Analysis of Linear Transformation Models with Covariate Measurement Error and Interval Censoring

Soutrik Mandal, Suojin Wang, and Samiran Sinha*
Department of Statistics, Texas A&M University, College Station, Texas 77843-3143
sinha@stat.tamu.edu*

Summary

Among several semiparametric models, the Cox proportional hazard model is widely used to assess the association between covariates and the time-to-event when the observed time-to-event is interval-censored. Often covariates are measured with error. To handle this covariate uncertainty in the Cox proportional hazard model with the interval-censored data flexible approaches have been proposed. To fill a gap and broaden the scope of statistical applications to analyze time-to-event data with different models, in this paper a general approach is proposed for fitting the semiparametric linear transformation model to interval-censored data when a covariate is measured with error. The semiparametric linear transformation model is a broad class of models that includes the proportional hazard model and the proportional odds model as special cases. The proposed method relies on a set of estimating equations to estimate the regression parameters and the infinite-dimensional parameter. For handling interval censoring and covariate measurement error, a flexible imputation technique is used. Finite sample performance of the proposed method is judged via simulation studies. Finally, the suggested method is applied to analyze a real data set from an AIDS clinical trial.

Key Words: Estimating equation; Linear transformation model; Martingale; Multiple imputation; Non-differential measurement error; Predictive density.

Running title: Transformation models with interval censoring and measurement error

1 Introduction

The Cox proportional hazard (CPH) model and the proportional odds (PO) model are routinely used as a time-to-event model for assessing the association between covariates and time-to-event. However, for a more flexible model fitting, here we consider a broader class of models. Suppose that given the covariates X and $\boldsymbol{Z}$, the time-to-event T follows the linear transformation model,

$$H(T) = -X\beta_1 - \boldsymbol{Z}^T\boldsymbol{\beta}_2 + e, \tag{1}$$

where H is an unknown monotone transformation function on $(0, \infty)$ with $H(T) \to -\infty$ as $T \to 0$, e is the error with a completely known distribution function and is independent of both covariates X and $\boldsymbol{Z}$. The linear transformation model (1) reduces to the CPH model and the PO model when e follows the extreme-value distribution and the logistic distribution, respectively. In this paper we propose a new method for estimating the regression parameters $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2^T)^T$ and the infinite-dimensional parameter H when T is interval-censored and X is subject to measurement error. This method is robust towards the distribution of X. In particular, we consider *case 2* interval censoring where instead of observing T, we observe a random interval. Along with the intervals, we also observe an indicator variable denoting whether the event occurs within the interval, or after the right end point of the interval.[1] Since X is subject to measurement error, X is not observed in the data. Instead, replicated measurements of a surrogate (proxy) variable are observed.

The motivation comes from an AIDS clinical trial data set. In order to test efficacy and compare four different drugs, HIV positive subjects were randomly assigned to one of the four drugs. On average, the subjects were followed over for 33 months, and the primary end point of the study was the occurrence of AIDS, death, or at least 50% drop in the CD4 counts from the baseline measurement. After the treatment initiation, subjects were supposed to be examined after 2, 4, and 8 weeks and then every 12 weeks thereafter.[2] Naturally, any event that happened between two examination times is interval-censored. Our interest is in modeling the time (in days) to occurrence

of the primary end point from the date treatment started. Since CD4 count is considered to be a marker for antiretroviral treatment responses and HIV disease progression, its baseline measurement is considered to be a covariate in our model. However, the actual CD4 cell count is difficult to measure. So multiple measurements before the treatment started are considered to be replicated observations on the erroneous surrogate variable for the true CD4 count. Analysis of these data is complex due to the presence of two sources of uncertainty – the time-to-event falls in an interval and instead of the actual CD4 counts, measurements on a surrogate variable are observed.

We first discuss the existing literature on the right-censored time-to-event data with covariate measurement error. There are two main approaches for handling covariate measurement error, namely, functional and structural approaches. In the functional paradigm the unobserved X is treated as unknown constant, while in the structural paradigm the unobserved X is treated as a stochastic variable and a probability distribution is assumed.[3] In the functional paradigm, Prentice,[4] Nakamura,[5] Huang and Wang[6] have developed several flexible approaches for handling right censoring and covariate measurement error in the CPH model. For the PO model, Sinha and Ma[7] proposed a functional approach to handle covariate measurement error. Cheng and Wang[8] and Sinha and Ma[9] proposed two different approaches for handling covariate measurement error in the linear transformation model. While Cheng and Wang[8] imposed partly parametric model assumptions on X and the measurement error U and required to estimate the censoring distribution, Sinha and Ma[9] required a parametric model only for the distribution of X.

For interval-censored data, Song and Ma[10] proposed a functional approach to handle covariate measurement error for the CPH model. They used a multiple imputation technique to impute the time-to-event that falls within an interval, and then analyzed the imputed data sets by using the conditional score approach for right-censored data.[11] Wen and Chen[12] proposed a functional inference procedure with interval censoring and covariate measurement error when T follows the PO model. They did not make any assumptions on the distribution of X, but assumed that U follows a normal

distribution. In contrast, we propose a methodology to handle covariate measurement error when T is subject to interval censoring and follows the linear transformation model. So far, to the best of our knowledge, the issue of measurement error and interval censoring in the linear transformation model has not been investigated, and this is the main contribution of this paper.

Here we briefly describe our methodology that has three basic components. First, we impute unobserved covariate X from a conditional model given all the observed variables including the information on the time-to-event T. To avoid model misspecification we use a mixture model for the conditional distribution of X with unknown number of mixing components. Second, we impute T given all the information including the covariate imputed in the first step. In the third step we treat the imputed data set as a right-censored data, and analyze it using the semiparametric approach proposed in Chen et al.[13] To account for uncertainty of the imputed values, we apply multiple imputations, and then combine the results in the end. Imputation methods are not new in handling interval-censored data. Pan[14] considered the multiple imputation method for handling interval censoring under the CPH model.

Before concluding this section, we would like to highlight the novel points of this article that accommodates the CPH model, the PO model, and beyond. For handling measurement error, we use a structural approach with a flexible imputation model that can accommodate a wide range of distributions of the covariate X. To the best of our knowledge, we are the first to use an imputation technique for handling covariate measurement error in interval-censored data. Second, we employ the structurally simple estimating equations of Chen et al.[13] in the context of interval-censored data, so that the estimation of the parameters is straightforward.

2 Background, model and notation

Suppose that the data are collected from n independent subjects randomly drawn from an underlying population. The data from the ith subject are $(L_i, R_i, \Delta_i, W_{i1}, \ldots, W_{im}, \boldsymbol{Z}_i)$, $i = 1, \ldots, n$. For each subject, if the right censoring indicator $\Delta = 1$, then the time-to-event T satisfies $L < T \leq R$,

and when $\Delta = 0$, the subject is right-censored above L, and $L < T < \infty$. Here $\boldsymbol{Z}$ is a $p \times 1$ vector of error free covariates, and the vector of observations W denotes the replicated measurements on a surrogate variable for X. The surrogacy implies that conditional on the true covariate X, the surrogate variable W is independent of the response T. We assume that the observed surrogate W is related to the unobserved covariate X through the additive measurement error model:

$$W_{ij} = X_i + U_{ij},\ j = 1, \ldots, m, \tag{2}$$

where measurement error U_{ij}'s are assumed to be independently and identically distributed (iid) Normal$(0, \sigma_u^2)$ variables. Furthermore, we assume that U is completely independent of other observed variables.

Now, we discuss two standard approaches, naive (NV) and regression calibration (RC), for handling covariate measurement error for any model. In the naive approach, all covariates are assumed to be error free, and unobserved X_i is replaced by the average of the W observations, $\overline{W}_i = \sum_{j=1}^{m} W_{ij}/m$, in the estimation method. However, this approach does not produce consistent estimators for $\boldsymbol{\beta}$ and H. The reason is explained in Section S.1 of the Supplementary materials.

In the regression calibration (RC) approach, an unobserved X is replaced by its predicted value $\widehat{X}$ that is a linear function of the surrogate variable W and the error-free covariates $\boldsymbol{Z}$. The expression for the predictor $\widehat{X}$ is given in Section S.2 of the Supplementary materials. Although the RC method works reasonably well when the conditional distribution of X given $\boldsymbol{Z}$ and the vector of W-observations is approximately normal and the measurement error σ_u^2 is small, it is generally not a consistent method.

3 Estimation methodology

3.1 Imputation of unobserved covariate

As mentioned previously there are three major steps in the proposed methodology: imputation of unobserved X, imputation of T for the subjects that are not right-censored, and finally estimation

of $\boldsymbol{\beta}$ and H based on the imputed right-censored data set. Thus, X and T must be imputed from the predictive distribution given the observed data. Let $\underset{\sim}{W} = (W_1, \ldots, W_m)^T$. The predictive distribution of X and T conditional on $\underset{\sim}{W}, \boldsymbol{Z}$ and the censored interval of T is

$$\begin{aligned} f(T, X|\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R) &= f(T|\underset{\sim}{W}, X, \boldsymbol{Z}, L < T \leq R) f(X|\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R) \\ &= f(T|X, \boldsymbol{Z}, L < T \leq R) f(X|\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R), \end{aligned}$$

where $f(T|X, \boldsymbol{Z}, L < T \leq R)$ is the conditional density of T given $X, \boldsymbol{Z}$ and the censored interval and $f(X|\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R)$ is conditional distribution of X given the surrogate $\underset{\sim}{W}$, covariates $\boldsymbol{Z}$, and the censored interval. The second equality in the above display holds because the surrogate variable and the time-to-event are assumed to be independent conditional on the true covariate X.

