Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–27
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280215580668
smm.sagepub.com
SAGE

Semiparametric Bayesian analysis of censored linear regression with errors-in-covariates

Samiran Sinha and Suojin Wang

Abstract

The accelerated failure time (AFT) model is a well-known alternative to the Cox proportional hazard model for analyzing time-to-event data. In this paper we consider fitting an AFT model to right censored data when a predictor variable is subject to measurement errors. First, without measurement errors, estimation of the model parameters in the AFT model is a challenging task due to the presence of censoring, especially when no specific assumption is made regarding the distribution of the logarithm of the time-to-event. The model complexity increases when a predictor is measured with error. We propose a non-parametric Bayesian method for analyzing such data. The novel component of our approach is to model (1) the distribution of the time-to-event, (2) the distribution of the unobserved true predictor, and (3) the distribution of the measurement errors all non-parametrically using mixtures of the Dirichlet process priors. Along with the parameter estimation we also prescribe how to estimate survival probabilities of the time-to-event. Some operating characteristics of the proposed approach are judged via finite sample simulation studies. We illustrate the proposed method by analyzing a data set from an AIDS clinical trial study.

Keywords

Buckley–James estimator, Dirichlet process prior, functional approach, measurement errors, mixture distributions, posterior inference

1 Introduction

Right censored time-to-event data are often analyzed by fitting a Cox proportional hazard (CPH) model. Although fitting the CPH model and obtaining the estimate of the relative risk parameters via the partial likelihood method are easy, model parameter interpretation requires the understanding of instantaneous hazard. On the other hand, the accelerated failure time (AFT) model is easy to interpret. In the AFT model, the logarithm of the time-to-failure T is assumed

Texas A&M University, College Station, TX, USA

Corresponding author:
Suojin Wang, Department of Statistics, Texas A&M University, College Station, TX 77843, USA.
Email: sjwang@stat.tamu.edu

to be a linear function of the covariates and an error term which is assumed to be free from the covariates. That means, for the AFT model,

$$\log(T) = \beta_1^{\mathrm{T}} Z + \beta_2 X + e \tag{1}$$

where the error e is assumed to follow a distribution with finite variance and is assumed to be independent of the covariates $(Z^{\mathrm{T}}, X)^{\mathrm{T}}$. Here Z is assumed to be a vector of error-free covariates while the continuous scalar covariate X is not observed in the data. Instead, multiple replications of an erroneous unbiased surrogate W for X are observed in the data. By unbiasedness we mean $E(W|X) = X$, and by surrogate we mean $f(T|W, X, Z) = f(T|X, Z)$ (Carroll[1]). In the error-free case (i.e. when X is accurately observed) the regression parameters $\beta = (\beta_1^{\mathrm{T}}, \beta_2)^{\mathrm{T}}$ are difficult to estimate due to the presence of censoring especially when the distribution of e is left unspecified. There are several choices for fitting an AFT model to right censored data, such as Buckley–James estimating equations,[2] modified Buckley–James estimating equations proposed by Lai and Ying,[3] some more recent approaches proposed by Lin and his co-authors,[4,5] and the empirical likelihood approach of Zhou and Li.[6] Although our interest is in the semiparametric AFT model where e is left unspecified, the AFT model can be fitted assuming some flexible parametric models (generalized Gamma, log-logistic, splines, etc.) for e (Cox et al.[7]).

In the error-free AFT model context, Christensen and Johnson[8] first considered the Dirichlet process (DP) prior for nonparametric modeling of the time-to-event, and proposed an elegant semi-Bayesian approach for estimating survival curves and the finite dimensional regression coefficient. Later, Kuo and Mallick[9] considered a mixture of the DP prior on e, and Walker and Mallick[10] proposed to use the Polya tree prior on e and a noninformative prior on the regression coefficients β. The last two papers considered full Bayesian inferences using the Markov chain Monte Carlo (MCMC) method.

In this paper we consider fitting an AFT model (1) to right censored data when the scalar covariate X is measured with error and with repeated measurements at the baseline. The motivation comes from a clinical study on AIDS. One of the important indicators for the time to AIDS or death of HIV-infected people is the CD4 count at the baseline examination before any treatment starts. The true CD4 count cannot be measured. Therefore, multiple measurements of a surrogate variable for CD4 count at the baseline are considered as the erroneous measurements of the true CD4 count. The goal is to estimate the regression coefficients utilizing the erroneous measurements for CD4 counts. While errors-in-covariate are a common issue in clinical or observational studies, fitting an AFT model when the predictor is measured with error has received little attention from researchers. He et al.[11] proposed a simulation and extrapolation (SIMEX) approach for estimating model parameters when the time-to-event data are subject to right censoring and a covariate is measured with error. They assumed that (1) the distribution of e belongs to a known parametric family, and (2) the errors associated with the covariate follow a normal distribution. These assumptions limit the application of their SIMEX method. Another paper in this context is by Ma and Yin.[12] They considered a broader issue by proposing a novel method of handling covariate measurement errors in a semiparametric quantile regression model. However, they require that the censoring mechanism and the actual time-to-event are marginally independent.

In order to circumvent these issues we propose a general method where (1) we do not make any parametric assumption regarding the distribution of e, (2) we do not make any parametric assumption regarding the distribution of the unobserved covariate, (3) we do not make any parametric assumption regarding the distribution of the measurement errors U in W. All of these

three issues are handled by a novel application of the non-parametric Bayesian methods. In particular, in a likelihood framework, the distributions of e, X, and U are modeled non-parametrically using a mixture of a finite-dimensional Dirichlet process (FDDP), a special case of stick-breaking prior.[13] In addition to a non-parametric modeling of e, since our approach does not make any parametric assumption regarding the distributions of X and U, the method can be considered as a *functional approach* in view of the modern measurement error literature. Since we use a parametric prior on the unknown regression parameter β along with nonparametric prior models for the distributions of e, X, and U, we call the proposed method semiparametric. The novelty of the proposed approach lies in the robustness of the procedure through non-parametric modeling of several nuisance densities. When a distribution is modeled by a DP mixture of kernel densities (we have taken them to be normal kernels), the distribution is essentially modeled by a mixture of infinitely many kernel densities, where the mixing proportions and the parameters of the kernel densities are random. This structure of the prior model for a density, in principle, leads to a posterior that is weakly consistent for the true density (Theorems 5.6.1–5.6.3 of Ghosh and Ramamoorthi[14]). This posterior consistency not only holds when the true density is a mixture of normals, but also when the true density has a compact support, such as the uniform distribution. In our set-up, instead of using a DP, for computational convenience we use a FDDP as a close approximation of the DP. Since we are modeling three nuisance distributions non-parametrically, our results are generally robust towards the distributions of e, U, and X. In the simulation studies, we numerically show the robustness of the proposed method by considering different types of distributions for e, X, and U, and comparing with some partly semiparametric approaches. In the partly semiparametric methods, one of three nuisance infinite-dimensional parameters is treated parametrically, and the results show that lack of proper modeling of at least one nuisance parameter may result in biased estimates of the regression parameters.

Previously, Müller and Roeder[15] used a non-parametric Bayesian approach for handling errors in a covariate in case–control studies that do not involve censored subjects. Gustafson et al.[16] considered a parametric Bayesian method for handling errors in a covariate in case–control studies. Further, Sinha et al.[17] considered a non-parametric Bayesian approach for handling errors in a covariate in the logistic regression model while the effect of the covariate was modeled as a non-parametric function. However, to the best of the authors' knowledge, our current problem is unique in that no one has addressed it before. Overall, our non-parametric Bayesian approach is useful not only for estimating the regression parameters β, but also for estimating the survival probabilities and the quantiles of the failure time distribution.

A brief outline of the remainder of the article is as follows. Section 2 contains basic models and assumptions. Section 3 discusses likelihood and priors. Posterior computation and parameter estimation are given in Section 4. Section 5 outlines some other statistical inferences using the posterior samples. Sections 6 and 7 are devoted to simulation studies and the analysis of a real data set from an AIDS clinical trial study, respectively. Concluding remarks are given in Section 8. The details of the MCMC steps and some further data analysis are relegated to the appendix.

