = \sum_{i=1}^{n} \sum_{k=1}^{m} 1_{ik}^* Z_{ik}^* \left[\eta_{ik} - \lambda_k \exp\{\beta \omega_{ik} \hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)\} \right], \tag{10}$$

$$\frac{\partial Cl_n(\beta, \gamma, \Lambda; \theta_W)}{\partial \lambda_k} = \sum_{i=1}^{n} 1_{ik}^* \left[\frac{\eta_{ik}}{\lambda_k} - \exp\{\beta \omega_{ik} \hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)\} \right], \tag{11}$$

for $k = 1, \ldots, m$, where $Z_{ik}^* = ((\omega_{ik}\hat{X}_{ik} + \dot{O}_i(\beta, t_k, \theta_W))^T, Z_i^T)^T$ and $\dot{O}_i(\beta, t_k, \theta_W)$ is the derivative of $O_i(\beta, t_k, \theta_W)$ with respect to β.

In the M-step, we calculate λ_k based on the score (11):

$$\lambda_k = \frac{\sum_{i=1}^n 1_{ik}^* \hat{E}(\eta_{ik})}{\sum_{i=1}^n 1_{ik}^* \exp\left\{\beta\omega_{ik}\hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)\right\}}, \tag{12}$$

for $k = 1, \ldots, m$, where $\hat{E}(\eta_{ik})$ denotes the posterior mean given the observed data. We then plug (12) into (10) and solve the score equations for β and γ:

$$\sum_{i=1}^n \sum_{k=1}^m 1_{ik}^* \hat{E}(\eta_{ik}) \left[Z_{ik}^* - \frac{\sum_{j=1}^n 1_{jk}^* \exp\left\{\gamma^T Z_j + \beta\omega_{jk}\hat{S}_{jk} + O_j(\beta, t_k, \theta_W)\right\} Z_{jk}^*}{\sum_{j=1}^n 1_{jk}^* \exp\left\{\gamma^T Z_j + \beta\omega_{jk}\hat{S}_{jk} + O_j(\beta, t_k, \theta_W)\right\}} \right] = 0. \tag{13}$$

The M-step estimators of λ_k $(k = 1, \ldots, m)$ and (β, γ) are obtained from (12) and (13).

In the E-step, we calculate the posterior mean $\hat{E}(\eta_{ik})$ of η_{ik} conditional on the observed data $(\tilde{v}_i, T_i, \hat{X}_{i\cdot}, Z_i, A_i = 0, B_i = 1)$ for $\Delta_i = 1$ and $(\tilde{v}_i, L_i, R_i, \hat{X}_{i\cdot}, Z_i, C_i = 0, D_i > 0)$ for $\Delta_i = 0$. For $\Delta_i = 1$, $\hat{E}(\eta_{ik}) = 0$ for $v_{ir} < t_k < T_i$ and $\hat{E}(\eta_{ik}) = 1$ for $v_{ir} < t_k = T_i$. For $\Delta_i = 0$, $\hat{E}(\eta_{ik}) = E(\eta_{ik}|\tilde{v}_i, L_i, R_i, \hat{X}_{i\cdot}, Z_i, C_i = 0, D_i > 0)$. It follows that $\hat{E}(\eta_{ik}) = 0$ for $v_{ir} \le t_k \le L_i$, and

$$\begin{aligned}\hat{E}(\eta_{ik}) &= E\left(\eta_{ik}\middle|\tilde{v}_i, L_i, R_i, \hat{X}_{i\cdot}, Z_i, C_i = 0, D_i > 0\right) \\ &= \frac{\lambda_k \exp\left\{\beta\omega_{ik}\hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)\right\}}{1 - \exp\{-\sum_{L_i < t_k \le R_i} 1_{ik}^* \lambda_k \exp\left\{\beta\omega_{ik}\hat{X}_{ik} + \gamma^T Z_i + O_i(\beta, t_k, \theta_W)\right\}\}},\end{aligned} \tag{14}$$

for $v_{ir} \le t_k$ and $L_i < t_k \le R_i$ with $R_i < \infty$.

The estimators of $(\lambda_k, k = 1, \ldots, m)$ and (β, γ) are obtained by iterating between the E and M steps until convergence, which are denoted by $(\hat{\lambda}_k, k = 1, \ldots, m)$ and $(\hat{\beta}, \hat{\gamma})$. We estimate $\Lambda(\cdot)$ by $\hat{\Lambda}(\cdot)$, which is the step function with jump size $\hat{\lambda}_k$ at t_k, $k = 1, \ldots, m$. This EM procedure assumes that the measurement error model parameters θ_W are known. In practice, they are usually unknown. These parameters can be estimated by existing methods for estimating a linear mixed effects model. In the numerical studies, we obtain the maximum likelihood estimates $\hat{\theta}_W$ using the *lmer* function in the R package *lme4* [27]. The aforementioned EM procedure is then carried out by replacing θ_W with $\hat{\theta}_W$. Therefore, $(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}(\cdot))$ is a plug-in estimator that maximizes $\log L_n(\beta, \gamma, \Lambda; \hat{\theta}_W)$ for $(\beta, \gamma) \in \mathcal{B}$ and $\Lambda \in \mathcal{C}$, where $\mathcal{B}$ is a known compact set in R^{d+1} and $\mathcal{C}$ is the set of step functions with nonnegative jumps at t_k, $k = 1, \ldots, m$.

The following theorem summarizes the asymptotic properties of the estimators $(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}(\cdot))$. The proof is outlined in Web Appendix A.

Theorem 1 *Under Conditions* (A1)-(A3) *and* (B1)-(B4) *given in the Appendix,* $(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}(t))$ *converges almost surely to* $(\beta, \gamma, \Lambda(t))$ *uniformly in* $t \in [\zeta, \tau]$, *and*

$\sqrt{n}\big(\hat{\beta}-\beta, \hat{\gamma}-\gamma, \hat{\Lambda}(t)-\Lambda(t)\big)$ *converges in distribution to a mean zero Gaussian process for* $t \in [\zeta, \tau]$.

3.3 Variance Estimation

The proposed estimator $(\hat{\beta}, \hat{\gamma})$ for model (1) is the profile likelihood estimator by profiling out the baseline Λ with the plugging in of $\hat{\theta}_W$ for θ_W. Define the profile log likelihood

$$\mathrm{pl}_n(\beta^*;\theta_W) = \max_{\Lambda\in\mathcal{C}} \log L_n(\beta, \gamma, \Lambda;\theta_W),$$

where $\beta^* = (\beta, \gamma^T)^T$, $L_n(\beta, \gamma, \Lambda;\theta_W)$ is given in (7) and $\mathcal{C}$ is the set of step functions with nonnegative jumps at t_k, $k = 1, \ldots, m$. Then $\hat{\beta}^* = \mathrm{argmax}_{\beta^*\in\mathcal{B}}\, \mathrm{pl}_n(\beta^*;\hat{\theta}_W)$. When θ_W is known, the profile likelihood approach can be used to estimate the covariance matrix of $\hat{\beta}$ [28]. With the plug-in estimator $\hat{\theta}_W$, the estimator of the variance of $\hat{\beta}^* = (\hat{\beta}, \hat{\gamma}^T)^T$ needs to account for the variation of $\hat{\theta}_W$.