In this subsection we discuss a robust method of imputing X from the predictive distribution $f(X|\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R)$. To sample X we use the data augmentation technique of Tanner and Wong.[15] We simulate the parameters from their posterior distribution, and given the parameter value we sample unobserved X from its conditional distribution. The conditional distribution of X given $\underset{\sim}{W}, \boldsymbol{Z}, L < T \leq R$ and parameters contains two parts: 1) $f(\underset{\sim}{W}|X, \sigma_u^2)$, the conditional density of $\underset{\sim}{W}$ given X and σ_u^2 which is a product of normal density functions, and 2) $f(X|\boldsymbol{Z}, L < T \leq R, \boldsymbol{\theta}^\dagger)$, the conditional distribution of X given $\boldsymbol{Z}, L < T \leq R$, and parameters. To accommodate a wide range of distributions, we model $f(X|\boldsymbol{Z}, L < T \leq R, \boldsymbol{\theta}^\dagger)$ as a k'-component mixture of normal distributions where the mixture density is

$$f(X|\boldsymbol{Z}, L < T \leq R, \boldsymbol{\theta}^\dagger) = \sum_{l=1}^{k'} \frac{\pi_l}{\sqrt{2\pi\sigma_l^2}} \exp\left\{ -\frac{(X - \boldsymbol{\gamma}_l^T \tilde{\boldsymbol{Z}})^2}{2\sigma_l^2} \right\}. \tag{3}$$

Here $\boldsymbol{\theta}^\dagger$ consists of the mixing proportions $\pi_1, \ldots, \pi_{k'-1}$, the variance components $\sigma_1^2, \ldots, \sigma_{k'}^2$, and the regression parameters $\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_{k'}^T$. The mixing proportions π_l's are assumed to be positive and they add up to one, and $\tilde{\boldsymbol{Z}}$ is a vector of potential predictors for X as functions of $\boldsymbol{Z}$ and the observed

time interval. Specifically we take

$$\tilde{\boldsymbol{Z}} = [1, \boldsymbol{Z}^T, S^*(L), S^*(R), S^*(L)\boldsymbol{Z}^T, S^*(R)\boldsymbol{Z}^T, S^*(L)S^*(R)]^T, \tag{4}$$

with $S^*(L)$ and $S^*(R)$ defined as $\log\{\widehat{S}(L)\}$ and $\log\{1 + \widehat{S}(R)\}$, respectively, and $\widehat{S}$ denotes the estimated survival function from the data $(L_i, R_i, \Delta_i, i = 1, \ldots, n)$ using Turnbull's algorithm.[16] The detailed steps are given in Section S.3 of the Supplementary materials. Besides $\tilde{\boldsymbol{Z}}$, for modeling the dependence of X on $[L, R]$ and $\boldsymbol{Z}$, various other functional forms of the regressor variables can be taken. Note that for a mixture of normals model, the likelihood function is not bounded. Therefore, either penalized maximum likelihood or Bayesian estimation is recommended.[17] Let $\boldsymbol{\theta}$ be the vector of parameters consisting of σ_u^2 and $\boldsymbol{\theta}^\dagger$. We adopt a Bayesian procedure and use Gibbs sampling to estimate $\boldsymbol{\theta}$, and impute unobserved X from its predictive distribution.

First, we describe the prior distributions for each component of $\boldsymbol{\theta}$. We assume InvGamma(a_u, b_u) and Dirichlet$(\alpha, \ldots, \alpha)$ prior on σ_u^2 and $(\pi_1, \ldots, \pi_{k'})$, respectively. Also, we use Normal$(\boldsymbol{\mu}_{\gamma_l}, \Sigma_{\gamma_l})$ prior on the lth regression coefficient $\boldsymbol{\gamma}_l$, and InvGamma(a_σ, b_σ) prior on σ_l^2 for $l = 1, \ldots, k'$. We also introduce latent class-indicators $\psi_1, \ldots, \psi_n$ such that the prior probability that ψ_i takes the value l is π_l for $l = 1, \ldots, k'$. Now, in each iteration of the Gibbs sampling technique, we sample each parameter, latent class-indicators, and unobserved $X_1, \ldots, X_n$ from their respective conditional distribution. To reduce autocorrelation among the sampled observations and for quicker convergence we use partially collapsed Gibbs sampler. The details are given in Section S.4 of the Supplementary materials. We repeat this sampling (iterations) for a large number of times, say M. We then discard the first few thousand iterations as burn-in run. The remaining samples are used to compute the parameter estimates. We repeat this model fitting for different choices of k', and then based on the minimum BIC criteria we choose the optimal k'. Note that BIC is equal to $k^*\log(n) - 2\log(\widehat{\mathcal{L}}_1)$, where k^* is $2k' + k'(3p + 4)$ and it denotes the number of parameters in the model for the vector of W observations given $\tilde{\boldsymbol{Z}}$. Additionally, $\widehat{\mathcal{L}}_1$ denotes the likelihood $\mathcal{L}_1$ of the vector of W observations

given the error free covariates $\boldsymbol{Z}$ and the time interval $[L, R]$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{MAP}$. The expression of $\mathcal{L}_1$ is

$$\begin{aligned}\mathcal{L}_1 &= \prod_{i=1}^{n} \int f(W_{i1}|X_i) \times \cdots \times f(W_{im}|X_i) f(X_i|\tilde{\boldsymbol{Z}}_i) dX_i \\ &= \prod_{i=1}^{n} \frac{1}{(\sqrt{2\pi}\sigma_u)^m} \exp\left\{-\frac{\sum_{j=1}^{m}(W_{ij} - \overline{W}_i)^2}{2\sigma_u^2}\right\} \sum_{l=1}^{k'} \frac{\pi_l}{\sqrt{2\pi(\sigma_u^2/m + \sigma_l^2)}} \exp\left\{-\frac{(\overline{W}_i - \gamma_l^T \tilde{\boldsymbol{Z}}_i)^2}{2(\sigma_u^2/m + \sigma_l^2)}\right\}.\end{aligned}$$

Since X is unobserved in the data, it is integrated out in $\mathcal{L}_1$. For the optimal k', we select m_x^* posterior samples for each X_i as imputed X-values, after thinning the remaining MCMC samples after the burn-in run. Thinning is used to reduce auto-correlation among the successive imputed values.

3.2 Imputation (augmentation) of T

We now describe the imputation of T when imputed X values are available. Imputed X will be denoted by X^*. Also, we shall impute T only when the time-to-event is interval-censored with the upper limit of the interval finite. Denote the known hazard and cumulative hazard functions of e by $\lambda(\cdot)$ and $\Lambda(\cdot)$, respectively. Suppose that for a not right-censored, we divide the interval $(L, R]$ into $m^\dagger$ equal small width intervals, $(r_0, r_1], \ldots, (r_{m^\dagger - 1}, r_{m^\dagger}]$ with $L = r_0$, $R = r_{m^\dagger}$, and then impute the latent T from the following discrete distribution

$$\text{pr}(T = r_l | T \in (L, R], X^*, \boldsymbol{Z}) = \frac{\exp[-\Lambda\{H(r_{l-1}) + X^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}] - \exp[-\Lambda\{H(r_l) + X^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}]}{\exp[-\Lambda\{H(L) + X^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}] - \exp[-\Lambda\{H(R) + X^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}]}$$

for $l = 1, \ldots, m^\dagger$. The denominator of the above expression is the probability that T lies in the interval $(L, R]$ while the numerator is the probability that T lies in the interval $(r_{l-1}, r_l]$. This imputation technique of dividing the interval into a set of grid points was used by Pan.[14] The number of intervals $m^\dagger$ may vary across subjects, and we discuss the choice of $m^\dagger$ in the simulation section.

3.3 Estimation for right-censored data with known covariates

Let T^* denote the imputed time for an uncensored subject. Define $V = T^*$ for $\Delta = 1$ and $V = L$ for $\Delta = 0$. Thus, after imputation, the data on the ith subject are $(V_i, \Delta_i, X_i^*, \boldsymbol{Z}_i)$ for $i = 1, \ldots, n$.

In this subsection we discuss the third major step of the proposed methodology – estimation of $\boldsymbol{\beta}$ and H based on the imputed right-censored data set according to Chen et al.[13]'s approach.