2 Basic models and assumptions

Suppose we observe the data $(V_i, \Delta_i, W_{ij}, j = 1, \ldots, m, Z_i),\ i = 1, \ldots, n$, where $V_i = \min(T_i, C_i)$, and the time-to-failure T_i is assumed to be independent of the censoring time C_i conditional on the observed covariates $(W_{i1}, \ldots, W_{im}, Z_i)$ and the binary variable $\Delta_i = I(T_i \leq C_i)$ denotes the censoring indicator. For non-parametric modeling of the measurement error distribution we require the

number of replications to be at least two (i.e. $m \geq 2$). Even for handling a more restrictive scenario, such as a symmetric error distribution, one needs $m \geq 2$ to identify the error distribution.[18] Without repeated measurements on W, one needs to specify the distribution of the error for any structural or functional approach. We assume that T_i follows model (1), and $e \sim F_e$ which is unknown. Furthermore, assume that Z_i is a vector of error-free covariates, and the surrogate variable $W_i^{\mathrm{T}} = (W_{i1}, \ldots, W_{im})$ is related to the unobserved latent variable X_i via the classical additive measurement error model

$$W_{ij} = X_i + U_{ij}, \quad \text{for } j = 1, \ldots, m, \ m \geq 2$$

where U_{ij} are independent and identically distributed (i.i.d.) following a mean zero distribution F_U with a finite variance and are independent of $(V_i, \Delta_i, X_i, Z_i)$. Furthermore, conditional on Z the unobserved X is assumed to follow a distribution $F_X(\cdot|Z)$ which is also unknown. It is known that the naive analysis of the data by replacing X_i by $\bar{W}_i = \sum_{j=1}^{m} W_{ij}/m$ will, in principle, yield a biased estimator of β, and consequently the estimator of the survival function is biased.[12,19]

It is worth mentioning that without measurement errors and assuming the response variables are subject to only right censoring, the Buckley–James estimator of β is obtained by solving

$$S(\beta; V, X, Z) = \sum_{i=1}^{n} \begin{pmatrix} Z_i \\ X_i \end{pmatrix} \{\tilde{T}_i(\beta) - \bar{\tilde{T}}(\beta) - (Z_i - \bar{Z})^{\mathrm{T}}\beta_1 - (X_i - \bar{X})\beta_2\} = 0$$

where

$$\tilde{T}_i(\beta) = \Delta_i \log(V_i) + (1 - \Delta_i)(Z_i^{\mathrm{T}}\beta_1 + X_i\beta_2) + \frac{\int_{e_i(\beta)}^{\infty} u \mathrm{d}\hat{F}_e(u, \beta)}{1 - \hat{F}_e(e_i(\beta), \beta)}, \quad \bar{\tilde{T}}(\beta) = \frac{1}{n}\sum_{i=1}^{n} \tilde{T}_i(\beta), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i, \quad \hat{F}_e(t, \beta) = 1 - \prod_{i:\, e_i(\beta) < t} \left[1 - \frac{\Delta_i}{\sum_{j=1}^{n} I\{e_j(\beta) \geq e_i(\beta)\}}\right]$$

and $e_i(\beta) = \log(V_i) - Z_i^{\mathrm{T}}\beta_1 - X_i\beta_2$. The estimating equation $S(\beta; V, X, Z) = 0$ is based upon the normal equations of the least squares method and is then adjusted for censoring (see Buckley and James[2] for details). The estimating function $S(\beta; V, X, Z)$ involves a non-smooth function $\hat{F}_e(\cdot, \beta)$ making the estimating function non-continuous and non-monotone in β. Commonly, in the traditional *functional approach* of handling covariate measurement errors where unobserved X is treated as an unknown constant, one seeks an estimating function $S^*(\beta; V, W, Z)$ such that $E\{S^*(\beta; V, W, Z)|V, X, Z\} = S(\beta; V, X, Z)$. However, due to the presence of $\hat{F}_e(t, \beta)$ in $\tilde{T}_i(\beta)$, it is not obvious how to construct such function $S^*(\beta; V, W, Z)$. Alternatively, for this problem with four infinite-dimensional nuisance parameters: (a) the distribution of e; (b) the distribution of the censoring process; (c) the distribution of X given Z; and (d) the distribution of the measurement errors, it would be interesting to investigate the existence and computational feasibility of an efficient estimator along the lines of Ma and Li[20] and Ma and Carroll.[21] The most challenging aspect will be handling the censoring process that may depend on Z. To circumvent these issues we propose a likelihood-based approach with only a few general regularity assumptions on these nuisance distributions, and statistical inferences are made using the MCMC method.

3 Likelihood and priors

In the Bayesian analysis, likelihood function takes a key role. For this purpose we assume that e, X, and U are absolutely continuous random variables and define $S_e(\cdot) = 1 - F_e(\cdot)$, $f_e(e) = dF_e(e)/de$, $f_X(x|Z) = \mathrm{d}F_X(x|Z)/\mathrm{d}x$, and $f_U(u) = \mathrm{d}F_U(u)/\mathrm{d}u$. Then the likelihood of the observed data ignoring the components related to the censoring is

$$L_{\text{obs}} = \prod_{i=1}^{n} \int S_e^{1-\Delta_i}(\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \beta_2 X_i) f_e^{\Delta_i}(\log(V_{i1}) - \beta_1^{\mathrm{T}} Z_i - \beta_2 X_i) f_X(X_i|Z_i) \prod_{j=1}^{m} f_U(W_{ij} - X_i)\, \mathrm{d}X_i$$

For non-parametric modeling of F_e, $F_X(\cdot|Z)$, F_U, oftentimes a DP mixture model is used that can essentially capture any shape for the distribution of the underlying variable. However, the computation involving a DP prior is time consuming, and it is proportional to the sample size. For efficient computation we shall use a FDDP prior. Before we describe the FDDP, we provide a general definition of the stick-breaking process.

A stick-breaking process is a random probability measure $\mathcal{P}$ defined as $\mathcal{P}(A) = \sum_{k=1}^{N} p_k \delta_{Y_k}(A)$ for a measurable set A. Here $\delta_{Y_k}(\cdot)$ denotes a measure concentrated at Y_k, Y_k are i.i.d. from a distribution H, N is the number of components, and p_k are random probabilities such that $0 \le p_k \le 1$ and $\sum_{k=1}^{N} p_k = 1$. Since p_k and Y_k are random, $\mathcal{P}(A)$ is also random. The name stick breaking comes due to the structure of random weights p_k, where

$$p_1 = V_1, \quad p_2 = (1 - V_1)V_2, \ldots, \quad p_k = (1 - V_1)\cdots(1 - V_{k-1})V_k$$

and $V_k \overset{\text{indep}}{\sim} \text{Beta}(a_k, b_k)$ are assumed to be independent of Y_k. This process allows finite and infinite values of N. As special cases it includes DP, Poisson-DP, Dirichlet-multinomial process, etc.[13] For a DP, $a_k = 1$, $b_k = \alpha$, and $N = \infty$, and it is denoted by $DP(\alpha H)$. A FDDP, denoted by $DP_N(\alpha H)$, has a finite number of components N and $(p_1, \ldots, p_N) \sim \text{Dirichlet}(\alpha/N, \ldots, \alpha/N)$.[22] Theorem 2 of Ishwaran and Zarepour[22] states that for any real valued measurable integrable function g, $\text{DP}_N(\alpha H)(g) \to \text{DP}(\alpha H)(g)$ in distribution as $N \to \infty$. They also described a convenient mechanism of selecting N, the maximum possible cluster size. We shall use N_e, N_u, and N_x to denote N for the FDDPs corresponding to e, U, and X, respectively.