Let $U(\beta^*;\hat{\theta}_W) = \frac{\partial}{\partial\beta^*}\mathrm{pl}_n(\beta^*;\hat{\theta}_W)$. Then $U(\hat{\beta}^*;\hat{\theta}_W) = 0$. By (1) in the proof of Theorem 1 in the Web Appendix A, we have

$$\hat{\beta}^* - \beta^* = -\Big(\frac{\partial U(\beta^*;\theta_W)}{\partial\beta^*}\Big)^{-1}\Big[U(\beta^*;\theta_W) + \frac{\partial U(\beta^*;\theta_W)}{\partial\theta_W}(\hat{\theta}_W - \theta_W)\Big] + o_p(n^{-1/2}). \tag{15}$$

Under the measurement error model (2), the estimator $\hat{\theta}_W$ admits the approximation $\hat{\theta}_W - \theta_W = J^{-1}\sum_{i=1}^n \xi_i + o_p(n^{-1/2})$, where ξ_i are iid random vectors with mean zero and J is a positive definite matrix. Under Conditions (A1)-(A3) given in the Appendix, $U(\beta^*;\theta_W)$ and $\hat{\theta}_W - \theta_W$ are uncorrelated. Therefore, the two summands in (15) are asymptotically independent.

The covariance matrix of $\hat{\beta}^*$ equals

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}^*) = &\Big(\frac{\partial U(\beta^*;\theta_W)}{\partial\beta^*}\Big)^{-1} + \Big(\frac{\partial U(\beta^*;\theta_W)}{\partial\beta^*}\Big)^{-1}\frac{\partial U(\beta^*;\theta_W)}{\partial\theta_W}\mathrm{Cov}\\ &(\hat{\theta}_W)\Big(\frac{\partial U(\beta^*;\theta_W)}{\partial\theta_W}\Big)^{T}\Big(\frac{\partial U(\beta^*;\theta_W)}{\partial\beta^*}\Big)^{-1} + o_p(n^{-1}).\end{aligned} \tag{16}$$

Thus $\mathrm{Cov}(\hat{\beta}^*)$ can be consistently estimated by replacing β^* with $\hat{\beta}^*$, θ_W with $\hat{\theta}_W$ and $\mathrm{Cov}(\hat{\theta}_W)$ with its estimator $\widehat{\mathrm{Cov}}(\hat{\theta}_W)$. The details of derivations for the variance estimation are given in Web Appendix A.

The $(j,k)^{\text{th}}$ element of matrix $\frac{\partial U(\hat{\beta}^*;\hat{\theta}_W)}{\partial\beta^*}$ is estimated by

$$\frac{\mathrm{pl}_n(\hat{\beta}^*;\hat{\theta}_W) - \mathrm{pl}_n(\hat{\beta}^* + h_n e_k;\hat{\theta}_W) - \mathrm{pl}_n(\hat{\beta}^* + h_n e_j;\hat{\theta}_W) + \mathrm{pl}_n(\hat{\beta}^* + h_n e_k + h_n e_j;\hat{\theta}_W)}{h_n^2},$$

where e_j and e_k are the j^{th} and k^{th} canonical vector in R^{d+1}, respectively, and h_n is at the order of $n^{-1/2}$.

Similarly, the $(j,k)^{\text{th}}$ element of matrix $\frac{\partial U(\beta^*;\theta_W)}{\partial \theta_W}$ is estimated by

$$\frac{\text{pl}_n(\hat{\beta}^*;\hat{\theta}_W) - \text{pl}_n(\hat{\beta}^* + h_n e_k;\hat{\theta}_W) - \text{pl}_n(\hat{\beta}^*;\hat{\theta}_W + h_n u_j) + \text{pl}_n(\hat{\beta}^* + h_n e_k;\hat{\theta}_W + h_n u_j)}{h_n^2},$$

where e_k is the k^{th} canonical vector in R^{d+1} and u_j is the j^{th} canonical vector in R^q with q the dimension of θ_W.

The R package *merDeriv* developed by [29] for generalized linear mixed models can be used to estimate $\text{Cov}(\hat{\theta}_W)$ [27].

To calculate $\text{pl}_n(\beta^*;\theta_W)$, we apply the proposed EM algorithm with β^* and θ_W held fixed. For any given values of β^* and θ_W, the procedure iterates between (12) for λ_k and (14) for $\hat{E}(\eta_{ik})$. For fast convergence, one can take the estimate $\hat{\lambda}_k$ of the jump size of of the cumulative baseline function $\Lambda(\cdot)$ for model (5) as the initial value. The step size h_n in calculating the second order differences can be taken as $h_n = Cn^{-1/2}$, where C is a constant that can be calibrated depending on data applications. Although there has been no existing study examining the optimal choice of h_n, our simulation studies show that $h_n = 5n^{-1/2}$ works well.

We summarize the steps for implementing the proposed method as follows:

1. Obtain the the maximum likelihood estimates $\hat{\theta}_W$ of the parameters $\theta_W = (\vartheta, G, \sigma^2)$ under the measurement error model (2).
2. Calculate the estimated longitudinal covariates $\hat{X}_i(t) = f^T(t)\hat{\theta}_i(t)$, where $\hat{\theta}_i(t)$ is the least squares estimator of θ_i given by (3), $i = 1, \ldots, n$.
3. Estimate the parameters (β, γ, Λ) in the measurement error induced hazard model (5) using the EM algorithm described in Section 3.2, where θ_W is replaced by $\hat{\theta}_W$.
4. Estimate the covariance matrix of $\hat{\beta}^* = (\hat{\beta}, \hat{\gamma})$ using $\text{Cov}(\hat{\beta}^*)$ given by (16).

4 A Diagnostic Testing Procedure for the Measurement Error Model

This section presents a diagnostic procedure to examine validity of the measurement error model (2). An invalid model can introduce additional bias and diminish the benefits of dealing with the measurement errors. The proposed test procedure provides a formal procedure to check for the model assumptions for the longitudinal covariate.

For each individual i, let $\hat{e}_{ij} = W_{ij} - f(v_{ij})^T\hat{\theta}_i\mathbf{1}_{B_{ij}}^T$ and $e_{ij} = W_{ij} - \theta_i^T f(v_{ij})\mathbf{1}_{B_{ij}}^T$, where $\hat{\theta}_i = \hat{\theta}_i(\tau)$. The regression residual process is defined as $\hat{\epsilon}_i(v_{ij}) = \hat{e}_{ij}\mathbf{1}_{B_{ij}}$. Let $\hat{\sigma}$ be the estimator of σ given in (4) under model (2). We introduce the following weighted residual process for individual i,

$$H_{iM_i(t)} = \sum_{j=1}^{M_i(t)} \frac{B_{ij}^{-1/2}\hat{\epsilon}_i(v_{ij})}{\hat{\sigma}\Big(1 - B_{ij}f^T(v_{ij})\{\tilde{F}_i^T(\tau)\tilde{F}_i(\tau)\}^{-1}f(v_{ij})\Big)^{1/2}}.$$