Define the counting process $N(\cdot)$ and the at-risk process $Y(\cdot)$ as $I(V \leq \cdot, \Delta = 1)$ and $I(V \geq \cdot)$, respectively, and suppose that $\boldsymbol{Z}^*$ denotes the vector $(X^*, \boldsymbol{Z}^T)^T$. Then $\mathcal{M}(\vartheta; \boldsymbol{\beta}, H)$ defined as $N(\vartheta) - \int_0^{\vartheta} Y(u) d\Lambda\{H(u) + X^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}$ is a martingale with respect to filtration $\mathcal{F}(\vartheta)$ defined as the sigma algebra on the covariate X^* and $\boldsymbol{Z}$, and the counting and at-risk processes up to time ϑ. Next we estimate $\boldsymbol{\beta}$ by solving the following estimating equations

$$\boldsymbol{S}_{\beta}(\boldsymbol{\beta}, H) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i^* d\mathcal{M}_i(u; \boldsymbol{\beta}, H) = \sum_{i=1}^{n} \boldsymbol{Z}_i^* [\Delta_i - \Lambda\{H(V_i) + X_i^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}] = \mathbf{0}, \tag{5}$$

and $H(u)$ at the ordered failure times $\vartheta_1 < \cdots < \vartheta_K$ (K denotes the number of distinct failure times in the data, ϑ_1: the smallest failure time, ϑ_K: the largest failure time) by solving

$$S_H(\boldsymbol{\beta}, H)(\vartheta_k) = \sum_{i=1}^{n} [dN_i(\vartheta_k) - Y_i(\vartheta_k) d\Lambda\{H(\vartheta_k) + X_i^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}] = 0. \tag{6}$$

Equation (6) is solved as follows. Suppose that d_k denotes the observed number of events/failures at time ϑ_k. We set $H(t)$ equal to $-\infty$ for all t less than ϑ_1 and from (6) obtain $H(\vartheta_1)$ by solving the equation

$$d_1 = \sum_{i=1}^{n} Y_i(\vartheta_1) \Lambda\{H(\vartheta_1) + X_i^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}$$

as $\Lambda(-\infty) = 0$. Next, for a given $H(\vartheta_{k-1})$, $H(\vartheta_k)$ will be obtained by solving the equation $d_k = \sum_{i=1}^{n} Y_i(\vartheta_k)[\Lambda\{H(\vartheta_k) + X_i^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\} - \Lambda\{H(\vartheta_{k-1}) + X_i^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}]$, for $k = 2, \ldots, K$. Prompted by a reviewer's comment, here we point out a difference between Wen and Chen[12]'s estimation technique and our method. Besides the difference in the models, Wen and Chen[12] used a conditional score approach to estimate the regression parameter, and used a self-consistent method to estimate the non-parametric component of their PO model.

Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{H}$ denote the estimates of $\boldsymbol{\beta}$ and H, respectively. Also, for $\vartheta_{k-1} \leq t < \vartheta_k$, we set $\widehat{H}(t) = \widehat{H}(\vartheta_{k-1})$. Under some regularity conditions, the asymptotic variance of $\sqrt{n}\widehat{\boldsymbol{\beta}}$ can be

consistently estimated by $\Sigma(\widehat{\boldsymbol{\beta}}, \widehat{H}) = A^{-1}A_M A^{-T}$, where the expression of A and A_M are given in the Appendix at the end of this paper. Note that A is obtained by taking the derivative of the estimating equation with respect to $\boldsymbol{\beta}$. The middle term A_M is obtained as an estimator of the variance of the estimating equation after proper linearization. The details can be found in Chen et al.[13]

3.4 Complete steps of estimation

We first impute unobserved X and interval-censored T from their prediction distributions. We form a pseudo right-censored data with the imputed values of X and T, while $\boldsymbol{Z}$ remains unchanged in every pseudo data set. We then apply a semiparametric method to estimate model parameters. We repeat this procedure for multiple imputed values of X and T, and then in the end we combine multiple estimates of parameters to obtain the final estimate. Here we summarize the steps of parameter estimation.

Step 1. Impute each unobserved X value m^* times according to the method given in Section 3.1. We use X^*_{i,k_x} as the k_xth imputed value for X_i, $i = 1, \ldots, n$ and $k_x = 1, \ldots, m^*_x$;

Step 2. Initialize $\boldsymbol{\beta}$ and H and denote them by $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{H}^{(0)}$, and set $\widehat{\boldsymbol{\beta}}_{k_x} = \widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{H}_{k_x} = \widehat{H}^{(0)}$;

Step 3. Given X^*_{i,k_x} and the current value of $\widehat{\boldsymbol{\beta}}_{k_x}$ and $\widehat{H}_{k_x}$, impute T_i according to the method in Section 3.2 when $\Delta_i = 1$, and define $V_{i,k_t,k_x} = T^*_{i,k_t}$, where T^*_{i,k_t} denotes the k_tth imputed value for T_i for $i = 1, \ldots, n$ and $k_t = 1, \ldots, m^*_t$;

Step 4. Define $D_{k_t,k_x} = \{V_{i,k_t,k_x}, \Delta_i, X^*_{i,k_x}, \boldsymbol{Z}_i, i = 1, \ldots, n\}$. Estimate $\boldsymbol{\beta}$ and H following the method in Section 3.3 and using data D_{k_t,k_x}. Denote the corresponding estimates by $\widehat{\boldsymbol{\beta}}_{k_t,k_x}$ and $\widehat{H}_{k_t,k_x}$, respectively;

Step 5. Compute $\widehat{\boldsymbol{\beta}}_{k_x} = (1/m^*_t)\sum_{k_t=1}^{m^*_t} \widehat{\boldsymbol{\beta}}_{k_t,k_x}$ and $\widehat{H}_{k_x} = (1/m^*_t)\sum_{k_t=1}^{m^*_t} \widehat{H}_{k_t,k_x}$;

Step 6. Repeat Steps 3–5 until $\widehat{\boldsymbol{\beta}}_{k_x}$ converges;

Step 7. The final estimates are combinations of the estimates from different imputed data sets, and

they are

$$\widehat{\boldsymbol{\beta}}_c = \frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*}\widehat{\boldsymbol{\beta}}_{k_x}, \qquad \widehat{H}_c = \frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*}\widehat{H}_{k_x}. \tag{7}$$

Importantly, there are two major steps: (a) X imputation through the MCMC mechanism, and (b) estimation of $\boldsymbol{\beta}$ and H along with T imputation. Note that the imputation of X is done independently of the $\boldsymbol{\beta}$ and H parameter estimation. On the other hand, the augmentation of T and estimation of $\boldsymbol{\beta}$ and H are amalgamated together. The convergence of the MCMC chain can be checked at least by the trace plot that shows the parameter values for each iteration of the MCMC chain, and the Gelman-Rubin diagnostic plot to check if the observed MCMC samples are reasonably close to the target posterior distribution. Once the MCMC chain reaches the equilibrium, the sampled X values can be considered to come from the predictive density. In Step 4 above, we estimate H from Equation (6) for a given $\boldsymbol{\beta}$. Then we estimate $\boldsymbol{\beta}$ by solving Equation (5) using the Newton-Raphson method while treating H as known with H being set to the last estimated value. Next, we estimate H again with the updated estimate of $\boldsymbol{\beta}$ and continue these steps until the estimate of $\boldsymbol{\beta}$ converges with a relative tolerance of 1%.

3.5 Large sample properties

Our estimator $\widehat{\boldsymbol{\beta}}_c$ converges in probability to $\boldsymbol{\beta}^*$. If the parametric imputation model for X is misspecified, then $\boldsymbol{\beta}^*$ may differ from the true parameter $\boldsymbol{\beta}$. Otherwise $\boldsymbol{\beta}^* = \boldsymbol{\beta}$. The asymptotic distributional properties of the estimator are summarized as follows with the proof given in Section S.6 of the Supplementary materials.

Theorem 1. *Under some regularity conditions, as $n \to \infty$ the distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}^*)$ converges to a mean-zero normal distribution, and the asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}^*)$ is*

$$E\left\{D_1^{-1}S_1(\boldsymbol{\beta}^*, \boldsymbol{\beta}^*, \boldsymbol{\theta}) + D_1^{-1}D_2\psi_1(\boldsymbol{\theta})\right\}^{\otimes 2} + (m_x^*)^{-1}D_1^{-1}D_2\{\mathcal{I}_1(\boldsymbol{\theta})\}^{-1}D_2^T D_1^{-T}.$$

Corollary 1. *The asymptotic variance of* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}^*)$ *can be consistently estimated by*

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{D}_1^{-1}S_i(\widehat{\boldsymbol{\beta}}_c,\widehat{\boldsymbol{\beta}}_c,\widehat{\boldsymbol{\theta}}_{MAP})+\widehat{D}_1^{-1}\widehat{D}_2\psi_i(\widehat{\boldsymbol{\theta}}_{MAP})\right\}^{\otimes 2}+(m_x^*)^{-1}\widehat{D}_1^{-1}\widehat{D}_2\{\widehat{\mathcal{I}}_1(\widehat{\boldsymbol{\theta}}_{MAP})\}^{-1}\widehat{D}_2^T\widehat{D}_1^{-T}.$$

Here $\widehat{\boldsymbol{\theta}}_{MAP}$ denotes the maximum a posteriori estimator of $\boldsymbol{\theta}$. The expression of the terms are given in the end of this paper. Here we provide some intuition behind the result. Define $\boldsymbol{S}_\beta(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}),k_x,k_t) = (S_{\beta_1}(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}),k_x,k_t),\ \boldsymbol{S}_{\beta_2}^T(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}),k_x,k_t))^T$ where the two components are nothing but $S_{\beta_1}(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}))$ and $\boldsymbol{S}_{\beta_2}(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}))$ with added components k_x and k_t to denote the corresponding dataset. Here $\widehat{H}(t,\boldsymbol{\beta})$ denotes the solution of Equation (6) for a given $\boldsymbol{\beta}$. To prove the asymptotic results, first note that due to Lemma 1 of Wang and Robins,[18] our estimator $\widehat{\boldsymbol{\beta}}_c$ is asymptotically equivalent to $\widetilde{\boldsymbol{\beta}}$ that is the solution of $n^{-1/2}(m_x^*m_t^*)^{-1}\sum_{k_x=1}^{m_x^*}\sum_{k_t=1}^{m_t^*}\boldsymbol{S}_\beta(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}),k_x,k_t)=\mathbf{0}$. In proving the weak convergence results we take into account three aspects: 1) given X, variability of the regression parameter estimator, 2) uncertainty due to the estimation of $\boldsymbol{\theta}$, and 3) variability of the posterior samples of $\boldsymbol{\theta}$. This asymptotic variance can also be used in the Wald test for $\boldsymbol{\beta}$.