Now we assume that

$$\begin{aligned} e_i|\theta_{ie} &\overset{\text{indep}}{\sim} \text{Normal}(\theta_{ie,1}, \theta_{ie,2}),\ \theta_{ie} = (\theta_{ie,1}, \theta_{ie,2}) \\ \theta_{ie}|P_e &\overset{\text{i.i.d.}}{\sim} P_e \text{ (a random distribution)} \\ P_e &\sim \text{DP}_{N_e}(\alpha_e H_{0e}), \quad i = 1, \ldots, n \end{aligned}$$

The random probability measure P_e follows a FDDP with the base probability measure H_{0e} on $R \times R^+$, and H_{0e} is viewed as the center of the random process P_e. The above assumption implies that for any measurable set A, the prior expectation of the probability that $\theta_e \in A$ under the probability measure $\text{DP}_{N_e}(\alpha_e H_{0e})$ is $E\{\text{pr}_{P_e}(\theta_e \in A)\} = \text{pr}_{H_{0e}}(\theta_e \in A)$, and $\text{var}\{\text{pr}_{P_e}(\theta_e \in A)\} = \text{pr}_{H_{0e}}(\theta_e \in A)\{1 - \text{pr}_{H_{0e}}(\theta_e \in A)\}(1 + \alpha_e/N_e)(1 + \alpha_e)^{-1}$. Thus, larger values of α_e lead to smaller variance of $\text{pr}_{P_e}(\theta_e \in A)$. Therefore, α_e can be interpreted as a precision parameter. We assume that under H_{0e}, $\theta_{e,2} \sim \text{IG}(a_{\sigma_e}, b_{\sigma_e})$ (IG $\equiv$ Inverse Gamma), and conditional on $\theta_{e,2}$, $\theta_{e,1} \sim \text{Normal}(m_e, \tau_e \theta_{e,2})$. Using the prior assumption on θ_{ie} we can now write

$$f_e(e|\, p_{1e}, \ldots, p_{N_e e}, Y_{1e}, \ldots, Y_{N_e e}) = \sum_{k=1}^{N_e} p_{ke} \frac{1}{\sqrt{2\pi\theta_{e,2}}} \exp\left\{-\frac{(e - \theta_{e,1})^2}{2\theta_{e,2}}\right\} \delta_{Y_{ke}}(\theta_e)$$

$$(p_{1e}, \ldots, p_{N_e e}) \sim \text{Dirichlet}(\alpha_e/N_e, \ldots, \alpha_e/N_e)$$
$$Y_{ke} \overset{\text{i.i.d.}}{\sim} H_{0e}$$

In Appendix A1 we give a brief discussion connecting the non-parametric kernel smoothing and DP mixture ideas for density estimation.

We model $F_X(\cdot|Z)$ as a FDDP mixture of normal distributions. In other words, we assume that

$$X_i|Z_i, \gamma_1, \theta_{ix} \overset{\text{indep}}{\sim} \text{Normal}(\theta_{ix,1} + Z_i^{\mathrm{T}}\gamma_1, \theta_{ix,2}),\ \theta_{ix} = (\theta_{ix,1}, \theta_{ix,2})^{\mathrm{T}}$$
$$\theta_{ix} \overset{\text{i.i.d.}}{\sim} P_x \text{ (a random distribution)}$$
$$P_x \sim \text{DP}_{N_x}(\alpha_x H_{0x})$$

where H_{0x} is the base probability measure on $R \times R^+$ and α_x is the precision parameter. Under H_{0x}, we assume that $\theta_{x,2} \sim \text{IG}(a_{\sigma_x}, b_{\sigma_x})$, and conditional on $\theta_{x,2}$, $\theta_{x,1} \sim \text{Normal}(m_x, \tau_x\theta_{x,2})$. An alternative statement of the above model is that $f_X(x|Z_i, \gamma_1) = \int f_{X,P_x}(x|Z_i, \gamma_1) d\text{DP}_{N_x}(P_x; \alpha_x H_{0x})$, where $f_{X,P_x}(x|Z_i, \gamma_1) = \int (1/\sqrt{2\pi\theta_{ix,2}}) \exp\{-(x - Z_i^T\gamma_1 - \theta_{ix,1})/2\theta_{ix,2}\}\, dP_x(\theta_{ix})$. If we knew the true P_x, say P_{x0}, we would model the distribution of X as $f_{X,P_{x0}}(\cdot)$ without requiring a FDDP prior on P_x. Furthermore,

$$f_X(x|Z, \gamma_1, p_{1x}, \ldots, p_{N_x x}, Y_{1x}, \ldots, Y_{N_x x}) = \sum_{k=1}^{N_x} p_{kx} \frac{1}{\sqrt{2\pi\theta_{x,2}}} \exp\left\{-\frac{(x - \theta_{x,1} - Z_i^{\mathrm{T}}\gamma_1)^2}{2\theta_{x,2}}\right\} \delta_{Y_{kx}}(\theta_x)$$
$$(p_{1x}, \ldots, p_{N_x x}) \sim \text{Dirichlet}(\alpha_x/N_x, \ldots, \alpha_x/N_x)$$
$$Y_{kx} \overset{\text{i.i.d.}}{\sim} H_{0x}$$

Now we model $F_U(\cdot)$. Although symmetric measurement error is a commonly used assumption,[18,23] for $m \geq 2$ one can still identify the distributions of U and X as long as U has mean zero and finite variance.[24] Therefore, in our development we use the weaker assumption that the distribution of $U_{ij} = W_{ij} - X_i$ has mean zero and finite variance and model it as a finite-dimensional centered Dirichlet process (CDP) mixture of a normal kernel.[25] For notational convenience we shall use the index $l = m(i-1) + j$ to identify duplex (i, j) for $i = 1, \ldots, n$, $j = 1, \ldots, m$. Thus, we consider

$$U_{ij}|\theta_{lu} \overset{\text{indep}}{\sim} \frac{1}{\sqrt{2\pi\theta_{lu,2}}} \exp\left\{-\frac{(U_{ij} - \theta_{lu,1})^2}{2\theta_{lu,2}}\right\},\ \theta_{lu} = (\theta_{lu,1}, \theta_{lu,2})^{\mathrm{T}}$$
$$\theta_{lu}|P_u \overset{\text{i.i.d.}}{\sim} P_u \text{ (a random distribution with mean zero)}$$
$$P_u \sim \text{CDP}_{N_u}(\alpha_u H_{0u})$$

More specifically, the zero mean of U_{ij} is ensured by the fact that $E(\theta_{lu}|P_u) = 0$ when P_u is randomly drawn from the $\text{CDP}_{N_u}(\alpha_u H_{0u})$. Now we write

$$f_U(W - X|\, p_{1u}, \ldots, p_{N_u u}, Y_{1u}, \ldots, Y_{N_u u}) = \sum_{k=1}^{N_u} p_{ku} \frac{1}{\sqrt{2\pi\theta_{u,2}}} \exp\left\{-\frac{(W - X - \theta_{u,1})^2}{2\theta_{u,2}}\right\} \delta_{Y_{ku}}(\theta_u)$$

$$
\begin{aligned}
(p_{1u}, \ldots, p_{N_u u}) &\sim \text{Dirichlet}(\alpha_u/N_u, \ldots, \alpha_u/N_u) \\
Y_{ku} &= Y^*_{ku} - \mu^*_u \\
\mu^*_u &= \sum_{k=1}^{N_u} p_{ku} Y^*_{ku} \\
Y^*_{ku} &\overset{\text{i.i.d.}}{\sim} H_{0u}
\end{aligned}
$$

Here H_{0u} is the base probability measure on $R \times R^+$. Under H_{0u} we assume that the second component of Y^*_{ku}, $Y^*_{ku,2} \sim \text{IG}(a_{\sigma_u}, b_{\sigma_u})$, and conditional on $Y^*_{ku,2}$, the first component of Y^*_{ku}, $Y^*_{ku,1} \sim \text{Normal}(m_u, \tau_u Y^*_{ku,2})$. We further assume that a priori $\beta_1 \sim \text{Normal}(\mu_{\beta_1}, \Sigma_{\beta_1})$, $\beta_2 \sim \text{Normal}(\mu_{\beta_2}, \sigma^2_{\beta_2})$, and $\gamma_1 \sim \text{Normal}(\mu_{\gamma_1}, \Sigma_{\gamma_1})$. On α_e, α_u, and α_x we put Gamma$(a_{\alpha_e}, b_{\alpha_e})$, Gamma$(a_{\alpha_u}, b_{\alpha_u})$, and Gamma$(a_{\alpha_x}, b_{\alpha_x})$ priors, respectively. Also, we assume that a priori $\tau_e \sim \text{IG}(g_e, h_e)$, $\tau_u \sim \text{IG}(g_u, h_u)$, and $\tau_x \sim \text{IG}(g_x, h_x)$. We use IG$(\eta_e, \zeta_e)$, IG$(\eta_u, \zeta_u)$, and IG$(\eta_x, \zeta_x)$ priors on b_{σ_e}, b_{σ_u}, and b_{σ_x}, respectively.