In the following we construct the test based on the differences of the weighted residual processes. Let $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_K \le \tau$ be the grid points on $[0, \tau]$. We set $M_i(0) = 0$ and $H_{iM_i(0)} = 0$. Define $D_{ik} = H_{iM_i(\tau_k)} - H_{iM_i(\tau_{k-1})}$ for $1 \le k \le K$, and $H_n = \sum_{i=1}^n \left(D_{i1}, D_{i2}, \ldots, D_{iK}\right)^T$. We propose the test statistic

$$Q = H_n^T \Sigma_H^{-1} H_n, \tag{17}$$

where Σ_H is the covariance matrix of H_n. The diagonal of Σ_H includes

$$\sum_{i=1}^n \mathrm{Var}(D_{ik}) = \sum_{i=1}^n \left(M_i(\tau_k) - M_i(\tau_{k-1})\right) - 2\sum_{i=1}^n \sum_{M_i(\tau_{k-1}) < l < m \le M_i(\tau_k)} \Psi_{i,lm},$$

for $1 \le k \le K$, and the off-diagonal elements of Σ_H are given by

$$\sum_{i=1}^n \mathrm{Cov}(D_{ij}, D_{ik}) = -\sum_{i=1}^n \sum_{M_i(\tau_{j-1}) < l \le M_i(\tau_j)} \sum_{M_i(\tau_{k-1}) < m \le M_i(\tau_k)} \Psi_{i,lm},$$

for $j \ne k$. Here

$$\Psi_{i,jk} = \frac{B_{ij}^{-1/2} B_{ik}^{-1/2} f^T(v_{ij}) \{\tilde{F}_i^T(\tau)\tilde{F}_i(\tau)\}^{-1} f(v_{ik})}{\left(1 - B_{ij} f^T(v_{ij}) \{\tilde{F}_i^T(\tau)\tilde{F}_i(\tau)\}^{-1} f(v_{ij})\right)^{1/2} \left(1 - B_{ik} f^T(v_{ik}) \{\tilde{F}_i^T(\tau)\tilde{F}_i(\tau)\}^{-1} f(v_{ik})\right)^{1/2}}.$$

Theorem 2 *Under model* (2)*, the test statistic* Q *has a chi-square distribution with* K *degrees of freedom.*

By Theorem 2*, the test rejects model* (2) *at significance level* α *if* $Q > \chi^2_{K,1-\alpha}$*. The proof of Theorem* 2 *is given in the Web Appendix A.*

It is easy to show that the test statistic Q *has an asymptotic chi-square distribution with* K *degrees of freedom as long as the random effects* ν_i *and the measurement errors* $e_{ij,b}$ *are iid with mean zero and finite variances. The proposed test provides a method to test the form of the within-individual patterns defined by the basis function* $f(t)$*. It does not test the normality assumptions of* θ_i *and the errors* e_{ij}*. Many existing tests such as the Kolmogorov-Smirnov test, Shapiro-Wilk test, and Anderson-Darling test can be used to test for normality. Testing of the normality assumption of* θ_i *can be conducted based on* $\{\hat{\theta}_i(\tau), i = 1, \ldots, n\}$*, while testing of the normality assumption of* e_{ij} *can be conducted based on* $\{\hat{e}_{ij}, j = 1, \ldots, M_i, i = 1, \ldots, n\}$*. Diagnostic tools such as Q-Q plots can be used to compliment the formal test procedures for real data applications.*

The proposed test is not overly sensitive to the choice of K*. We suggest* $3 \le K \le 8$ *and that the grid points* $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_K \le \tau$ *be evenly spaced in* $[0, \tau]$*. Our simulation results show that the test performs well.*

5 Simulation Studies

We evaluate the proposed method via simulation studies. Let n be the sample size. For $i = 1, \ldots, n$, the failure time T_i is generated from the proportional hazards model

$$\lambda(t|\bar{X}_i(t), Z_i) = \lambda(t)\exp\{\beta X_i(t) + \gamma Z_i\}, \tag{18}$$

where $\lambda(t) = 1/(2+t)$, $\beta = 0.5$, $\gamma = -\log(2)$, $Z_i \sim Ber(0.3)$, and $X_i(t)$ has the form $X_i(t) = (\nu_0 + b_{0i}) + (\nu_1 + b_{1i})t$, with $\nu_0 = 1$, $\nu_1 = 0.5$, $(b_{0i}, b_{1i}) \sim N(0, G)$, and $G = [0.02, -0.01; -0.01, 0.02]$. Let $Unif(0, a)$ denote a uniform random variable on $(0, a)$ and $Ber(p)$ a Bernoulli random variable with success probability p. We simulate the measurement times $(v_{i1}, \ldots, v_{i,M_i})$ for $X_i(\cdot)$ as follows. We first generate the measurement times as the cumulative sums of independent $Unif(0, 0.2)$ random variates until $\tau/6$ is reached, and then keep adding up independent $Unif(0, 0.4)$ random variates until τ. We simulate partly interval censored data for individual i as follows. We first generate the number of monitoring times $K_i \sim Ber(0.8) + 1$. If $K_i = 1$, we generate one monitoring time $U_{i1} \sim Unif(0, \tau/2)$; define $(L_i, R_i] = (0, U_{i1}]$ if $T_i \le U_{i1}$ and $(L_i, R_i] = (U_{i1}, \infty)$ if $T_i > U_{i1}$. If $K_i = 2$, we generate two monitoring times $U_{i1} \sim Unif(0, \tau/2)$ and $U_{i2} \sim \min\{0.1 + U_{i1} + Unif(0, 3\tau/4), \tau\}$; define $(L_i, R_i] = (0, U_{i1}]$ if $T_i \le U_{i1}$, $(L_i, R_i] = (U_{i1}, U_{i2}]$ if $U_{i1} < T_i \le U_{i2}$, and $(L_i, R_i] = (U_{i2}, \infty)$ if $T_i > U_{i2}$. If $R_i = \infty$, we set $\Delta_i = 0$; if $R_i < \infty$, we generate $\Delta_i \sim Ber(p)$ with $p = 0.25$ or 0.75. If $\Delta_i = 1$, the failure time T_i is exactly observed. The length of study is taken to be $\tau = 3$ yielding about 40% right censoring. The error-prone measurements $W_{i,b}(v_{ij})$ are generated from the model

$$W_{i,b}(v_{ij}) = X_i(v_{ij}) + e_{ij,b}, \quad b = 1, \ldots, B_{ij}, \tag{19}$$

where $X_i(v_{ij}) = (\nu_0 + b_{0i}) + (\nu_1 + b_{1i})v_{ij}$ is specified above, $e_{ij,b} \sim N(0, \sigma^2)$ with $\sigma = 0.1$ or 0.2 and the number of repeated measurements of $X_i(v_{ij})$ is $B_{ij} = B = 1$ or 3 for all i, j.