Further discussion is warranted regarding the asymptotic results. The parameter $\boldsymbol{\beta}^*$ used in Theorem 1 can be characterized as the solution of $E[(m_x^*m_t^*n)^{-1}\sum_{k_x=1}^{m_x^*}\sum_{k_t=1}^{m_t^*}\boldsymbol{S}_\beta(\boldsymbol{\beta},\widehat{H}(\cdot,\boldsymbol{\beta}),k_x,k_t)]=\mathbf{0}$. If the assumed model for X is the true X generating process then $\boldsymbol{\beta}^*=\boldsymbol{\beta}$. We have to accept the fact that it is not possible to guarantee that the assumed model and the true data generating process are identical. However, we had proposed to use a flexible finite mixture of normals to model the conditional distribution of X given $\boldsymbol{Z}$ and observed time interval. There we used the minimum BIC criteria to choose the optimal number of mixing components. This in turn allows us to obtain the best fitted model for the observed data. Moreover, to avoid misspecification of the dependence of X on other variables, we took a flexible structure for $\widetilde{\boldsymbol{Z}}$ as given in (4). Under this flexible model, it is reasonable to assume that our estimator estimates $\boldsymbol{\beta}$ (or in other words $\boldsymbol{\beta}^*=\boldsymbol{\beta}$).

As σ_u^2 gets small, D_2 tends to be small. In the no measurement error case, $D_2=0$. In this case we do not impute X values (i.e., $X_i=W_{i,1}=X_{i,1}^*(\boldsymbol{\theta})$, $m_x^*=1$). As expected, the asymptotic distribution

of the estimator $\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is then the same as that of $-n^{-1/2}\sum_{i=1}^{n} D_1^{-1} S_i(\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\theta})$ which is normal with mean zero and a proper covariance matrix, where $S_i(\boldsymbol{\beta},\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{k_t=1}^{m_t^*} S_i(\boldsymbol{\beta}, V_{i,k_t,1}(\boldsymbol{\beta}), X_{i,1}^*(\boldsymbol{\theta}))/m_t^*$. The expression of the summand $S_i(\boldsymbol{\beta}, V_{i,k_t,1}(\boldsymbol{\beta}), X_{i,1}^*(\boldsymbol{\theta}))$ is given in the Appendix.

Alternative to the asymptotic variance formula of Corollary 1, here we also provide the variance formula by Rubin's[19] (p. 76) approach,

$$\Omega = \frac{1}{m_x^* m_t^*}\sum_{k_x=1}^{m_x^*}\sum_{k_t=1}^{m_t^*}\Sigma(\widehat{\boldsymbol{\beta}}_{k_t,k_x}, \widehat{H}_{k_t,k_x}) + (1+\frac{1}{m_x^* m_t^*})\sum_{k_x=1}^{m_x^*}\sum_{k_t=1}^{m_t^*}\frac{(\widehat{\boldsymbol{\beta}}_{k_t,k_x}-\widehat{\boldsymbol{\beta}}_{k_x})(\widehat{\boldsymbol{\beta}}_{k_t,k_x}-\widehat{\boldsymbol{\beta}}_{k_x})^T}{m_x^* m_t^* - 1}. \tag{8}$$

Formula (8) has two parts. The first part assesses the variability of the estimates for a given imputed data set while the second term measures the variability between the estimates for different imputed data sets. In most practical applications, the difference between Rubin's variance formula and the more accurate variance formula is negligible unless the imputation model is grossly wrong.[20] Due to ease of implementation, in our numerical studies we have used the variance formula (8).

4 Simulation study

First we considered the case where intervals for each non-censored subject were of equal length. Each simulated data set consisted of $n = 200$ subjects. For each subject, a scalar Z was simulated from Bernoulli(0.5) and X was simulated from two different distributions: 1) Normal$(0,1)$ and 2) $\{\text{Gamma}(2,2)-4\}/\sqrt{8}$ which is referred to as the modified gamma (MG) distribution. To generate time-to-event T, $\log(T)$ was set equal to $-\beta_1 X - Z\beta_2 + e$, where $\beta_1 = -1$ and $\beta_2 = 1$ and the error e followed the extreme-value distribution with the cumulative distribution function

$$F(e) = \begin{cases} 1-\exp\{-\exp(e)\}, & \text{for } r=0; \\ 1-\exp[-\log\{r\exp(e)+1\}/r], & \text{for } r>0. \end{cases} \tag{9}$$

For simulating interval-censored data, let $0 = v_0 < v_1 < \cdots < v_8 < v_9 = \infty$ be the eight scheduled visits.[21] Then for each subject, v_1 was simulated from Uniform$(0,\ \vartheta^\dagger)$, and the next seven follow-up times were generated using the formula: $v_t = v_1 + (t-1)d$, $t = 2,\ldots,8$. For 30% right censoring, d was set at 0.1, and for r equal to 0 and 1, $\vartheta^\dagger$ was set to 0.1 and 2.1, respectively. A subject was

considered right-censored (i.e., $\Delta = 0$), if the corresponding T fell above v_8. Otherwise the subject was considered interval-censored ($\Delta = 1$) with the interval (v_t, v_{t+1}) when T lies between v_t and v_{t+1}. To obtain W, we set $W = X + U$, where measurement error $U = \sigma_u U^*$ with two distributions for U^*: 1) Normal$(0, 1)$ and 2) modified gamma (MG). We took two values for σ_u^2, 0.25 and 0.5. We considered two replicated measurements for W that were obtained by adding two independent copies of U to a given X value.

Each data set was analyzed by four methods. For reference, the interval-censored data were first analyzed using the true X. This method is referred to as the no measurement error case (NM). In NM, T imputation (Section 3.2) and parameter estimation (Section 3.3) were repeated until the estimates converged. In the next three approaches, instead of X, W were used. In NV, each unobserved X was replaced by the average of the two replicated W's while in RC, X was replaced by $\widehat{X}$ discussed in Section 2. Finally, in the proposed imputation based method (IM), the parameters were estimated according to the steps given in Section 3.4. For IM, we set $m_x^* = 20$ and $m_t^* = 10$. In IM, for sampling from the predictive distribution, we used InvGamma$(1, 1)$ and Dirichlet$(1, \ldots, 1)$ prior for σ_u^2 and $(\pi_1, \ldots, \pi_{k'})$, respectively. For σ_l^2 and $\boldsymbol{\gamma}_l$ we used InvGamma$(1, 1)$ and Normal$(\boldsymbol{\mu}_{\gamma_l}, 5I_7)$ prior, respectively, for $l = 1, \ldots, k'$. Each $\boldsymbol{\mu}_{\gamma_l}$ was a vector of length 7 ($3p+4 = 3\times1+4$) whose first element was chosen to be the $\{100l/(k'+1)\}$th quantile of $\overline{W}$ and the rest of the components were randomly drawn from the standard normal distribution, and I_7 represents the identity matrix of order 7.

In each scenario, we present bias (B), simulation standard error (S), estimated standard error (E) and 95% coverage probability (C) using the Wald confidence interval based on the converged datasets out of 1000 replications. Approximately 1-1.5% data sets had convergence issue. For IM, the estimated standard error was calculated using formula (8) while for NM and NV it was calculated using formula (8) but without any imputation for X. For RC, the estimated standard errors were calculated using the formula given in Section S.5 of the Supplementary materials. The bias in NM serves as the benchmark. We are mainly concerned with the estimation bias of β_1. Results of Table

1 indicate that the bias in NV is substantially large and the 95% coverage probability is quite low for $\sigma_u^2 = 0.5$. Although RC shows satisfactory performance when X follows the normal distribution and σ_u^2 is 0.25, its bias substantially increases and the coverage probability markedly decreases when σ_u^2 is increased to 0.5. Even for $\sigma_u^2 = 0.25$, RC performs poorly when X follows the MG distribution across different r values. In contrast, the bias in IM is generally substantially smaller than that in NV and RC. Also, the empirical coverage probability for IM is reasonably close to the nominal level. For all the methods, the estimated standard errors are reasonably close to the simulation standard deviations indicating that the asymptotic standard error formula is valid. The standard error of IM is generally larger than that of NV and RC because of the uncertainity of the imputed X values.