Further notation is needed for posterior computation. Define $\Theta^{\text{T}}_e = (\theta_{1e}, \ldots, \theta_{ne})$, $\Theta^{\text{T}}_x = (\theta_{1x}, \ldots, \theta_{nx})$, and $\Theta^{\text{T}}_u = (\theta_{1u}, \ldots, \theta_{Mu})$, where $M = n \times m$. Let ϕ_e be an $N_e \times 2$ matrix that contains N_e distinct elements of Θ_e. Similarly define ϕ_x and ϕ_u. For updating random elements of Θ_e, define configuration indicators $s^{\text{T}}_e = (s_{1e}, \ldots, s_{ne})$ such that $s_{ie} = j$ if $\theta_{ie} = \phi_{je}$, $j = 1, \ldots, N_e$, $i = 1, \ldots, n$. Also define the size of the jth cluster $n^e_j = \sum_{i=1}^n I(s_{ie} = j)$, for $j = 1, \ldots, N_e$. Thus, $0 \le n^e_j \le n$ and $\sum_{j=1}^{N_e} n^e_j = n$. Similarly, define $n^x_j = \sum_{i=1}^n I(s_{ix} = j)$ that satisfies $0 \le n^x_j \le n$ and $\sum_{j=1}^{N_x} n^x_j = n$, and $n^u_j = \sum_{l=1}^M I(s_{lu} = j)$ with $0 \le n^u_j \le M$ and $\sum_{j=1}^{N_u} n^u_j = M$.

Since knowing s_e and ϕ_e is equivalent to knowing Θ_e, in the MCMC method Θ_e is updated via resampling s_e and ϕ_e. Similarly, s_x, s_u can be defined, and Θ_x is updated by resampling s_x and ϕ_x and Θ_u is updated by resampling s_u and ϕ_u. From now on, we shall write θ_{ie} as $\phi^{\text{T}}_{s_{ie}e} = (\phi_{s_{ie}e,1}, \phi_{s_{ie}e,2})$. Similarly, we shall use $\phi_{s_{ix}x}$ and $\phi_{s_{lu}u}$ instead of θ_{ix} and θ_{lu}.

4 Posterior computation and parameter estimation

Inference regarding the parameters are made from the respective posterior distribution. Using the MCMC method we draw random numbers from the posterior distribution.

Define $T^*_i = \log(T_i)$. When $\Delta_i = 0$ the value of T^*_i is unknown. Then it will be treated as an unknown parameter in our Bayesian computation and resampled conditional on the observed data and the other parameters. The important feature of the following MCMC technique is that all of the conditional distributions except the one related to α_e, α_u, and α_x are in the form of standard well-known distributions. We follow Ishwaran and James[13] for updating the parameters related to the stick-breaking priors.

In the MCMC method we repeat the steps 1–8 (given in Appendix A2) for a large number (e.g. 20,000) of iterations. Along with the unknown parameters and hyperparameters we shall resample all X_i for $i = 1, \ldots, n$, and T^*_i for those i where $\Delta_i = 0$.

After discarding the first few thousand samples (e.g. 5,000) as burn-in (see, e.g., Cowles and Carlin[27]), we shall consider the remaining MCMC samples as the random numbers from the joint posterior distribution of the parameters. These sampled observations will be used for calculating parameter estimates and other statistics.

5 Other statistical inferences based on posterior samples

5.1 Estimation of survival probabilities

In addition to the estimation of β in the AFT model (1), another key objective in this context is to estimate the survival probability $\mathrm{pr}(T \geq t_0 | X_0, Z_0, \Theta)$ for given t_0, X_0, and Z_0, where Θ denotes the set of all parameters. Let $\pi(\Theta|\boldsymbol{D})$ be the generic notation for the posterior distribution of Θ given the observed data $\boldsymbol{D}$. Then a random number from the posterior distribution of the survival probability can be obtained by computing $\mathrm{pr}(T \geq t_0 | X_0, Z_0, \Theta)$ when Θ is randomly drawn from $\pi(\Theta|\boldsymbol{D})$. A Bayes estimator of this survival probability is the posterior mean $\widehat{\mathrm{pr}}(T \geq t_0 | X_0, Z_0, \Theta) = \int \mathrm{pr}(T \geq t_0 | X_0, Z_0, \Theta)\pi(\Theta|\boldsymbol{D})\, d\Theta$ that can be estimated by taking the Monte Carlo average of

$$1 - \sum_{k=1}^{N_e} p_{ke} \Phi \left\{ \frac{\log(t_0) - \beta_1^{\mathrm{T}} Z_0 - \beta_2 X_0 - \theta_{ke,1}}{\sqrt{\theta_{ke,2}}} \right\}$$

over B (e.g. $B = 10{,}000$ or more) MCMC samples of $(\beta_1^{\mathrm{T}}, \beta_2, \theta_{ke}^{\mathrm{T}}, p_{ke})$ drawn from their joint posterior distribution.

5.2 Model selection

In clinical studies we are also interested in testing hypotheses, such as H_0: $\beta_2 = 0$ versus H_1: $\beta_2 \neq 0$. In the Bayesian set-up one can conduct hypothesis testing by calculating the Bayes factor $BF = \mathrm{pr}(\boldsymbol{D}|H_1)\mathrm{pr}(H_0)/\{\mathrm{pr}(\boldsymbol{D}|H_0)\mathrm{pr}(H_1)\}$, where $\mathrm{pr}(H_0)$ and $\mathrm{pr}(H_1)$ are the prior probabilities of H_0 and H_1, $\mathrm{pr}(\boldsymbol{D}|H_k) = \int \mathrm{pr}(\boldsymbol{D}|\Theta_k)\pi(\Theta_k)\, d\Theta_k$ for $k = 0, 1$ with Θ_k being the finite- and infinite-dimensional parameter under the hypothesis H_k, and $\pi(\Theta_k)$ is the corresponding prior distribution. Usually BF larger than 10 indicates a strong evidence for the alternative model specified by H_1. Following Newton and Raftery[28] we shall calculate the marginal probability or likelihood $\mathrm{pr}(\boldsymbol{D}|H_k)$ using the harmonic mean:

$$\mathrm{pr}(\boldsymbol{D}|H_k) = \left[E_{\pi(\Theta_k|\boldsymbol{D})} \left\{ \frac{1}{L(\boldsymbol{D}|\Theta_k)} \right\} \right]^{-1}$$

that can be estimated by $[B^{-1} \sum_{b=1}^{B} L^{-1}(\boldsymbol{D}|\Theta_k^{(b)})]^{-1}$, where $(\Theta_k^{(1)}, \ldots, \Theta_k^{(B)})$ are B MCMC samples from the posterior distribution $\pi(\Theta_k|\boldsymbol{D})$. Since under the Bayesian set-up unobserved X_i is also considered as an unknown parameter, the likelihood is, under H_0,

$$\begin{aligned} L(\boldsymbol{D}|\Theta_0) = & \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\phi_{s_{ie}e,2}}} \exp\left[-\frac{\{\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \phi_{s_{ie}e,1}\}^2}{2\phi_{s_{ie}e,2}} \right] \right)^{\Delta_i} \\ & \times \left\{ 1 - \Phi\left(\frac{\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \phi_{s_{ie}e,1}}{\sqrt{\phi_{s_{ie}e,2}}} \right) \right\}^{1-\Delta_i} \\ & \times \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\phi_{s_{lu}u,2}}} \exp\left\{ -\frac{(W_{ij} - X_i - \phi_{s_{lu}u,1})^2}{2\phi_{s_{lu}u,2}} \right\} \frac{1}{\sqrt{2\pi\phi_{s_{ix}x,2}}} \exp\left\{ -\frac{(X_i - \phi_{s_{ix}x,1} - Z_i^{\mathrm{T}}\gamma_1)^2}{2\phi_{s_{ix}x,2}} \right\} \end{aligned}$$

and similarly under H_1,

$$L(\boldsymbol{D}|\Theta_1) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\phi_{s_{ie}e,2}}} \exp\left[-\frac{\{\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \beta_2 X_i - \phi_{s_{ie}e,1}\}^2}{2\phi_{s_{ie}e,2}} \right] \right)^{\Delta_i}$$
$$\times \left\{ 1 - \Phi\left(\frac{\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \beta_2 X_i - \phi_{s_{ie}e,1}}{\sqrt{\phi_{s_{ie}e,2}}} \right) \right\}^{1-\Delta_i}$$
$$\times \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\phi_{s_{lu}u,2}}} \exp\left\{ -\frac{(W_{ij} - X_i - \phi_{s_{lu}u,1})^2}{2\phi_{s_{lu}u,2}} \right\} \frac{1}{\sqrt{2\pi\phi_{s_{ix}x,2}}} \exp\left\{ -\frac{(X_i - \phi_{s_{ix}x,1} - Z_i^{\mathrm{T}}\gamma_1)^2}{2\phi_{s_{ix}x,2}} \right\}$$