We compare four methods: (i) the proposed method; (ii) the ideal method using true $X(t)$ which is not available in practice; (iii) the naive method that ignores measurement error and uses $W(t)$ directly, where $W(t)$ at any time t is evaluated via last value carried forward from the longitudinal measurements (the average is used if there are replicates for $W(t)$); (iv) the naive method using $\hat{X}(t)$ by simply replacing $X(t)$ with $\hat{X}(t)$ in the proportional hazards model. For the variance estimation based on the profile likelihood method, we take $h_n = 5n^{-1/2}$. The results are similar with other choices of h_n, such as $n^{-1/2}$ and $10n^{-1/2}$, which is also noted in [9]. We consider the sample size $n = 400$ and 600. The estimation results for (β, γ) based on 500 simulations are presented in Table 1 for $B = 1$ and in Table S1 of Web Appendix B for $B = 3$, where Bias is the average point estimate minus the true parameter value, SSD is the sample standard deviation of point estimates, ESE is the average of estimated standard errors and CP is the coverage proportion of the 95% confidence interval.

We can see from Tables 1 and S1 that (i) for all scenarios considered, the proposed method yields unbiased estimates with reasonable estimated standard errors and coverage proportions; (ii) the sample standard deviation of the

Table 1 Simulation results for (β, γ) under models (18) and (19) when there are no repeated measurements of $X_i(t)$, i.e., $B = 1$. The random effects $(b_{0i}, b_{1i}) \sim N(0, G)$ with $G = [0.02, -0.01; -0.01, 0.02]$. Each entry is based on 500 replicates

			$p = 0.25$							
			$\beta = 0.5$				$\gamma = -\log(2)$			
n	σ	Method	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
400	0.1	Proposed	0.045	0.554	0.565	0.962	−0.017	0.166	0.154	0.922
		$X(t)$	0.010	0.443	0.453	0.970	−0.016	0.162	0.152	0.930
		$W(t)$	−0.072	0.417	0.424	0.958	−0.015	0.162	0.152	0.928
		$\hat{X}(t)$	−0.356	0.248	0.189	0.433	−0.016	0.167	0.155	0.918
	0.2	Proposed	0.077	0.703	0.719	0.958	−0.017	0.166	0.154	0.920
		$X(t)$	0.010	0.443	0.453	0.970	−0.016	0.162	0.152	0.930
		$W(t)$	−0.216	0.342	0.347	0.908	−0.014	0.161	0.152	0.930
		$\hat{X}(t)$	−0.440	0.150	0.112	0.150	−0.014	0.168	0.154	0.914
600	0.1	Proposed	0.040	0.455	0.452	0.950	−0.001	0.130	0.126	0.944
		$X(t)$	−0.010	0.375	0.365	0.946	0.001	0.129	0.124	0.938
		$W(t)$	−0.082	0.352	0.341	0.936	0.001	0.129	0.124	0.932
		$\hat{X}(t)$	−0.393	0.183	0.117	0.245	0.002	0.131	0.126	0.940
	0.2	Proposed	0.073	0.569	0.572	0.956	−0.000	0.130	0.126	0.944
		$X(t)$	−0.010	0.375	0.365	0.946	0.001	0.129	0.124	0.938
		$W(t)$	−0.215	0.290	0.280	0.866	0.002	0.128	0.124	0.932
		$\hat{X}(t)$	−0.457	0.110	0.067	0.053	0.002	0.130	0.126	0.942
			$p = 0.75$							
			$\beta = 0.5$				$\gamma = -\log(2)$			
n	σ	Method	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
400	0.1	Proposed	0.022	0.535	0.537	0.954	−0.017	0.167	0.156	0.934
		$X(t)$	0.017	0.430	0.437	0.960	−0.014	0.159	0.150	0.936
		$W(t)$	−0.128	0.380	0.377	0.942	−0.013	0.159	0.150	0.938
		$\hat{X}(t)$	−0.405	0.197	0.125	0.267	−0.015	0.166	0.156	0.933
	0.2	Proposed	0.030	0.652	0.662	0.948	−0.016	0.167	0.156	0.934
		$X(t)$	0.017	0.430	0.437	0.960	−0.014	0.159	0.150	0.936
		$W(t)$	−0.296	0.289	0.282	0.804	−0.012	0.158	0.150	0.938
		$\hat{X}(t)$	−0.466	0.111	0.068	0.049	−0.014	0.167	0.157	0.932
600	0.1	Proposed	−0.004	0.423	0.432	0.952	0.002	0.132	0.128	0.938
		$X(t)$	−0.011	0.359	0.354	0.950	0.003	0.126	0.123	0.944
		$W(t)$	−0.146	0.310	0.306	0.912	0.004	0.126	0.123	0.944
		$\hat{X}(t)$	−0.437	0.136	0.081	0.107	0.004	0.131	0.128	0.936
	0.2	Proposed	0.004	0.517	0.532	0.958	0.003	0.132	0.128	0.944
		$X(t)$	−0.011	0.359	0.354	0.950	0.003	0.126	0.123	0.944
		$W(t)$	−0.304	0.236	0.229	0.725	0.005	0.125	0.123	0.940
		$\hat{X}(t)$	−0.477	0.075	0.042	0.013	0.007	0.130	0.128	0.940

proposed estimator of β decreases when the degree of measurement error represented by σ decreases and when the sample size n, the number of repeated measurements B of $X(t)$, or the proportion of exact observations p increases; (iii) as expected, the ideal method that uses true $X(t)$ is more efficient than the proposed method and the efficiency gain increases with the degree of measurement error given by σ; (iv) the naive method that ignores measurement error and uses $W(t)$ directly yields acceptable results when $\sigma = 0.1$, but has large bias when $\sigma = 0.2$; particularly, it tends to underestimate β; (v) the naive method that replaces $X(t)$ with $\hat{X}(t)$ in the proportional hazards model gives severely biased estimates of β for all scenarios considered (and it should be noted that this method underestimates β due to the attenuating effect); and (vi) all methods perform well for the estimation of γ.

We have also investigated the computational cost of the proposed method. For the simulation setup $n = 400$, $p = 0.75$, $\sigma = 0.1$ and $B = 1$, it takes about 409 seconds (132 for parameter estimation and 277 for variance estimation) to implement the proposed method on a MacBook Pro (3.1GHz Quad-Core Intel Core i7).

To evaluate the robustness of our method to the normality assumption on random effects in the measurement error model, we generate the random effects b_{0i} and b_{1i} from $Unif(-0.25, 0.25)$ independently while keeping the other settings the same as in Tables 1 and S1. The results are presented in Table 2 for $B = 1$ and Table S2 of Web Appendix B for $B = 3$. One can see that the proposed method performs well in such situations.

In addition, we obtain the estimate of the baseline hazard function $\lambda(t)$ using kernel smoothing with the Gaussian kernel and bandwidth 0.1. Figures $S1 \sim S4$ in Web Appendix B plot the estimated baseline hazard functions based on the simulation results of Tables 1, 2, S1 and S2, respectively. One can see that the proposed method and the ideal method yield unbiased estimates of the baseline hazard function $\lambda(t)$ except for t close to 0, while the naive methods yield biased estimates.