Next, mimicking the AIDS data set, we simulated datasets with $n = 500$, unequal lengths of intervals, and approximately 90% right-censored subjects. To do so, ordered examination times $v_1 < \cdots < v_8$ were simulated for each subject as before. Subjects were then allowed to miss the first four scheduled visits with probability 0.3 and the last four with probability 0.5. The interval $(L, R]$ was the shortest interval between two non-missed visit times that contained T. Here, we chose $\vartheta^\dagger = 0.0001$, and for $r = 0$, $d = 0.12$ and for $r = 1$, $d = 0.15$ to maintain 90% censoring. The generation of the rest of the variables remained the same as in the equal-length setup. The results for r equal to 0 and 1 cases are presented in Table 2. When X follows the MG distribution, regardless of the distribution of U (or U^*), IM performs much better compared to NV and RC in terms of bias and coverage. However, when X follows a normal distribution, the results indicate similar performance by RC and IM while both are superior to NV. To assess the performance of IM beyond the CPH ($r = 0$) and PO ($r = 1$) models, data sets were also generated for a general $r = 2$ case with $\vartheta^\dagger = 0.0001$ and $d = 0.15$ to maintain 90% censoring. To save space, only the results for both X and U following the MG distribution are presented (Table 3). Again IM shows superior performance.

Note that the cases where X or U follows the MG distribution are violations of our model assumptions on the distribution of X and U. Thus these results and some additional simulation

studies presented in Section S.7 of the Supplementary materials help judge the sensitivity of IM towards those assumptions. Following a reviewer's comment, we also compared the proximity of the estimated and empirical variances for the IM method. For the normal-normal scenario of Table 2 with $r = 0$ and $\sigma_u^2 = 0.5$, the estimated and empirical variances for $\widehat{\beta}_1$ were 0.029 and 0.024, respectively. The 80% and 95% coverage probabilities were 0.834 and 0.967, respectively. For the MG-MG scenario of Table 2 with $r = 0$ and $\sigma_u^2 = 0.5$, the estimated and empirical variances for $\widehat{\beta}_1$ were 0.087 and 0.095, respectively, and the 80% and 95% coverage probabilities were 0.806 and 0.947, respectively. These results along with S, E, C presented in the tables provide evidence in favor of the numerical validity of the variance estimation technique used in this paper.

5 Real data example

Now we analyze the motivating data set from the ACTG 175 trial, a randomized, double-blinded, placebo controlled clinical trial to compare nucleoside monotherapy with combination therapy in HIV-infected subjects. In our analysis, we consider only 516 subjects who had received zidovudine alone ($Z = 1$) or the combination therapy zidovudine plus didanosine ($Z = 0$). Out of the 516 subjects, 50 subjects experienced the event (defined in the introduction) in the trial. The time-to-event T is the length of time from when the treatment started to time when the event occurred. Here, the logarithm of the unobserved true CD4 cell count is denoted by X, whereas the logarithm of the observed CD4 cell count are the surrogate measurements, W. Since there are two replications of W, we use $m = 2$. The estimated measurement error variance for our data was approximately 44% of the variance of true X. This follows from the following facts:

$$\text{var}(W_{i,1} - W_{i,2}) = \text{var}(U_{i,1} - U_{i,2}) = \text{var}(U_{i,1}) + \text{var}(U_{i,2}) = 2\sigma_u^2, \tag{10}$$

and $\text{var}(\overline{W}_i) = \text{var}(X_i) + \text{var}(\overline{U}_i)$, where $\overline{W}_i$ and $\overline{U}_i$ denote $(W_{i,1} + W_{i,2})/2$ and $(U_{i,1} + U_{i,2})/2$, respectively. Next, σ_u^2 is estimated by the sample variance of n pairwise differences of the replicated W observations divided by 2, and σ_x^2 is estimated by subtracting the estimator of $\sigma_u^2/2$ from the

sample variance of n average values of replicated W observations. Once these are estimated we can estimate the noise to signal ratio, σ_u^2/σ_x^2. The estimated error variance is within the range of the values of σ_u^2 considered in our simulation study. We model T by setting $H(T)$ equal to $-X\beta_1 - Z\beta_2 + e$, where e is assumed to follow the distributions given in (9), and H is an unknown non-decreasing function of T. Here our primary goal is to draw inference on β_1 and β_2.

We now analyze the data using the proposed IM method and compare its performance with the existing NV and RC methods. Since true X is never recorded in this real data, we cannot apply NM. For the IM method, we use $m_x^* = 20$ and $m_t^* = 10$ and analyze the data for $r = 0, 1, 2$. We use the same flexible prior distributions for the IM approach as used in the simulation study. The left panel of Table 4 contains the results of the analyses. In all three methods and for different r, high baseline CD4 count seems to act as a protective factor on the time-to-event. Also, the combination therapy seems to have an advantage over the monotherapy for elongating time-to-event. The estimated regression coefficients for log(CD4) differ across the methods. It is worth mentioning that in IM the regression parameter estimate corresponding to the treatment when $r = 1$ is reasonably close to the corresponding estimate of Wen and Chen.[12] Also, the negative sign of the regression parameter estimate for $\log(CD4)$ is consistent with the findings in other articles[9,12] that analyzed this data set. Although different authors are using the data from the same clinical study, it is quite difficult to verify if authors are using the same subset of the main data set. This is likely to result in different parameter estimates in different papers. While the NV estimates are different from that of RC and IM, the RC and IM estimates of β_1 are somewhat close. This intuitively indicates that the underlying distribution of X is likely to be normal because in Table 2 the biases for the RC and IM methods were quite close when X followed a normal distribution. We point out that in IM a two-component normal mixture model was used for modelling the distribution of X and the number of components was determined based on the minimum BIC criteria. For this selected model, we have verified the convergence of the MCMC chain through trace and Gelman-Rubin diagnostic plots given in Section

S.8 of the Supplementary materials. There was no convergence problem for estimating $\boldsymbol{\beta}$ either.

For the purpose of further illustration that parameter estimates could be really different between the RC and IM methods, we have replaced the observed W-values by simulated W's in this data set, so that the underlying distribution of X is more skewed than the observed data set. We simulated W's by $W = X + U$, where U's were simulated from the modified gamma distribution, and for the censored cases, X was generated from Gamma$(0.5, 0.7)$ (the mean of X was 0.35) and for the non-censored cases, X was simulated from Gamma$(0.5, 1.6)$ (the mean of X was 0.8). Next, we analyzed this partly simulated data using the NV, RC, and IM methods. The results given in the right panel of Table 4 clearly show appreciable differences in the regression parameter estimates for X across the methods and for different $r = 0, 1, 2$ values. The results of our simulation studies suggest that these differences are reflective of the smaller bias for the IM method. This indicates the advantage of using the IM method to estimate the regression parameters over existing methods when the underlying latent variable has a skewed distribution.

6 Conclusion

We have developed a semiparametric methodology for analyzing the linear transformation model for interval-censored data when a covariate is measured with error. As mentioned, linear transformation model contains standard choices of proportional hazards and proportional odds models as special cases. It allows us to study the performance of both standard and non-standard models using the methodology in this paper. The proposed method is fairly robust towards the distribution of X. As demonstrated in our simulation studies, the proposed method works well for scenarios including the case of high percentage of right censoring and for slightly asymmetric measurement error distributions. While in some cases the regression calibration shows satisfactory performance, overall the performance of the proposed method is superior across different scenarios.

Technically, the proposed methodology can be extended to the scenario where multiple covariates are measured with error. For that, model (3) needs to be replaced by a mixture of multivariate dis-

tributions or a series of conditional distributions that are used in multivariate imputation by chained equations (MICE[22]), and the measurement error should be modeled by a multivariate distribution. However, some novelty is needed for handling a high-dimensional covariate measured with error. In addition, the modeling technique may vary depending on whether internal or external validation data are available. The proposed methodology can also be extended to case-K interval-censored data. In addition, it can be applied to the scenario where the data contain a mixture of exact, right-censored and interval-censored time-to-event observations.

Alternative to our semiparametric methodology, one may develop a methodology based on the nonparametric maximum likelihood method of Zeng and Lin.[23] However, we chose to use the semiparametric methodology given in Chen et al.[13] for regression parameter estimation due to the simplicity of its computational algorithm. Particularly, the recursive procedure in estimating the H function is easy to implement. In summary, the significance of the proposed work lies in the fact that we are the first to provide a solution of this important problem. We have created an R package, named *icemelt*, for the proposed approach, and it is currently available at `http://www.stat.tamu.edu/~sinha/research.html`, and soon it will be available through `https://cran.r-project.org/`.

7 Supplementary materials

The materials referenced in Sections 2, 3.1, 3.5, 4, and 5 are available in the Supplementary materials.

Acknowledgments

The authors wish to thank the Associate Editor and referees for their valuable comments and suggestions that substantially improved earlier versions of this manuscript. S. Wang's research was supported in part by the Simons Foundation Mathematics and Physical Sciences - Collaboration Grants for Mathematicians Program Award 499650.

REFERENCES

1. Huang J, Wellner J. Interval censored survival data: A review of recent progress. In: Springer-Verlag. 1997 (pp.123–169).

2. Hammer S, Katzenstein D, Hughes M, et al. Bayesian modelling of measurement error in predictor variables using item response theory. *N Engl J Med.* 1996;335:1081–1090.

3. Carroll R, Ruppert D, Stefanski L, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, FL: CRC press. 2006.

4. Prentice R. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika.* 1982;69:331–342.

5. Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics.* 1992;48:829–838.

6. Huang Y, Wang C. Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *J Am Stat Assoc.* 2000;95:1209–1219.

7. Sinha S, Ma Y. Analysis of proportional odds models with censoring and errors-in-covariates. *J Am Stat Assoc.* 2016;111:1301–1312.

8. Cheng S, Wang N. Linear transformation models for failure time data with covariate measurement error. *J Am Stat Assoc.* 2001;96:706–716.