In the real data analysis this Bayes factor approach will also be used for model comparisons where we compute marginal probability of $\boldsymbol{D}$ under a given model. One numerical problem in calculating $\sum_{b=1}^{B} 1/L(\boldsymbol{D}|\Theta_k^{(b)})$ is that oftentimes $l_k^{(b)} = \log\{L(\boldsymbol{D}|\Theta_k^{(b)})\}$ is a large positive or negative number in the order of 1000, making it impossible to calculate the quantity. Thus, we adopt the following approximation using the Taylor series expansion:

$$\frac{B}{\sum_{b=1}^{B} \exp(-l_k^{(b)})} \approx \frac{B}{\sum_{b=1}^{B} \{\exp(-\mu_{l_k}) - (l_k^{(b)} - \mu_{l_k})\exp(-\mu_{l_k}) + 0.5(l_k^{(b)} - \mu_{l_k})^2 \exp(-\mu_{l_k})\}}$$
$$= \exp(\mu_{l_k})(1 + 0.5\sigma_{*,k}^2)^{-1}$$

where $\mu_{l_k} = \sum_{b=1}^{B} l_k^{(b)}/B$ and $\sigma_{*,k}^2 = \sum_{b=1}^{B} (l_k^{(b)} - \mu_{l_k})^2/B$. Hence, based on this approximation $\log\{\mathrm{pr}(\boldsymbol{D}|H_k)\} \approx \mu_{l_k} - \log(1 + 0.5\sigma_{*,k}^2)$.

6 Simulation studies

6.1 Simulation design

While a violation of model assumptions may lead to biased estimates of the parameters, the amount of bias depends on the degree of violation, and intricate interplay among the several model assumptions and their violations. We conducted simulation experiments with several scenarios, but due to limited space we shall discuss mainly two scenarios that clearly show the advantage of the proposed method in terms of bias whereas for the other scenarios the semiparametric and partly semiparametric (we shall discuss it in the next paragraph) approaches are comparable. We point out that inconsistency of partly semiparametric methods are manifested via large bias in the parameter estimates. We simulated a cohort of size $n = 200$ and 300, by simulating $Z \sim \mathrm{Normal}(0, 1)$, and then X and e in the following scenarios. Finally, we obtained T by setting $\log(T) = 1 + Z + 2X + e$. Two ($m = 2$) erroneous measurements W_{i1} and W_{i2} were obtained by adding U_{i1} and U_{i2} with X_i, for $i = 1, \ldots, n$. For scenario 1, $e \sim \mathrm{Exponential}(1)$, $X \sim 0.2Z + (1/3)\mathrm{Normal}(0, 0.7^2) + (2/3)\mathrm{Normal}(2, 0.3^2)$, and $U \sim \mathrm{Gamma}(1, 1) - 1$. To create approximately 25% and 50% censored data the censoring variable was simulated as $C = 0.5X^2 + \mathrm{Unif}(0, 2,000)$ and $C = 0.5X^2 + \mathrm{Unif}(0, 400)$, respectively. For scenario 2, $e \sim t_3$, $X \sim \{\mathrm{Gamma}(6, 0.5) - 3\}/1.22$, $U \sim \mathrm{Normal}(0, 0.71^2)$, and C followed two distributions: $C = 0.5Z^2 + 0.5X^2 + \mathrm{Unif}(0, 40)$ and $C = 0.5Z^2 + 0.5X^2 + \mathrm{Unif}(0, 5)$, for 25% and 50% censoring, respectively. For both scenarios we

took $\text{var}(U)/\text{var}(X) \times 100\,\% = 50\,\%$ to closely match with the noise-to-signal ratio of the real data. Note that in these scenarios C violates our assumption by making it depend on unobserved X variable. The results when C does not depend on X are similar, thus is omitted. Also, we have intentionally taken non-normal distribution for e, U, and X, to show the robustness of our approach. For completeness, we also ran additional simulations with normal distributions for X, e, and U. The results indicate that SP, SPPE, SPPU, and SPPX worked equally well in this case. The details are omitted.

6.2 Methods for the analyses

The observed data were $(V_i, \Delta_i, Z_i, W_{ij},\ j = 1, 2,\ i = 1, \ldots, n)$, and X was no longer used in the analysis stage. The first method is the naive method, where we used $\bar{W}_i = \sum_{j=1}^{2} W_{ij}/2$ in place of X_i in the Buckley–James method and used an existing program to compute the estimates (`bj` within the `R` package `rms`), and this approach will be referred to as the naive method. Next, we analyzed the data using the regression calibration (RC) approach. Here we assume that $\bar{W}_i|X_i \sim \text{Normal}(X_i, \sigma^2_{w|x})$ and $X_i|Z_i \sim \text{Normal}(\gamma_0 + \gamma_1 Z_i, \sigma^2_{x|z})$ which imply $X_i|\bar{W}_i, Z_i \sim \text{Normal}[(\sigma^{-2}_{x|z} + \sigma^{-2}_{w|x})^{-1}\{(\gamma_0 + \gamma_1 Z_i)\sigma^{-2}_{x|z} + \bar{W}_i\sigma^{-2}_{w|x}\}, (\sigma^{-2}_{x|z} + \sigma^{-2}_{w|x})^{-1}]$. We then analyzed the data with X_i being replaced by $\hat{X}_i = (1/\hat{\sigma}^2_{x|z} + 1/\hat{\sigma}^2_{w|x})^{-1}\{(\hat{\gamma}_0 + \hat{\gamma}_1 Z_i)/\hat{\sigma}^2_{x|z} + \bar{W}_i/\hat{\sigma}^2_{w|x}\}$ in the Buckley–James method. Here $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are the estimated coefficients obtained by regressing $\bar{W}_i$ on Z_i, $i = 1, \ldots, n$, $\hat{\sigma}^2_{w|x} = (2n)^{-1}\sum_{i=1}^{n}(W_{i1} - W_{i2})^2$ and $\hat{\sigma}^2_{x|z} = (n-2)^{-1}\sum_{i=1}^{n}(\bar{W}_i - \hat{\gamma}_0 - \hat{\gamma}_1 Z_i)^2 - \hat{\sigma}^2_{w|x}$. Next, we analyzed the data using the proposed method which is referred to as the semiparametric method (SP) where we treated all three infinite-dimensional nuisance parameters non-parametrically.

One may analyze these data sets using several parametric and partly semiparametric approaches. In principle, these approaches may produce biased results when the parametric assumptions are violated. For the sake of comparisons, here we also analyzed the data sets using three partly semiparametric approaches denoted by SPPE, SPPU, SPPX, where two of the three nuisance parameters were treated non-parametrically while the third was treated parametrically. The SPPE model is the same as SP model except that e is modeled parametrically as $e \sim \text{Normal}(\theta_{e,1}, \theta_{e,2})$, $\theta_{e,2} \sim \text{IG}(a_{\sigma_e}, b_{\sigma_e})$, $\theta_{e,1}|\theta_{e,2} \sim \text{Normal}(m_e, \tau_e\theta_{e,2})$, $\tau_e \sim \text{IG}(g_e, h_e)$, $b_{\sigma_e} \sim \text{IG}(1, 1)$. The SPPU model is the same as the SP model except that U is modeled parametrically as $U = W - X \sim \text{Normal}(0, \theta_u)$, $\theta_u \sim \text{IG}(a_{\sigma_u}, b_{\sigma_u})$, $b_{\sigma_u} \sim \text{IG}(1, 1)$. The SPPX model is the same as the SP except that X given Z is modeled parametrically as $X|Z \sim \text{Normal}(\theta_{x,1} + \gamma_1^{\mathrm{T}} Z_i, \theta_{x,2})$, $\theta_{x,2} \sim \text{IG}(a_{\sigma_x}, b_{\sigma_x})$, $\theta_{x,1}|\theta_{x,2} \sim \text{Normal}(m_x, \tau_x\theta_{x,2})$, $b_{\sigma_x} \sim \text{IG}(1, 1), \tau_x \sim \text{IG}(g_x, h_x)$.