We also conduct a simulation study to examine the empirical sizes and powers of the proposed test for $X_i(v_{ij}) = (\vartheta + v_i)^T f(v_{ij})$ under the measurement error model (2), for $i = 1, \ldots, n$, $j = 1, \ldots, M_i$ and $b = 1, \ldots, B$, where $f(t)$ is an $r \times 1$ vector of basis functions, ϑ is a vector of fixed parameters, and $v_i (i = 1, \ldots, n)$ are iid $N(0, G)$. We set the null model to be $f(t) = (1, t)$ and generate data from the following four models:

I: $f(t) = (1, t)$, $\vartheta = (1, 0.5)$ and $G = [0.02, -0.01; -0.01, 0.02]$
II: $f(t) = (1, t, t^2)$, $\vartheta = (1, 0.5, 0.01)$ and $G = [0.02, -0.01, 0; -0.01, 0.02, 0; 0, 0, 0.02]$
III: $f(t) = (1, t, t^2)$, $\vartheta = (1, 0.5, 0.02)$ and $G = [0.02, -0.01, 0; -0.01, 0.02, 0; 0, 0, 0.02]$
IV: $f(t) = (1, t, t^2)$, $\vartheta = (1, 0.5, 0.02)$ and $G = [0.02, -0.01, -0.01; -0.01, 0.02, -0.01; -0.01, -0.01, 0.02]$

Here for easy presentation G is the covariance matrix with rows separated by semicolons.

Table 2 Simulation results for (β, γ) under models (18) and (19) when there are no repeated measurements of $X_i(t)$, i.e., $B = 1$. The random effects b_{0i} and b_{1i} are independent $Unif(-0.25, 0.25)$. Each entry is based on 500 replicates

			$p = 0.25$							
			$\beta = 0.5$				$\gamma = -\log(2)$			
n	σ	Method	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
400	0.1	Proposed	0.055	0.393	0.392	0.936	−0.007	0.168	0.154	0.938
		$X(t)$	0.015	0.340	0.341	0.938	−0.007	0.164	0.152	0.930
		$W(t)$	−0.021	0.324	0.335	0.960	−0.007	0.164	0.152	0.930
		$\hat{X}(t)$	−0.309	0.230	0.164	0.453	−0.006	0.165	0.154	0.944
	0.2	Proposed	0.093	0.472	0.470	0.936	−0.008	0.168	0.154	0.942
		$X(t)$	0.015	0.340	0.341	0.938	−0.007	0.164	0.152	0.930
		$W(t)$	−0.131	0.284	0.294	0.944	−0.006	0.164	0.152	0.934
		$\hat{X}(t)$	−0.413	0.150	0.108	0.178	−0.006	0.166	0.154	0.944
600	0.1	Proposed	0.045	0.296	0.316	0.958	−0.014	0.135	0.126	0.932
		$X(t)$	0.011	0.261	0.276	0.970	−0.013	0.135	0.124	0.938
		$W(t)$	−0.030	0.245	0.271	0.968	−0.012	0.135	0.124	0.936
		$\hat{X}(t)$	−0.352	0.173	0.108	0.260	−0.011	0.133	0.126	0.929
	0.2	Proposed	0.079	0.354	0.378	0.958	−0.013	0.135	0.126	0.930
		$X(t)$	0.011	0.261	0.276	0.970	−0.013	0.135	0.124	0.938
		$W(t)$	−0.142	0.210	0.237	0.942	−0.011	0.134	0.124	0.932
		$\hat{X}(t)$	−0.432	0.106	0.063	0.049	−0.009	0.132	0.126	0.929
			$p = 0.75$							
			$\beta = 0.5$				$\gamma = -\log(2)$			
n	σ	Method	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
400	0.1	Proposed	0.031	0.381	0.380	0.940	−0.004	0.170	0.156	0.936
		$X(t)$	0.017	0.327	0.334	0.952	−0.004	0.161	0.150	0.936
		$W(t)$	−0.074	0.291	0.311	0.962	−0.004	0.161	0.150	0.940
		$\hat{X}(t)$	−0.356	0.199	0.121	0.318	−0.000	0.169	0.156	0.933
	0.2	Proposed	0.038	0.446	0.443	0.940	−0.004	0.170	0.156	0.936
		$X(t)$	0.017	0.327	0.334	0.952	−0.004	0.161	0.150	0.936
		$W(t)$	−0.224	0.234	0.250	0.888	−0.003	0.161	0.150	0.936
		$\hat{X}(t)$	−0.441	0.116	0.069	0.065	0.000	0.169	0.156	0.928
600	0.1	Proposed	0.018	0.288	0.307	0.964	−0.012	0.135	0.128	0.936
		$X(t)$	0.008	0.257	0.271	0.960	−0.011	0.135	0.123	0.932
		$W(t)$	−0.080	0.231	0.252	0.954	−0.010	0.134	0.123	0.936
		$\hat{X}(t)$	−0.398	0.142	0.075	0.136	−0.009	0.133	0.128	0.944
	0.2	Proposed	0.028	0.338	0.358	0.956	−0.011	0.136	0.128	0.938
		$X(t)$	0.009	0.257	0.271	0.960	−0.011	0.135	0.123	0.932
		$W(t)$	−0.227	0.185	0.203	0.830	−0.009	0.134	0.123	0.932
		$\hat{X}(t)$	−0.461	0.077	0.041	0.006	−0.009	0.133	0.128	0.943

We simulate the measurement times $(v_{i1}, \ldots, v_{i,M_i})$ for $X(\cdot)$ as in Tables 1 and S1. Specifically, we first generate the measurement times as the cumulative sums of $Unif(0, 0.2)$ until $\tau/6$ is reached, and then keep adding up $Unif(0, 0.4)$ variates until τ. We set the length of follow-up to be $\tau = 3$ and take equally spaced grid points in $[0, \tau]$ with $K = 1, 3, 5, 7$. We conduct the test and compute the p-value for each simulated dataset. We calculate the empirical size for Model I and the power for Model II, III and IV as the proportion of p-values ≤ 0.05 for 500 simulated datasets. The results are presented in Table 3 for $B = 1$ and in Table S3 of the Web Appendix B for $B = 3$. The empirical sizes under Model I are around the 0.05 nominal level for all cases and remain the same for different values of σ. The powers under Model II, III and IV are fairly high given the small effect sizes 0.01 and 0.02. The power increases with the number of repeated measurements B, the sample size n and the effect size. Moreover, the power decreases when σ increases, and seems to be similar when using different numbers of grid points K in the test.