9. Sinha S, Ma Y. Semiparametric analysis of linear transformation models with covariate measurement errors. *Biometrics.* 2014;70:21–32.

10. Song X, Ma S. Multiple augmentation for interval-censored data with measurement error. *Statist Med.* 2008;27:3178–3190.

11. Tsiatis A, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika.* 2001;88:447–458.

12. Wen C, Chen Y. Functional inference for interval-censored data in proportional odds model with covariate measurement error. *Stat Sin.* 2014;24:1301–1317.

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

14. Pan W. A multiple imputation approach to Cox regression with interval-censored data. *Biometrics.* 2000;56:199–203.

15. Tanner M, Wong W. The calculation of posterior distributions by data augmentation. *J Am Stat Assoc.* 1987;82:528–540.

16. Turnbull B. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J R Stat Soc Series B Stat Methodol.* 1976;38:290–295.

17. Ciuperca G, Ridolfi A, Idier J. Penalized maximum likelihood estimator for normal mixtures. *Scand Stat Theory Appl.* 2003;30:45–59.

18. Wang N, Robins J. Large-sample theory for parametric multiple imputation procedures. *Biometrika.* 1998;85:935–948.

19. Rubin D. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: Wiley. 2004.

20. Hughes R, Sterne J, Tilling K. Comparison of imputation variance estimators. *Stat Methods Med Res.* 2016;25:2541–2557.

21. Chen L, Sun J. A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Comput Stat Data Anal.* 2010;54:1109–1116.

22. Buuren vS, Groothuis-Oudshoorn K. MICE: Multivariate imputation by chained equations in R. *J Stat Softw.* 2010;45:1–68.

23. Zeng D, Lin D. Maximum likelihood estimation in semiparametric models with censored data. *J R Stat Soc Series B Stat Methodol.* 2007;69:507–564.

Appendix

Expressions of the terms involved in $\Sigma(\widehat{\boldsymbol{\beta}}, \widehat{H})$

$$A = \frac{1}{n}\sum_{k=2}^{K}\sum_{i=1}^{n}\{\boldsymbol{Z}_i^* - \boldsymbol{\mu}_z(\vartheta_k, \widehat{\boldsymbol{\beta}})\}(\boldsymbol{Z}_i^*)^T\dot{\lambda}\{\widehat{\beta}_1 X_i^* + \boldsymbol{Z}_i^T\widehat{\boldsymbol{\beta}}_2 + \widehat{H}(\vartheta_k, \widehat{\boldsymbol{\beta}})\}Y_i(\vartheta_k)\{\widehat{H}(\vartheta_k, \widehat{\boldsymbol{\beta}}) - \widehat{H}(\vartheta_{k-1}, \widehat{\boldsymbol{\beta}})\},$$

$$A_M = \frac{1}{n}\sum_{k=2}^{K}\sum_{i=1}^{n}\{\boldsymbol{Z}_i^* - \boldsymbol{\mu}_z(\vartheta_k, \widehat{\boldsymbol{\beta}})\}^{\otimes 2}\lambda\{\widehat{\beta}_1 X_i^* + \boldsymbol{Z}_i^T\widehat{\boldsymbol{\beta}}_2 + \widehat{H}(\vartheta_k, \widehat{\boldsymbol{\beta}})\}Y_i(\vartheta_k)\{\widehat{H}(\vartheta_k, \widehat{\boldsymbol{\beta}}) - \widehat{H}(\vartheta_{k-1}, \widehat{\boldsymbol{\beta}})\},$$

$$\boldsymbol{\mu}_z(t, \boldsymbol{\beta}) = \frac{\sum_{i=1}^n Z_i^*\lambda\{\beta_1 X_i^* + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}(V_i, \boldsymbol{\beta})\}Y_i(t)B(t, V_i, \boldsymbol{\beta})}{\sum_{i=1}^n \lambda\{\beta_1 X_i^* + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}(t, \boldsymbol{\beta})\}Y_i(t)},$$

$$B(t, s, \boldsymbol{\beta}) = \exp\left[-\sum_{k: t\le\vartheta_{k-1}<\vartheta_k\le s}\frac{\sum_{i=1}^n[\dot{\lambda}\{\beta_1 X_i^* + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}(\vartheta_k, \boldsymbol{\beta})\}Y_i(\vartheta_k)]}{\sum_{i=1}^n[\lambda\{\beta_1 X_i^* + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}(\vartheta_k, \boldsymbol{\beta})\}Y_i(\vartheta_k)]}\{\widehat{H}(\vartheta_k, \boldsymbol{\beta}) - \widehat{H}(\vartheta_{k-1}, \boldsymbol{\beta})\}\right]$$

for $t \le s$, and $\dot{\lambda}(t)$ denotes the derivative of λ. Here the estimator of H is denoted by $\widehat{H}(\cdot, \widehat{\boldsymbol{\beta}})$.

Expressions of the terms involved in Theorem 1 and Corollary 1

Define $\mathcal{I}_1(\boldsymbol{\theta}) = E\{-(1/n)\partial^2\log(\mathcal{L}_1)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T\}$ and $S_i(\boldsymbol{\beta}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*} S_i(\boldsymbol{\beta}, V_{i,k_t,k_x}(\boldsymbol{\beta}), X_{i,k_x}^*(\boldsymbol{\theta}))$ $/m_t^* m_x^*$, where $S_i(\boldsymbol{\beta}, V_{i,k_t,k_x}(\boldsymbol{\beta}), X_{i,k_x}^*(\boldsymbol{\theta})) = \int_0^\tau\{\boldsymbol{Z}_{i,k_x}^* - \boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}d\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta}, H)$ with $\boldsymbol{Z}_{i,k_x}^* = (X_{i,k_x}^*, \boldsymbol{Z}_i^T)^T$, $\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta}, H) = N_{i,k_t,k_x}(u) - \int_0^u Y_{i,k_t,k_x}(\vartheta)d\Lambda\{H(\vartheta) + X_{i,k_x}^*\beta_1 + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}$, $N_{i,k_t,k_x}(u) = I(V_{i,k_t,k_x} \le u, \Delta_{i,k_t,k_x} = 1)$, $Y_{i,k_t,k_x}(u) = I(V_{i,k_t,k_x} \ge u)$, and $\boldsymbol{\mu}_{z,k_t,k_x}(t;\boldsymbol{\beta})$ being the $\boldsymbol{\mu}_z(t;\boldsymbol{\beta})$ for the (k_t, k_x)th imputed data set. Also, define $\widehat{H}_{k_t,k_x}(u, \boldsymbol{\beta})$ as the estimator of H for the (k_t, k_x)th imputed dataset for given $\boldsymbol{\beta}$ and this estimator is obtained by solving (6). Additionally,

$$D_1 = -\left(E\left[\frac{1}{m_t^* m_x^*}\sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*}\int_0^\tau\{\boldsymbol{Z}_{k_x}^* - \boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}Y_{k_t,k_x}(u)\left[\boldsymbol{Z}^{*T}_{k_x}\dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) + X_{k_x}^*\beta_1 + \boldsymbol{Z}^T\boldsymbol{\beta}_2\}\right.\right.$$

$$\times d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+\dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+X^*_{k_x}\beta_1+\boldsymbol{Z}^T\boldsymbol{\beta}_2\}\frac{\partial H_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+\lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})$$

$$+X^*_{k_x}\beta_1+\boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}\times\frac{\partial d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big]\Big]+E\Big[\int\frac{1}{m^*_x}\sum_{k_x=1}^{m^*_x}\int_0^\tau\{\boldsymbol{Z}^*_{k_x}-\boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}d\mathcal{M}_{k_t,k_x}(u;\boldsymbol{\beta},H)$$

$$\times\Delta\frac{\partial}{\partial\boldsymbol{\beta}}\log\{f(T_{k_t,k_x}|X^*_{k_x},\boldsymbol{Z},L<T_{k_t,k_x}\le R;\boldsymbol{\beta})\}f(T_{k_t,k_x}|X^*_{k_x},\boldsymbol{Z},L<T_{k_t,k_x}\le R;\boldsymbol{\beta})dT_{k_t,k_x}\Big]\Big)_{\boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

$$D_2 = \Big(E\Big[\frac{1}{m^*_t}\sum_{k_t=1}^{m^*_t}\int S(\boldsymbol{\beta},V_{k_t,k_x}(\boldsymbol{\beta}),X^*_{k_x})\frac{\partial}{\partial\boldsymbol{\theta}^T}\log\{f(X^*_{k_x}|L<T<R,W_1,\ldots,W_m,\boldsymbol{Z};\boldsymbol{\theta})\}$$

$$\times f(X^*_{k_x}|L<T<R,W_1,\ldots,W_m,\boldsymbol{Z};\boldsymbol{\theta})dX^*_{k_x}\Big]\Big)_{\boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

$$\widehat{D}_1 = -\frac{1}{n}\sum_{i=1}^n\Big(\frac{1}{m^*_tm^*_x}\sum_{k_t=1}^{m^*_t}\sum_{k_x=1}^{m^*_x}\int_0^\tau\{\boldsymbol{Z}^*_{i,k_x}-\boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}Y_{i,k_t,k_x}(u)$$

$$\times\Big[\boldsymbol{Z}^{*T}_{i,k_x}\dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+X^*_{i,k_x}\beta_1+\boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})$$