Although the general modeling technique and inference method of SP are described in Sections 3 and 4, some necessary details are described in this paragraph. Before each analysis we re-centered the W values by subtracting the sample mean of all of the W from W itself. For the Bayesian methods (SP, SPPE, SPPU, SPPX), posterior inference was made through the MCMC method with 20,000 iterations using the following priors and hyperparameters. We took the RC estimates of β_1 and β_2 as the prior mean of β_1 and β_2, and used five as the prior variance for β_1 and β_2. We used $\hat{\gamma}_1$ as the prior mean for γ_1 and used 2 times the corresponding standard error as the prior standard deviation of γ_1. We set $a_{\alpha_e} = a_{\alpha_u} = a_{\alpha_x} = b_{\alpha_e} = b_{\alpha_u} = b_{\alpha_x} = 1$ that lead to an Exponential(1) prior for the precision parameters that covers a wide range of plausible values. Then we set $\eta_e = \zeta_e = \eta_u = \zeta_u = \eta_x = \zeta_x = 1$, $g_e = h_e = g_u = h_u = g_x = h_x = 1$. By setting these parameters of the inverse gamma distribution to both be one, we allow very large finite variances of the distributions of the hyperparameters that in turn result in estimates that are less affected by

the prior choice. In addition, we set $m_e = m_u = m_x = 0$ that are involved in the base probability measure of the FDDPs. Consider the case of $\theta_e = (\theta_{e,1}, \theta_{e,2})^T$. Conditional on $\theta_{e,2}$ and τ_e, under the base measure of the FDDP, we assumed $\theta_{e,1} \sim \text{Normal}(m_e, \tau_e\theta_{e,2})$. For any choice of m_e, this normal distribution can cover a wide spectrum of values for appropriate choice of τ_e and $\theta_{e,2}$. Thus, even a completely wrong choice of m_e is compensated by flexible values of τ_e and $\theta_{e,2}$ supported by their almost non-informative prior distributions. Thus, following the same analogy for m_u and m_x as well, the choice of m_e, m_u, m_x does not need to be perfect. We initialize $\tau_e = \tau_u = \tau_x = 1$, $\alpha_e = \alpha_u = \alpha_x = 2$, $X_i = \hat{X}_i$, and set β to the RC estimates of β and γ_1 to $\hat{\gamma}_1$.

The "error" in approximating a DP by a FDDP can be measured via the L_1 difference between two marginal probabilities, one corresponding to the FDDP and the other corresponding to the DP. Had we observed $e_1, \ldots, e_n$, then based on our model assumption on the distribution of e of the AFT model we could write $f_{N_e}(e_1, \ldots, e_n) \equiv \int \{\int \prod_{i=1}^n f(e_i|\theta_{ie})\, dP_e(\theta_{ie})\} d\text{DP}_{N_e}(P_e; \alpha_e H_{0e})$, and had we assumed that $P_e \sim \text{DP}(\alpha_e H_{0e})$, then $f_\infty(e_1, \ldots, e_n) \equiv \int \{\int \prod_{i=1}^n f(e_i\theta_{ie})\, dP_e(\theta_{ie})\} d\text{DP}(P_e; \alpha_e H_{0e})$. Ishwaran and Zarepour[22] showed that $\int |f_{N_e}(e_1, \ldots, e_n) - f_\infty(e_1, \ldots, e_n)|\, de_1 \cdots de_n \approx 4n \exp\{-(N_e - 1)/\alpha_e\}$. With $N_e = 50$, $\alpha_e = 1$, and $n = 1,036$ (the sample size for the data) the error is 2.173×10^{-18}, and for $\alpha_e = 2$, it is 9.489×10^{-08}. Our numerical experience shows that all precision parameters α_e, α_u, and α_x are usually smaller than 1, and as long as the error $(= 4n \exp\{-(N_e - 1)/\alpha_e\})$ is in the order of 10^{-5}, the results do not vary much with the choice of N_e. We have used the same analogy in choosing N_u and N_x both equal to 50.

6.3 Results

We report the estimated bias (Bias), empirical standard deviations of the estimates (SD), and mean squared error (MSE) based on 500 replications. Tables 1 and 2 contain results for scenarios 1 and 2, respectively. All methods show finite sample bias, and for the naive, RC, and partly semiparametric methods usually bias (see the estimates of β_2) increases with the censoring percentage. However, the naive estimates are much more biased than any other methods. Although the RC method is generally inconsistent, for small values of β_2 ($\beta_2 < 1$), and relatively small measurement errors, the RC estimates are pretty satisfactory (not presented here). However, in scenarios 1 and 2, RC estimates are quite biased. Considering the bias of all methods in different scenarios, the SP method becomes the most robust approach. In Table 1, other than the naive and RC methods, the largest bias is seen in SPPX and then SPPU. However, SPPE where a normal model is used for e turned out to be comparable with the our proposed model SP. An intuitive explanation is that larger values of T involving larger values of e are likely to get censored more often. The tail probabilities of e, needed for handling censored data, are moderately well-approximated by tail probabilities of a normal distribution when the true distribution of e is Exponential (1). Table 2 clearly shows that if the model assumption regarding e is grossly violated that could affect adversely the parameter estimates, which is evident in large bias in the SPPE method. The second largest bias is shown in SPPX. SPPU performs as good as SP, as the normal distribution assumption on U holds true in this scenario.

Of course, as a price for the robustness, the SD of the SP method is often slightly larger than the competing methods but the MSEs are relatively comparable. The SD of the estimates decreases with sample size, and it increases with the percentage of censoring. The difference between the SP and other partly semiparametric approaches is not as large as the difference between the SP and RC methods. The reason lies in the fact that in other partly semiparametric methods two of the three nuisance parameters are non-parametrically modeled that reduces the degree of model violations. Of course, the bias reduction achieved in the SP compared with other partly semiparametric approaches indicates the supremacy of flexible modeling of all three nuisance parameters.

Table 1. Results of the simulation study where $\log(T) = Z + 2X + 1 + e$, $e \sim$ Exponential(1), $X \sim 0.2Z + (1/3)$ Normal(0,0.7^2) + (2/3)Normal(2,0.3^2), $C = 0.5X^2 + $ Unif(0,2,000) for 25% censoring, $C = 0.5X^2 + $ Unif(0,400) for 50% censoring, and $U \sim$ Gamma(1,1) − 1. Here $N_e = N_u = N_x = 50$.

		Bias	SD	MSE	Bias	SD	MSE
Method	Parameter	$n = 200$ & 25% censoring			$n = 200$ & 50% censoring		
Naive	β_1	0.152	0.122	0.038	0.127	0.138	0.035
	β_2	–0.529	0.102	0.291	0.590	0.122	0.183
RC	β_1	0.012	0.142	0.020	–0.024	0.162	0.026
	β_2	0.153	0.209	0.067	0.327	0.245	0.167
SP	β_1	0.001	0.098	0.009	0.011	0.108	0.012
	β_2	0.007	0.091	0.008	0.025	0.099	0.011
SPPE	β_1	–0.001	0.098	0.009	0.011	0.109	0.012
	β_2	0.009	0.093	0.009	0.027	0.100	0.011
SPPU	β_1	0.003	0.132	0.017	0.011	0.144	0.021
	β_2	0.035	0.122	0.016	0.074	0.137	0.024
SPPX	β_1	0.012	0.104	0.011	0.004	0.117	0.013
	β_2	0.076	0.117	0.019	0.159	0.128	0.042
		$n = 300$ & 25% censoring			$n = 300$ & 50% censoring		
Naive	β_1	0.156	0.095	0.034	0.134	0.112	0.034
	β_2	–0.529	0.083	0.287	0.409	0.095	0.078
RC	β_1	0.018	0.106	0.011	–0.019	0.121	0.021
	β_2	0.140	0.161	0.045	0.316	0.184	0.142
SP	β_1	0.008	0.076	0.006	0.019	0.089	0.012
	β_2	0.009	0.068	0.005	0.027	0.075	0.009
SPPE	β_1	0.004	0.081	0.007	0.019	0.091	0.012
	β_2	0.012	0.072	0.005	0.028	0.077	0.010
SPPU	β_1	0.003	0.100	0.010	0.014	0.110	0.016
	β_2	0.034	0.097	0.010	0.065	0.106	0.022
SPPX	β_1	0.017	0.084	0.007	0.011	0.095	0.013
	β_2	0.070	0.088	0.013	0.149	0.095	0.032

Prompted by a referee's comment, we re-ran the simulation study for scenario 1 with $N_e = N_u = N_x = 100$. The results are presented in Table 3. A close comparison between Tables 1 and 3 reveals that there are no qualitative differences in these results.