6 Application to ACTG 175

We apply the proposed method to the ACTG 175 trial, a randomized, double-blind phase II/III trial of antiretroviral regimens in persons living with HIV infection with CD4 cell count from 200 to 500 per cubic millimeter [1]. Between December 1991 and October 1992, 2467 individuals were recruited and followed until November

Table 3 Simulation results for the proposed test of the measurement error model (2) with $X_i(t) = (\vartheta + v_i)^T f(t)$ at significance level 0.05, when there are no repeated measurements of $X_i(t)$, i.e., $B = 1$. Each entry is based on 500 replicates

		Model I (size)				Model II (power)			
n	σ	K=1	K=3	K=5	K=7	K=1	K=3	K=5	K=7
200	0.05	0.040	0.062	0.054	0.050	0.746	0.676	0.668	0.666
	0.1	0.040	0.062	0.054	0.050	0.682	0.608	0.582	0.572
	0.2	0.040	0.062	0.054	0.050	0.520	0.432	0.390	0.368
400	0.05	0.050	0.048	0.066	0.042	0.832	0.764	0.756	0.760
	0.1	0.050	0.048	0.066	0.042	0.778	0.718	0.698	0.680
	0.2	0.050	0.048	0.066	0.042	0.638	0.568	0.508	0.488
		Model III (power)				Model IV (power)			
n	σ	K=1	K=3	K=5	K=7	K=1	K=3	K=5	K=7
200	0.05	0.930	0.890	0.878	0.876	0.932	0.902	0.902	0.904
	0.1	0.904	0.866	0.838	0.844	0.904	0.880	0.858	0.870
	0.2	0.798	0.746	0.686	0.648	0.822	0.740	0.688	0.664
400	0.05	0.984	0.974	0.974	0.964	0.984	0.982	0.976	0.974
	0.1	0.976	0.962	0.956	0.946	0.976	0.968	0.960	0.962
	0.2	0.938	0.898	0.890	0.868	0.950	0.906	0.898	0.884

1994. Among these, 1396 participants received antiretroviral therapy (ART) prior to the study while 1061 participants were ART naive. The objective of the trial was to compare the effectiveness of four antiretroviral regimens (zidovudine only, zidovudine + didanosine, zidovudine + zalcitabine, and didanosine only) in preventing disease progression to AIDS or death. An important prognostic biomarker for progression to the clinical endpoint is CD4 cell count per cubic millimiter of blood [30]. All ACTG 175 trial participants had CD4 cell count measured every 12 weeks starting at Week 8, and were followed for occurrence of the composite clinical endpoint of AIDS or death. The median number of measurement times is 12 with interquartile range (IQR) [8,14], and its histogram is given by Figure S8 in the Web Appendix.

The original analysis of [1] found zidovudine alone to be inferior to the other three therapies. Following [17] and [20], we consider two treatment groups, zidovudine alone and the combination of the other three therapies. We demonstrate the utility of the proposed method by investigating associations of treatment arm and time-dependent trajectory $\log_{10}$(CD4) with the composite clinical endpoint of AIDS or death. Let Z be the treatment indicator (TRT) with value 0 for zidovudine alone and 1 for other three regimens. Let $X(\cdot)$ be the error-prone time-dependent covariate $\log_{10}$(CD4) (measured without replicates, i.e., $B = 1$) and T be the time from enrollment to AIDS or death, whichever occurred first. We assume that the conditional hazard function of T given $\bar{X}(t)$ and Z follows the proportional hazards model (1),

$$\lambda(t|\bar{X}(t), Z) = \lambda(t) \exp\{\beta X(t) + \gamma Z\} = \lambda(t) \exp\left\{\beta \log_{10}(\text{CD4}) + \gamma \text{TRT}\right\},$$

where the regression coefficients β and γ can be interpreted as log hazard ratios and represent the association of time to AIDS or death with $\log_{10}$(CD4) and TRT, respectively. Also, we assume that the measurement error model for $X(t)$ is (2) with the quadratic basis function $f(t) = (1, t, t^2)$. Our analysis includes 1396 participants who received ART prior to the study. There were 215 composite endpoint cases (15.4%) with 167 AIDS events and 48 deaths. The time to the AIDS onset is interval censored while the time to death prior to AIDS is observed or right-censored. The observed data consist of exact, interval- and right-censored event times.

The true CD4 cell count values $X(t)$ are generally not attainable. The observed CD4 cell count is measured intermittently and is an error-prone time-dependent covariate. The naive approaches often replace $X(t)$ in model (1) with the observed $W(t)$ by last value carried forward that ignores the measurement errors or with a model-based estimate $\hat{X}(t)$ without modifying the hazard model for the induced error. The former naive approach – termed the "naive approach using $W(t)$" – imputes the CD4 values at each failure time using "last value carried forward" that substitutes the unavailable CD4 cell count with the last observed value prior to the failure time. The latter naive approach – termed the "naive approach using $\hat{X}(t)$" – replaces CD4 cell count values with the estimated $\hat{X}(t)$ in model (1) using the measurement error model (2) based on each individual's longitudinal profile prior to time t.

Fitting the quadratic measurement error model (2), we obtain average values of the individual-specific estimates of coefficients of $(2.515, -0.035, -0.047)$. The plot of the average fitted individual-specific curves along with the plot of the observed

Table 4 Analysis results for ACTG 175

	Proposed Method			Naive Using $W(t)$			Naive Using $\hat{X}(t)$		
Covariates	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
$\log_{10}$(CD4)	−2.472	0.116	<0.001	−2.668	0.119	<0.001	−0.240	0.051	<0.001
Treatment	−0.119	0.160	0.458	0.056	0.155	0.718	−0.338	0.159	0.034

$\log_{10}$(CD4) for 50 randomly selected individuals shows the downward trend in $\log_{10}$(CD4) (Figure S5). The results of analysis using the proposed method and the two naive methods are summarized in Table 4, where Est is the estimates of the regression parameters and SE is the estimated standard errors. The estimated regression coefficient (the standard error) of $\log_{10}$(CD4) using the proposed method is −2.472 (0.116), whose absolute value is much larger than that of the estimated regression coefficient (the standard error) −0.240 (0.051) obtained using the naive approach using $\hat{X}(t)$ but slightly less than that of the estimated regression coefficient (the standard error) −2.668 (0.119) using the naive approach using $W(t)$. The estimated regression coefficient of $\log_{10}$(CD4) is the log hazards ratio for every unit increase in $\log_{10}$(CD4) under the Cox model (equivalently every 10-fold increase in CD4 cell count with units number of cells per cubic milimeter of blood, cells/mm^3) and represents the association of $\log_{10}$(CD4) with the failure time. All methods suggest that lower values of $\log_{10}$(CD4) are significantly associated with higher risk of AIDS or death. The naive approach using $\hat{X}(t)$ yields an estimated association closer to zero (−0.240) as compared to the proposed method (−2.472) partly due to the attenuating effect $\omega_i(t)$ from the measurement errors. The naive approach using $W(t)$ yields a slightly stronger inverse association (−2.668) as compared to the proposed method because it tends to carry forward a too-large value of CD4 cell count. While there is no significant treatment effect after adjusting for $\log_{10}$(CD4) with the proposed method and the naive approach using $W(t)$, the naive approach using $\hat{X}(t)$ shows a significant treatment effect with p-value 0.034. Our study further confirms that the naive approaches can lead to biased estimates of the associations of interest for the variables measured in errors as well as biased estimates of treatment effects [2, 3, 31].

Figure 1 plots the estimated survival functions at four different combinations of covariates: two values of Z (0 or 1) and two curves of $\log_{10}$(CD4) (25$^{\text{th}}$ or 75$^{\text{th}}$ percentile of the estimated $\log_{10}$(CD4) at each time point). The plots show that the naive approach using $W(t)$ overestimates the survival probabilities and the naive approach using $\hat{X}(t)$ substantially underestimates the survival probabilities. The discrepancy is very large when $X(t)$ is the 75$^{\text{th}}$ percentile of $\log_{10}$(CD4). Plots of the estimated baseline hazard functions for the three methods considered are given in Figure S7 in Web Appendix B. It can be seen that the naive approach using $W(t)$ overestimates the baseline hazard function, while the naive approach using $\hat{X}(t)$ highly underestimates the baseline hazard function.