$$+\dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+X^*_{i,k_x}\beta_1+\boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}\frac{\partial\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+\lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})+X^*_{i,k_x}\beta_1+\boldsymbol{Z}_i^T\boldsymbol{\beta}_2\}$$

$$\times\frac{\partial d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big]\Big)_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c,\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MAP}}$$

$$+\frac{1}{n}\sum_{i=1}^n\Big[\int\frac{1}{m^*_x}\sum_{k_x=1}^{m^*_x}\int_0^\tau\{\boldsymbol{Z}^*_{i,k_x}-\boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}d\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta},\widehat{H}_{k_t,k_x})$$

$$\times\Delta_i\frac{\partial}{\partial\boldsymbol{\beta}}\log\{f(T_{i,k_t,k_x}|X^*_{i,k_x},\boldsymbol{Z}_i,L_i<T_{i,k_t,k_x}\le R_i;\boldsymbol{\beta})\}$$

$$\times f(T_{i,k_t,k_x}|X^*_{i,k_x},\boldsymbol{Z}_i,L_i<T_{i,k_t,k_x}\le R_i;\boldsymbol{\beta})dT_{i,k_t,k_x}\Big]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c,\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MAP}}$$

$$\widehat{D}_2 = \frac{1}{n}\sum_{i=1}^n\Big[\frac{1}{m^*_t}\sum_{k_t=1}^{m^*_t}\int S_i(\boldsymbol{\beta},V_{i,k_t,k_x}(\boldsymbol{\beta}),X^*_{i,k_x})\frac{\partial}{\partial\boldsymbol{\theta}^T}\log\{f(X^*_{i,k_x}|L_i<T_i<R_i,W_{i,1},\ldots,W_{i,m},\boldsymbol{Z}_i;\boldsymbol{\theta})\}$$

$$\times f(X^*_{i,k_x}|L_i<T_i<R_i,W_{i,1},\ldots,W_{i,m},\boldsymbol{Z}_i;\boldsymbol{\theta})dX^*_{i,k_x}\Big]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c,\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MAP}},$$

and with $\sigma^{*,2}_l=1/(m/\sigma^2_u+1/\sigma^2_l)$, $\pi^*_l=\pi_l(\sigma^*_l/\sigma_l)\exp[0.5\{\sigma^{*,2}_l(\overline{W}_i/\sigma^2_u/m+\gamma_l^T\widetilde{\boldsymbol{Z}}_i/\sigma^2_l)-(\gamma_l^T\widetilde{\boldsymbol{Z}}_i)^2/\sigma^2_l\}]$,

$$f(X^*_{i,k_x}|L_i<T_i<R_i,W_{i,1},\ldots,W_{i,m},\boldsymbol{Z}_i;\boldsymbol{\theta})=\sum_{l=1}^{k'}\frac{\pi^*_l}{\sqrt{2\pi}\sigma^*_l}\exp\left[-\frac{1}{2\sigma^{*,2}_l}\left\{X^*_{i,k_x}-\sigma^{*,2}_l\left(\frac{\overline{W}_i}{\sigma^2_u/m}+\frac{\gamma_l^T\widetilde{\boldsymbol{Z}}_i}{\sigma^2_l}\right)\right\}^2\right].$$

Table 1: Simulation results based on 1000 replications for $r = 0$ and 1 with $n = 200$, equal-length intervals and 30% right censoring on average. Here measurement error $U = \sigma_u U^*$. All entries are multiplied by 100. B ≡ bias, S ≡ standard deviation, E ≡ estimated standard error, C ≡ 95% coverage probability, $N \equiv$ Normal, $MG \equiv$ Modified Gamma, NM ≡ No measurement error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

			$X \sim N(0,1)$, $U^* \sim N(0,1)$								$X \sim N(0,1)$, $U^* \sim MG$							
			β_1				β_2				β_1				β_2			
r	σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
0	0.25	B	−1.4	14.1	3.7	2.4	1.1	−4.5	−4.5	−2.0	−1.4	12.8	1.7	1.2	1.1	−3.9	−3.8	−1.4
		S	11.6	10.7	11.9	12.4	19.2	19.0	19.0	19.6	11.6	10.7	12.0	12.6	19.2	19.0	18.9	19.6
		E	11.1	10.0	11.9	12.7	18.3	18.1	18.9	19.5	11.1	10.2	12.0	13.0	18.3	18.1	18.8	19.5
		C	94.1	66.7	92.9	94.0	94.5	92.8	93.4	94.7	94.1	71.4	93.7	94.7	94.5	93.6	93.2	95.5
0	0.5	B	−1.4	25.5	7.1	4.0	1.1	−8.1	−8.2	−3.6	−1.4	22.9	3.2	2.0	1.1	−7.1	−7.3	−2.7
		S	11.6	9.9	12.5	14.0	19.2	18.8	19.2	20.2	11.6	10.0	12.6	14.1	19.2	18.9	18.9	20.0
		E	11.1	9.2	12.7	14.3	18.3	18.0	19.1	20.4	11.1	9.5	12.9	14.7	18.3	18.0	19.0	20.5
		C	94.1	23.4	87.8	92.2	94.5	92.0	92.3	94.8	94.1	33.1	93.6	94.8	94.5	91.8	92.9	95.7
1	0.25	B	3.7	15.3	4.8	2.4	−4.0	−5.8	−5.1	−4.1	3.7	15.1	4.1	2.2	−4.0	−5.3	−5.8	−3.1
		S	17.7	16.3	18.4	19.3	32.5	32.0	32.5	32.7	17.7	16.7	18.1	19.9	32.5	32.2	32.1	32.2
		E	16.7	15.4	17.4	18.5	29.7	29.6	29.5	30.3	16.7	15.5	17.5	18.7	29.7	29.6	29.5	30.3
		C	92.3	78.3	92.2	93.5	93.6	93.8	92.3	93.7	92.3	79.5	93.3	92.9	93.6	93.6	92.0	94.3
1	0.5	B	3.7	24.9	5.7	1.5	−4.0	−6.4	−6.1	−4.0	3.7	23.8	4.4	0.9	−4.0	−6.8	−7.1	−2.9
		S	17.7	15.2	19.3	21.2	32.5	32.1	31.9	33.1	17.7	14.8	18.8	22.2	32.5	32.0	31.9	32.9
		E	16.7	14.4	18.0	20.4	29.7	29.3	29.4	31.0	16.7	14.5	18.2	20.7	29.7	29.3	29.4	31.0
		C	92.3	56.2	90.9	94.5	93.6	92.3	92.5	93.7	92.3	60.8	93.0	93.6	93.6	92.3	92.0	94.5

			$X \sim MG$, $U^* \sim N(0,1)$								$X \sim MG$, $U^* \sim MG$							
			β_1				β_2				β_1				β_2			
r	σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
0	0.25	B	−1.7	18.9	8.5	2.1	1.1	−3.8	−3.9	0.6	−1.7	17.2	6.9	0.1	1.1	−3.4	−3.6	1.1
		S	13.4	11.0	13.2	14.7	19.0	18.9	18.9	20.2	13.4	11.3	12.8	14.5	19.0	18.8	19.1	20.0
		E	12.9	10.8	12.8	14.9	17.9	17.7	18.5	19.4	12.9	11.0	12.9	15.3	17.9	17.7	18.4	19.4
		C	94.7	56.6	87.0	94.2	93.7	93.4	94.1	94.2	94.7	63.0	89.9	95.5	93.7	93.2	93.2	95.2
0	0.5	B	−1.7	31.6	13.8	5.4	1.1	−7.0	−7.2	0.0	−1.7	28.7	10.6	1.5	1.1	−6.3	−6.6	0.9
		S	13.4	9.7	13.3	16.0	19.0	18.8	18.9	21.1	13.4	10.1	13.1	16.0	19.0	18.5	19.2	20.7
		E	12.9	9.5	13.2	16.2	17.9	17.6	18.7	20.3	12.9	10.0	13.3	17.0	17.9	17.7	18.5	20.5
		C	94.7	11.3	77.9	91.3	93.7	92.0	92.6	94.3	94.7	20.5	83.8	95.3	93.7	92.2	91.7	95.8
1	0.25	B	1.0	15.2	5.1	−0.8	−4.3	−5.9	−6.0	−3.7	1.0	15.0	4.5	−0.8	−4.3	−5.9	−5.4	−3.6
		S	19.9	17.9	20.3	22.5	33.3	33.0	32.3	34.1	19.9	17.8	19.7	22.5	33.3	33.0	32.4	34.1
		E	18.0	16.1	18.1	20.7	29.8	29.7	29.7	30.5	18.0	16.2	18.2	20.6	29.8	29.7	29.7	30.5
		C	91.9	79.0	90.1	92.6	91.6	91.8	92.4	92.4	91.9	79.6	91.7	92.7	91.6	92.1	92.2	92.5
1	0.5	B	1.0	26.2	9.2	−2.2	−4.3	−7.2	−9.3	3.3	1.0	25.2	5.9	−2.2	−4.3	−6.7	−6.7	−3.3
		S	19.9	16.3	20.1	25.1	33.3	32.0	30.6	34.8	19.9	15.7	20.2	25.1	33.3	32.0	32.4	34.8
		E	18.0	14.7	18.6	23.2	29.8	29.4	34.4	31.3	18.0	14.8	18.7	23.2	29.8	29.4	29.6	31.3
		C	91.9	51.3	86.9	93.1	91.6	92.3	92.2	92.4	91.9	55.6	90.5	93.1	91.6	92.0	91.7	92.3

Table 2: Simulation results based on 1000 replications for $r = 0$ and 1 with $n = 500$, unequal-length intervals and 90% right censoring on average. Here measurement error $U = \sigma_u U^*$. All entries are multiplied by 100. B $\equiv$ bias, S $\equiv$ standard deviation, E $\equiv$ estimated standard error, C $\equiv$ 95% coverage probability, $N \equiv$ Normal, $MG \equiv$ Modified Gamma, NM $\equiv$ No measurement error, NV $\equiv$ Naive, RC $\equiv$ Regression calibration, IM $\equiv$ Imputation method.