In addition to the simulation studies described above, we conducted a small-scale simulation to compare the performance of the proposed method with the SIMEX approach. Here we took the semiparametric AFT model where e was left unspecified. In the first SIMEX approach (refer to as SIMEX1) we considered one of the two measurements of W as the erroneous measurement and estimated the measurement error variance with $\sum_{i=1}^{n}(W_{i1} - W_{i2})^2/(2n)$. SIMEX1 does not use all the available data. It is thus likely to lead to more bias. In the second SIMEX (refer to as SIMEX2) we used $\bar{W}_i = (W_{i1} + W_{i2})/2$ as the erroneous measurement and estimated the measurement error variance with $\sum_{i=1}^{n}(W_{i1} - W_{i2})^2/(4n)$. We used symmetric and asymmetric measurement error distributions, and used two different values for the measurement error variance $\mathrm{var}(U) = 0.5$ and $\mathrm{var}(U) = 1$. For comparisons we also present the naive and the RC along with our SP method. The results are given in Table 4. For obvious reasons SIMEX1 is worse than SIMEX2. When $\mathrm{var}(U) = 0.5$ the performance of SIMEX2 and SP are similar. However, for $\mathrm{var}(U) = 1$ the bias in

Table 2. Results of the simulation study where $\log(T) = Z + 2X + 1 + e$, and $e \sim t_3$, $X \sim \{\text{Gamma}(6,0.5) - 3\}/1.22$, $C = 0.5Z^2 + 0.5X^2 + \text{Unif}(0,40)$ for 25% censoring, $C = 0.5Z^2 + 0.5X^2 + \text{Unif}(0,5)$ for 50% censoring, and $U \sim \text{Normal}(0,0.7^2)$. Here $N_e = N_u = N_x = 50$.

		Bias	SD	MSE	Bias	SD	MSE
Method	Parameter	$n = 200$ & 25% censoring			$n = 200$ & 50% censoring		
Naive	β_1	–0.008	0.134	0.018	–0.007	0.159	0.025
	β_2	–0.466	0.142	0.238	–0.488	0.189	0.275
RC	β_1	–0.008	0.140	0.019	–0.007	0.164	0.027
	β_2	–0.069	0.197	0.044	–0.097	0.253	0.074
SP	β_1	0.002	0.133	0.017	0.009	0.152	0.023
	β_2	0.023	0.210	0.045	0.034	0.271	0.074
SPPE	β_1	0.009	0.143	0.021	0.033	0.172	0.031
	β_2	0.085	0.232	0.061	0.129	0.309	0.113
SPPU	β_1	0.003	0.134	0.018	0.008	0.153	0.023
	β_2	0.039	0.208	0.045	0.056	0.281	0.082
SPPX	β_1	–0.001	0.132	0.017	0.005	0.148	0.022
	β_2	–0.036	0.192	0.038	–0.052	0.238	0.059
		$n = 300$ & 25% censoring			$n = 300$ & 50% censoring		
Naive	β_1	–0.009	0.111	0.012	–0.016	0.133	0.018
	β_2	–0.467	0.111	0.231	–0.496	0.156	0.271
RC	β_1	–0.010	0.115	0.013	–0.017	0.137	0.019
	β_2	–0.076	0.150	0.028	–0.116	0.209	0.057
SP	β_1	–0.002	0.106	0.011	–0.010	0.126	0.016
	β_2	–0.013	0.139	0.019	–0.022	0.195	0.038
SPPE	β_1	0.006	0.115	0.013	0.022	0.141	0.023
	β_2	0.071	0.173	0.035	0.118	0.251	0.076
SPPU	β_1	–0.001	0.106	0.011	–0.010	0.127	0.016
	β_2	0.006	0.143	0.020	–0.000	0.201	0.040
SPPX	β_1	–0.007	0.104	0.011	–0.009	0.125	0.015
	β_2	–0.072	0.134	0.023	–0.097	0.191	0.045

SIMEX2 is much larger than SP. Large finite sample bias is a reflection of possible inconsistency of SIMEX. Although the bias of the SIMEX estimates seem to be not much affected by the non-normal measurement error, SIMEX is shown to be consistent only in a handful of cases with normal measurement errors and correct extrapolating function.

It is seen that while SP greatly reduces the bias, it is also accompanied by larger variance compared with the naive or the RC approach. This phenomenon is expected since any method that takes into account measurement errors in a covariate would generally result in larger uncertainty in the parameter estimators than the methods that fail to consider the measurement error issue in the analysis. Also, the variances of the estimators increase with the measurement error variance (see Table 4), and so are the MSEs. In our simulation, the bias of the inconsistent methods (such the naive and RC) is overwhelmingly larger than their corresponding variances, resulting in larger MSEs for the naive and RC methods compared with SP. Of course, there is no guarantee that SP would always have smaller MSE than the inconsistent approaches; eventually it all depends on whether the bias over weights the estimation variance (uncertainty) that in turn depends on the

Table 3. Results of the simulation study where $\log(T) = Z + 2X + 1 + e$, $e \sim$ Exponential(1),$X \sim 0.2Z + (1/3)$ Normal(0,0.7^2) + (2/3)Normal(2,0.3^2), $C = 0.5X^2 +$ Unif(0,2,000) for 25% censoring, $C = 0.5X^2 +$ Unif(0,400) for 50% censoring, and $U \sim$ Gamma(1,1) − 1. Here $N_e = N_u = N_x = 100$.

		Bias	SD	MSE	Bias	SD	MSE
Method	Parameter	$n = 200$ & 25% censoring			$n = 200$ & 50% censoring		
Naive	β_1	0.143	0.122	0.035	0.119	0.135	0.032
	β_2	–0.521	0.109	0.282	–0.400	0.118	0.174
RC	β_1	0.005	0.145	0.021	–0.036	0.162	0.027
	β_2	0.152	0.193	0.060	0.329	0.223	0.158
SP	β_1	0.013	0.097	0.009	0.017	0.104	0.011
	β_2	–0.016	0.146	0.022	0.012	0.124	0.015
SPPE	β_1	–0.001	0.108	0.012	0.009	0.113	0.013
	β_2	0.007	0.132	0.017	0.029	0.112	0.013
SPPU	β_1	0.016	0.127	0.016	0.019	0.136	0.019
	β_2	–0.012	0.108	0.012	0.048	0.120	0.017
SPPX	β_1	0.011	0.110	0.012	–0.002	0.123	0.015
	β_2	0.077	0.127	0.022	0.177	0.140	0.051

measurement error variance. Lastly, we would like to point out that although MSEs are presented in all of the tables, MSE may not be a good measure to compare consistent and inconsistent estimators.

7 Analysis of the data from an AIDS clinical trial study

This data set comes from a randomized, double-blind trial on AIDS known as ACTG 175 study. One of the study aims is to understand the effect of several antiretoviral drugs on human immunodeficiency virus-1 (HIV-1) infected people who had no history of an AIDS-defining illness other than minimal mucocutaneous Kaposi's sarcoma (see Hammert et al.[29] for details).