To examine appropriateness of the quadratic measurement error model (2), we conduct the model checking procedure. In particular, we consider the quadratic

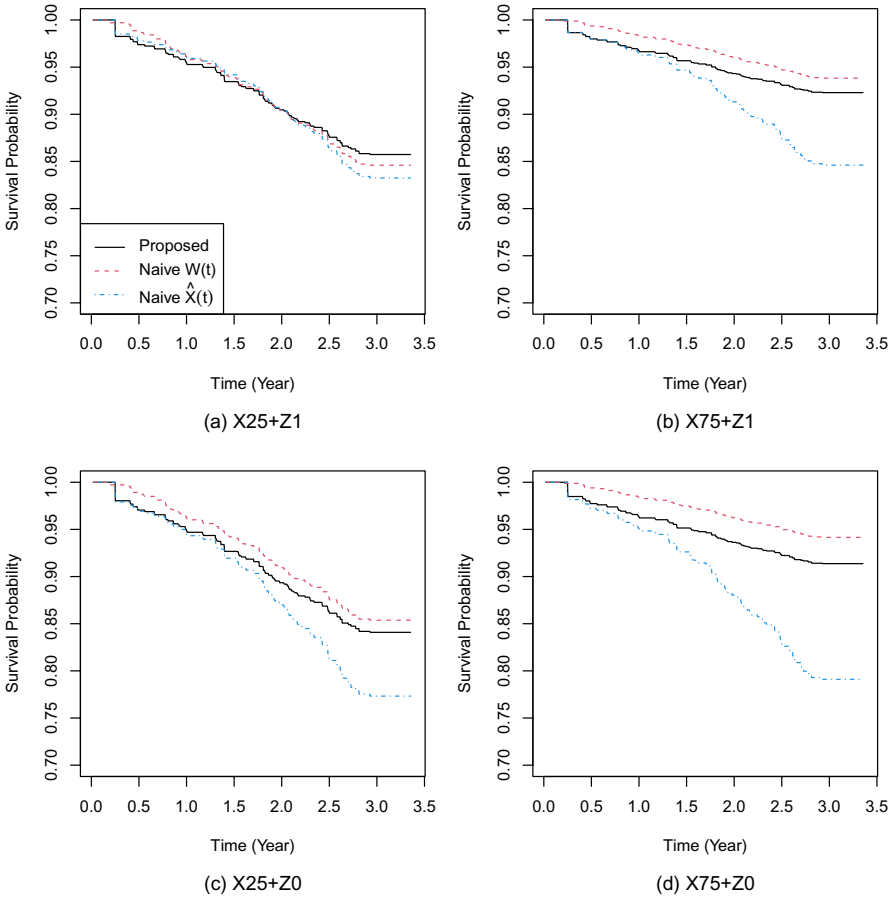

Fig. 1 Plots of the estimated survival functions at four different combinations of covariates. For example, 'X25+Z0' corresponds to the covariates combination with $X(t)$ being the 25$^{\text{th}}$ percentile of $\log_{10}$(CD4) and for zidovudine alone, and 'X25+Z1' corresponds to the covariates combination with $X(t)$ being the 25$^{\text{th}}$ percentile of $\log_{10}$(CD4) and for the other three treatment arms pooled

measurement error model corresponding to $f(t) = (1, t, t^2)$ in (2). We set the grid points to be the $(0, 25, 50, 75, 100)^{\text{th}}$ quantiles of the follow-up times. The quadratic measurement error model yields a p-value of 0.298 suggesting that the quadratic model fits reasonably well to the data. As a comparison, we also test fitness of the linear measurement error model with $f(t) = (1, t)$. The test yields a p-value close to zero. In addition, we note that the log-likelihood value at the final estimates of the proposed method under the linear measurement error model is −932.56, while the log-likelihood value under the quadratic model is −856.54. The analysis supports that the quadratic measurement error model fits the data better than the linear model.

We further examine the fit of the quadratic measurement error model via graphical tools. The residual plots for the quadratic measurement error model, including the normal Q-Q plot and histogram, are presented in Figure S6 of

Web Appendix B. The normal Q-Q plot suggests that the sample quantiles of the standardized residuals are close to the theoretical ones except for slight deviations at the two tails, while the histogram of the standardized residuals looks like a standard normal density curve except having slightly shorter tails. Furthermore, to evaluate the normality of random effects, we obtain the least-squares estimates of individual-specific coefficients in the quadratic measurement error model. The normal Q-Q plot and histogram of these estimated coefficients are given in Figure S6. These plots look satisfactory in general except for in the tails. The p-values from the Kolmogorov-Smirnov tests for normality are 0.0006 for the errors and < 0.0001 for the random effects of the quadratic measurement error model. The small p-values reflect lack of fit in the tail areas of the distributions though the normality assumptions seem reasonable overall based on the diagnostic plots. The very small p-values can also be a result of the large sample size. Nevertheless, as shown in the simulation studies, the proposed method seems to be robust to the normality assumption of the random effects.

7 Concluding Remarks

This article develops an estimation method for the Cox model based on partly interval censored failure time and a longitudinal covariate with measurement errors. The research is motivated by the ACTG 175 trial to understand the association of longitudinal CD4 cell count on the hazard of the composite clinical endpoint of AIDS or death, where the time to the composite endpoint is partly interval censored and the recorded values of CD4 cell count are error-prone measures of the unattainable true values. The proposed measurement error induced hazard approach is intuitively appealing and easy to interpret. The EM-algorithm is proposed to implement the maximum likelihood estimation with partly interval censored data. The developed method has broad applications. For example, COVID-19 vaccine efficacy trials will study longitudinal antibody biomarkers over time as correlates of the study endpoint acquisition of SARS-CoV-2 infection. This endpoint is a composite endpoint with the same structure as the AIDS/death composite endpoint, defined as the first event of asymptomatic SARS-CoV-2 infection measured by seroconversion from a blood sample (interval censored) and symptomatic virologically confirmed SARS-CoV-2 infection that is symptom-triggered and hence measured exactly [32].

This paper assumes that the measurement errors are independent identically distributed. In practice, the measurement errors may be correlated or the measurement error variance is heterogeneous over time. In this case, the likelihood method can be used to estimate model (2) by assuming a certain covariance structure for the measurement errors. This is an interesting scenario that is worth investigation in a future project. We have regarded $X_i(t)$ as a scalar. The method can be extended to multivariate longitudinal covariates. As with many works in the joint modeling framework, a limitation of the proposed method is the normality assumption of the random effects in the measurement error model for $X_i(t)$. A simulation study conducted to examine the robustness of the proposed method shows that the estimation bias remains small.

A formal test procedure is proposed to examine validity of the measurement error model.

7.1 Supplementary Information

Web Appendices A and B, referenced in Sections 3, 5 and 6, are included in the Supplementary Information and are available with this paper on the journal website.

Appendix: Technical Details

In this appendix, we present the regularity conditions needed for Propositions 1 and Theorem 1.