			$X \sim N$, $U^* \sim N(0,1)$								$X \sim N(0,1)$, $U^* \sim MG$							
			β_1				β_2				β_1				β_2			
r	σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
0	0.25	B	−1.3	11.6	0.5	0.9	3.4	2.3	2.3	2.6	−1.3	9.9	−1.4	−0.6	3.4	2.5	2.5	2.8
		S	14.8	13.7	15.4	15.0	30.9	31.0	30.9	31.0	14.8	14.1	15.9	15.3	30.9	30.9	30.9	31.0
		E	14.7	13.6	15.4	15.9	30.5	30.5	30.7	30.9	14.7	13.9	15.6	16.1	30.5	30.5	30.8	30.9
		C	94.6	84.8	95.4	96.5	95.7	96.0	95.9	95.6	94.6	86.5	95.3	96.7	95.7	96.1	95.6	95.9
0	0.5	B	−1.3	21.3	1.5	2.1	3.4	1.4	1.3	1.8	−1.3	18.2	−2.4	−0.8	3.4	1.7	1.8	2.1
		S	14.8	12.9	16.3	15.6	30.9	30.9	30.8	31.0	14.8	13.7	17.3	16.2	30.9	30.8	30.8	31.0
		E	14.7	12.8	16.2	16.9	30.5	30.4	31.4	31.2	14.7	13.3	16.6	17.4	30.5	30.4	30.6	31.3
		C	94.6	59.5	94.9	96.7	95.7	95.9	95.7	95.7	94.6	67.2	94.7	96.1	95.7	95.7	95.7	96.1
1	0.25	B	−1.3	11.7	0.6	−0.1	2.9	1.5	1.5	2.5	−1.3	10.3	−1.0	−1.4	2.9	1.7	1.7	2.8
		S	17.0	15.4	17.5	17.5	32.2	32.0	31.8	32.2	17.0	15.7	17.7	17.6	32.2	32.0	32.0	32.3
		E	16.9	15.5	17.5	18.2	32.1	31.9	32.0	32.5	16.9	15.7	17.7	18.4	32.1	31.9	32.1	32.5
		C	95.1	86.4	96.0	96.9	96.1	96.4	96.3	96.4	95.1	88.8	95.5	96.7	96.1	96.3	96.3	96.4
1	0.5	B	−1.3	21.6	1.8	0.4	2.9	0.4	0.4	2.1	−1.3	18.9	−1.6	−2.0	2.9	0.7	0.7	2.5
		S	17.0	14.4	18.2	18.5	32.2	31.9	31.7	32.3	17.0	15.0	18.9	18.7	32.2	31.8	31.8	32.3
		E	16.9	14.5	18.2	19.4	32.1	31.7	32.1	32.7	16.9	14.8	18.7	19.8	32.1	31.7	32.1	32.8
		C	95.1	65.7	95.5	96.8	96.1	95.9	96.1	96.4	95.1	71.9	95.0	96.8	96.1	96.0	95.4	96.1

			$X \sim MG$, $U^* \sim N(0,1)$								$X \sim MG$, $U^* \sim MG$							
			β_1				β_2				β_1				β_2			
r	σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
0	0.25	B	−4.2	21.7	11.7	2.0	5.4	5.0	4.9	5.6	−4.2	19.8	9.6	−0.7	5.4	5.0	4.9	5.5
		S	25.5	18.7	21.1	25.2	34.2	34.7	34.5	34.6	25.5	19.6	22.0	26.6	34.2	34.5	34.2	34.3
		E	24.2	19.1	21.5	25.7	32.2	32.2	32.4	32.6	24.2	19.7	22.2	27.0	32.2	32.2	32.4	32.6
		C	93.4	76.8	89.5	94.2	94.3	94.1	94.4	94.8	93.4	79.1	90.6	94.6	94.3	94.2	94.7	94.8
0	0.5	B	−4.2	35.6	19.2	5.2	5.4	4.7	4.6	5.4	−4.2	32.7	15.5	−1.6	5.4	4.7	4.6	5.4
		S	25.5	15.8	20.0	26.1	34.2	34.8	34.5	34.8	25.5	17.2	21.6	29.4	34.2	34.7	34.2	34.3
		E	24.2	16.5	20.7	27.4	32.2	32.2	32.3	32.8	24.2	17.5	22.0	30.3	32.2	32.2	32.6	32.9
		C	93.4	42.2	82.6	94.1	94.3	94.1	94.2	94.6	93.4	51.1	86.7	94.7	94.3	94.2	94.5	94.8
1	0.25	B	−4.6	19.8	9.6	−0.3	4.6	3.8	3.9	4.8	−4.6	18.6	8.3	−2.3	4.6	3.8	3.6	4.7
		S	25.4	19.2	21.6	26.3	33.4	33.6	33.4	33.6	25.4	19.8	22.3	27.3	33.4	33.4	33.1	33.4
		E	24.2	19.6	22.1	26.6	32.3	32.2	32.4	32.7	24.2	20.0	22.6	27.2	32.3	32.2	32.3	32.7
		C	93.4	79.2	91.3	94.7	95.3	95.1	95.2	95.5	93.4	81.3	92.1	95.4	95.3	94.3	94.7	94.6
1	0.5	B	−4.6	33.6	16.7	1.7	4.6	3.3	3.2	4.8	−4.6	31.4	14.0	−3.9	4.6	3.2	3.2	5.1
		S	25.4	16.3	20.6	28.0	33.4	33.6	33.4	33.8	25.4	17.4	21.8	30.7	33.4	33.4	33.1	33.6
		E	24.2	17.1	21.7	28.7	32.3	32.1	33.2	33.0	24.2	17.8	22.5	31.6	32.3	32.1	32.8	33.0
		C	93.4	48.5	85.7	95.1	95.3	94.9	95.2	95.2	93.4	56.3	88.3	95.8	95.3	94.0	94.6	94.8

Table 3: Simulation results based on 1000 replications with $n = 500$, unequal-length intervals and 90% right censoring on average. Measurement error $U = \sigma_u U^*$, and U^* follows the modified gamma distribution. All entries are multiplied by 100. B ≡ bias, S ≡ standard deviation, E ≡ estimated standard error, C ≡ 95% coverage probability, MN ≡ Mixture Normal, NM ≡ No measurement error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

		$X \sim MG$, $r = 2$							
		β_1				β_2			
σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM
0.25	B	−4.8	17.9	7.5	−3.0	4.8	4.0	3.9	5.1
	S	27.7	21.8	24.5	30.3	36.3	36.2	36.0	36.3
	E	26.3	21.9	24.7	30.3	35.5	35.2	35.4	35.8
	C	94.5	82.4	93.0	95.7	94.8	94.8	95.2	95.2
0.5	B	−4.8	32.4	13.0	−5.1	4.8	3.3	3.1	5.2
	S	27.7	18.0	24.1	34.4	36.3	36.6	36.0	36.6
	E	26.3	18.9	24.6	34.7	35.5	35.1	35.5	36.2
	C	94.5	55.5	89.5	95.9	94.8	94.8	95.1	95.6

Table 4: Results of the analysis of the ACTG data set (left panel) and the analysis of the ACTG data set where the values of the surrogate for CD4 were replaced by simulated data (right panel). Est ≡ Estimate, SE ≡ standard error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

		log(CD4)			Treatment			X			Treatment		
r	Method	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
0	NV	−2.72	0.50	< 0.01	0.75	0.31	0.016	−0.56	0.23	< 0.05	0.70	0.30	< 0.05
	RC	−3.29	0.60	< 0.01	0.74	0.31	0.016	−0.65	0.26	< 0.05	0.68	0.31	< 0.05
	IM	−3.16	0.62	< 0.01	0.73	0.31	0.019	−0.89	0.42	< 0.05	0.71	0.31	< 0.05
1	NV	−3.04	0.60	< 0.01	0.85	0.34	0.013	−0.64	0.25	< 0.05	0.73	0.35	< 0.05
	RC	−3.72	0.72	< 0.01	0.83	0.34	0.015	−0.69	0.28	< 0.05	0.69	0.34	< 0.05
	IM	−3.62	0.77	< 0.01	0.82	0.35	0.020	−0.96	0.46	< 0.05	0.77	0.34	< 0.05
2	NV	−3.39	0.70	< 0.01	0.93	0.38	0.015	−0.65	0.26	< 0.05	0.80	0.36	< 0.05
	RC	−4.14	0.85	< 0.01	0.93	0.38	0.015	−0.73	0.30	< 0.05	0.82	0.37	< 0.05
	IM	−4.05	0.93	< 0.01	0.92	0.40	0.021	−1.03	0.51	< 0.05	0.84	0.37	< 0.05