Let $\tilde{U}_i = (U_{i1}, \ldots, U_{i,K_i})$ denote the monitoring times for the failure event for individual i, where $0 = U_{i0} < U_{i1} < \cdots < U_{i,K_i} < U_{i,K_i+1} = \infty$. The monitoring times are the mechanism used to generate interval censored failure times $[L_i, R_i]$, where $L_i = \max\{U_{ik} : T_i > U_{ik}, k = 0, \ldots, K_i\}$ and $R_i = \min\{U_{ik} : T_i \le U_{ik}, k = 1, \ldots, K_i + 1\}$ with $U_{i0} = 0$ and $U_{i,K_i+1} = \infty$. We assume Conditions (A1)-(A3) below that require noninformative monitoring/measurement times and a nondifferential measurement error mechanism for the time-dependent covariates.

(A1) The monitoring times, measurement times and measurement errors are noninformative given the information already provided by Z_i and θ_i, i.e., T_i is independent of $(\Delta_i, \tilde{U}_i, \tilde{v}_i, \tilde{e}_i)$ given (Z_i, θ_i).
(A2) Measurement error $\tilde{e}_i$ is independent of $(T_i, \Delta_i, \tilde{U}_i, \tilde{v}_i, Z_i)$.
(A3) $X_i(t)$, $0 \le t \le \tau$, is a left continuous process.

More discussion on the assumptions of noninformative observation times and a nondifferential measurement error mechanism can be found in [21, 26].

The estimation of the induced hazard model (5) for partly interval censored data follows the EM procedure developed by [11]. In addition to Conditions (A1)-(A3), we assume the following regularity conditions similar to [11].

(B1) The true value of (β, γ) lies in the interior of a known compact set $\mathcal{B}$ in R^{d+1}, and the true value of $\Lambda(\cdot)$ is continuously differentiable with positive derivatives in $[\zeta, \tau]$, where $[\zeta, \tau]$ is the union of the supports of $\{\Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i\}$.
(B2) The vector of the basis functions $f(t)$ is left continuous and with bounded total variation over $[\zeta, \tau]$.
(B3) If $h(t) + \beta\omega_i(t)\hat{X}_i(t) + \gamma^T Z_i + O_i(\beta, t, \theta_W) = 0$ for all $t \in [\zeta, \tau]$ with a positive probability, then $h(t) = 0$ for $t \in [\zeta, \tau]$ and $(\beta, \gamma) = 0$.
(B4) $0 < P(\Delta_i = 0) < 1$, $P\big(L_i = \tau, R_i = \infty | \Delta_i = 0, \bar{X}_i(\tau), Z_i\big) \ge c_1$ and $P\big(R_i - L_i > c_2 | \Delta_i = 0, \bar{X}_i(\tau), Z_i\big) = 1$ for some positive constants c_1 and c_2. The conditional density of (L_i, R_i) given $(\bar{X}_i(\tau), Z_i)$, denoted by $g(u, v)$, has continuous second-order partial

derivatives with respect to u and v when $v - u > c_2$ and are continuously differentiable with respect to $(\bar{X}_i(\tau), Z_i)$.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12561-023-09372-y.

Acknowledgements The authors thank the reviewers for their constructive comments that have improved the paper. This research was partially supported by NIAID NIH award R37AI054165. Dr. Sun's research was also partially supported by National Science Foundation grant DMS1915829 and the Reassignment of Duties fund provided by the University of North Carolina at Charlotte. Dr. Zhou's research was partially supported by National Science Foundation grant DMS1916170. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M et al (1996) A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. N Engl J Med 335:1081–1090
2. Prentice RL (1982) Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69:331–342
3. Tsiatis AA, Davidian M (2001) A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika 88:447–458
4. Finkelstein DM (1986) A proportional hazards model for interval-censored failure time data. Biometrics 42:845–854
5. Huang J (1996) Efficient estimation for the proportional hazards model with interval censoring. Ann Stat 24:540–568
6. Huang J, Wellner JA (1997) Interval censored survival data: a review of recent progress. In proceedings of the first seattle symposium in biostatistics pp. 123–169 Springer
7. Zhang Y, Hua L, Huang J (2010) A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. Scand Stat Theory Appl 37:338–354
8. Wang L, McMahan CS, Hudgens MG, Qureshi ZP (2016) A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. Biometrics 72:222–231
9. Zeng D, Mao L, Lin D (2016) Maximum likelihood estimation for semiparametric transformation models with interval-censored data. Biometrika 103:253–271
10. Kim JS (2003) Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. J R Stat Soc Series B 65:489–502
11. Zhou Q, Sun Y, Gilbert PB (2021) Semiparametric regression analysis of partly interval-censored failure time data with application to an AIDS clinical trial. Stat Med 40:4376–4394
12. Nakamura T (1992) Proportional hazards model with covariates subject to measurement error. Biometrics 48:829–838
13. Hughes M (1993) Regression dilution in the proportional hazards model. Biometrics 49:1056–1066
14. Buzas J (1998) Unbiased scores in proportional hazards regression with covariate measurement error. J Stat Plan Inference 67:247–257
15. Huang Y, Wang CY (2000) Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. J Am Stat Assoc 95:1209–1219
16. Hu C, Lin DY (2002) Cox regression with covariate measurement error. Scand Stat Theory Appl 29:637–655

17. Song X, Davidian M, Tsiatis AA (2002) An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3(4):511–528
18. Song X, Huang Y (2005) On corrected score approach for proportional hazards model with covariate measurement error. Biometrics 61:702–714
19. Yi G, Lawless J (2007) A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. J Stat Plan Inference 137:1816–1828
20. Fu R, Gilbert PB (2017) Joint modeling of longitudinal and survival data with the cox model and two-phase sampling. Lifetime Data Anal 23:136–159
21. Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin 14:809–834
22. Song X, Ma S (2008) Multiple augmentation for interval-censored data with measurement error. Stat Med 27(16):3178–3190
23. Mandal S, Wang S, Sinha S (2019) Analysis of linear transformation models with covariate measurement error and interval censoring. Stat Med 38:4642–4655
24. Chen K, Jin Z, Ying Z (2002) Semiparametric analysis of transformation models with censored data. Biometrika 89:659–668
25. Wen CC, Chen YH (2014) Functional inference for interval-censored data in proportional odds model with covariate measurement error. Stat Sin 24:1301–1317
26. Yi G (2017) Statistical Analysis with Measurement Error or Misclassification. Springer, New York
27. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48
28. Murphy SA, Van der Vaart AW (2000) On profile likelihood. J Am Stat Assoc 95:449–465
29. Wang T, Merkle EC (2018) merDeriv: derivative computations for linear mixed effects models with application to robust standard errors. J Stat Softw 87(1):1–16
30. Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, Kingsley LA, Todd JA, Saah AJ, Detels R et al (1997) Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med 126(12):946–954
31. Dafni UG, Tsiatis AA (1998) Evaluating surrogate markers of clinical outcome when measured with error. Biometrics 54(4):1445–1462
32. Mehrotra DV, Janes HE, Fleming TR, Annunziato PW, Neuzil KM, Carpp LN, Benkeser D, Brown ER, Carone M, Cho I et al (2020) Clinical endpoints for evaluating efficacy in covid-19 vaccine trials. Ann Intern Med 22:M20-6169