Subjects were randomly assigned to one of the four therapies, 600 mg of zidovudine, 600 mg of zidovudine plus 400 mg of didanosine, 600 mg of zidovudine plus 2.25 mg of zalcitabine, and 400 mg of didanosine. In our analysis, the event is the development of AIDS or death, and T is defined as the time (in days) from the start of the treatment to the occurrence of the event. According to the ACTG 175 study, for this group of patients, AIDS and death were considered as the primary end points as both were related to at least 50% decline in CD4 counts.[29]

For our analysis we considered only $n = 1,036$ subjects who did not have antiretroviral treatment before this trial. Among them 262, 257, 260, 257 subjects received the above 4 treatments, respectively. These subjects had two (i.e. $m = 2$) baseline measurements of CD4 counts prior to the start of their treatment. Among the 1036 subjects 85 experienced the above event and the median and average follow-up time were approximately 27 and 32 months, respectively.

We fit model (1) to this data set, where the logarithm of the actual CD4 count at the baseline minus 5.89 is considered as X. The choice of 5.89 is to make the distribution centered around 0. Note that the exact CD4 count in the blood is impossible to measure mainly due to constant movement of these cells between blood and tissues. In addition, within a short time span (a few days) small changes may occur in the CD4 count due to physical activity, stress, good night's sleep, etc. Therefore, the two baseline measurements are considered to be two erroneous measurements W_{i1} and W_{i2} for X_i, $i = 1, \ldots, 1036$. The estimated noise-to-signal ratio is $\widehat{\text{var}}(U)/\widehat{\text{var}}(X) \times 100\,\% = 42\,\%$.

Table 4. Results of the simulation study where $\log(T) = Z + 2X + 1 + e$, $e \sim$ Normal(0,1), $X \sim 0.2Z + (1/3)$ Normal(0,0.7^2) + (2/3)Normal(2,0.3^2), $C = 0.5X^2 +$ Unif(0,500) for 25% censoring. Here $N_e = N_u = N_x = 50$.

		Bias	SD	MSE	Bias	SD	MSE
Method	Parameter	$\sigma_u^2 = 0.5$			$\sigma_u^2 = 1$		
$U \sim$ Normal(0,σ_u^2)							
Naive	β_1	0.095	0.118	0.023	0.159	0.135	0.043
	β_2	–0.291	0.098	0.094	–0.515	0.103	0.277
RC	β_1	0.014	0.124	0.016	0.019	0.155	0.024
	β_2	0.109	0.144	0.033	0.184	0.221	0.083
SIMEX1	β_1	0.045	0.144	0.022	0.122	0.165	0.043
	β_2	–0.095	0.141	0.029	–0.322	0.153	0.127
SIMEX2	β_1	0.011	0.122	0.015	0.046	0.144	0.023
	β_2	–0.021	0.121	0.015	–0.091	0.148	0.030
SP	β_1	0.016	0.132	0.017	0.030	0.155	0.025
	β_2	0.020	0.123	0.015	0.011	0.153	0.023
$U \sim \sigma_u\{$Gamma $(1,1) - 1\}$							
Naive	β_1	0.085	0.119	0.022	0.141	0.137	0.039
	β_2	–0.295	0.100	0.097	–0.517	0.113	0.280
RC	β_1	0.004	0.125	0.015	0.001	0.155	0.024
	β_2	0.105	0.136	0.029	0.180	0.209	0.076
SIMEX1	β_1	0.007	0.143	0.020	0.055	0.169	0.031
	β_2	–0.092	0.169	0.037	–0.300	0.221	0.139
SIMEX2	β_1	–0.003	0.122	0.015	0.013	0.144	0.021
	β_2	–0.027	0.114	0.014	–0.094	0.144	0.029
SP	β_1	0.004	0.117	0.013	0.009	0.132	0.017
	β_2	0.022	0.105	0.011	0.015	0.157	0.024
$U \sim \sigma_u\{0.5$Normal $(0.9,0.5^2) + 0.5$Normal$(-0.9,0.5^2)\}$							
Naive	β_1	0.098	0.119	0.023	0.165	0.137	0.046
	β_2	–0.305	0.093	0.102	–0.538	0.095	0.299
RC	β_1	0.013	0.123	0.015	0.018	0.152	0.023
	β_2	0.116	0.146	0.035	0.195	0.233	0.093
SIMEX1	β_1	0.052	0.142	0.022	0.132	0.165	0.044
	β_2	–0.104	0.127	0.027	–0.349	0.129	0.139
SIMEX2	β_1	0.013	0.123	0.015	0.058	0.143	0.022
	β_2	–0.022	0.119	0.015	–0.100	0.144	0.031
SP	β_1	0.008	0.125	0.015	0.015	0.138	0.019
	β_2	0.032	0.111	0.013	0.037	0.126	0.017

The three dummy variables corresponding to the four treatments are considered to be error-free covariates Z, where 600 mg of zidovudine was considered as the reference category. We analyzed the data using NV, RC, SP, SPPE, SPPU, and SPPX, and the results are presented in Table 5. For SP we used $N_e = N_u = N_x = 50$. For RC, the 95% confidence intervals were calculated based on 1000 bootstrap samples.

Based on the 95% confidence intervals and credible intervals all methods indicate that log(CD4) has a statistically significant effect on the time-to-event. More importantly, after adjusting for the

Table 5. Results for the ACTG AIDS clinical trial data. For the naive Buckley–James method the 95% interval refers to the Wald-type confidence interval whereas for the RC method the 95% interval refers to the percentile interval based on 1000 bootstrap samples. For the Bayesian methods the 95% intervals refer to the equal tail credible intervals. For the Bayesian methods we present the posterior mean of the parameters as the estimates. Here Z, Z + D, Z + Z, and D stand for zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively.

Method	Z + D (Ref: Z)	Z + Z (Ref: Z)	D (Ref: Z)	log(CD4)
Accelerated failure time model				
NV	0.333	0.411	0.263	1.001
	(–0.042, 0.708)	(0.013, 0.808)	(–0.101, 0.625)	(0.566, 1.433)
RC	0.332	0.407	0.265	1.217
	(0.044, 0.645)	(0.080, 0.755)	(–0.049, 0.554)	(0.563, 2.094)
SP	0.406	0.512	0.350	1.360
	(0.054, 0.771)	(0.164, 0.889)	(0.019, 0.702)	(0.864, 1.897)
SPPE	0.414	0.514	0.355	1.355
	(0.066, 0.777)	(0.161, 0.897)	(0.023, 0.701)	(0.866, 1.898)
SPPU	0.410	0.507	0.352	1.417
	(0.059, 0.776)	(0.139, 0.891)	(0.016, 0.693)	(0.902, 1.985)
SPPX	0.408	0.513	0.356	1.394
	(0.071, 0.780)	(0.156, 0.888)	(0.020, 0.700)	(0.879, 1.952)
Piecewise exponential model				
NVPE	–0.801	–1.010	–0.778	–1.927
	(–1.365, –0.262)	(–1.617, –0.443)	(–1.326, –0.237)	(–2.575, –1.284)
RCPE	–0.801	–1.005	–0.783	–2.32
	(–1.372, –0.267)	(–1.607, –0.432)	(–1.321, –0.246)	(–3.093, –1.542)
PCPE	–0.809	–1.014	–0.795	–2.528
	(–1.388, –0.264)	(–1.629, –0.430)	(–1.334, –0.252)	(–3.413, –1.651)

measurement errors, the estimate of the coefficient for CD4 counts, β_2, is quite different in the SP method from the naive estimate. Clearly, for β_2, the naive estimate is closer to zero than the estimates from the other methods, a trend that is also observed in the simulation studies. The results of SP also indicate that compared with zidovudine, the other three therapies have statistically significant effect on delaying the time-to-event. This result is consistent with the findings of Hammer et al.[29] Figure 1 shows the estimated survival probabilities and 95% pointwise credible intervals based on SP for each treatment group when CD4 counts were 232 and 539, the approximate 10th and 90th quantiles of baseline CD4 measurements of the subjects. We found that between the competing hypotheses, H_0: $\beta_2 = 0$ versus H_1: $\beta_2 \neq 0$, the data unequivocally support H_1 as the Bayes factor was much larger than 10. Note that the analysis of this data set by SP with 60,000 MCMC iterations took approximately 3 minutes on a 2.8 GHz Intel Xenon X5560 processor.

We also analyzed the data using a piecewise exponential (PE) model, i.e. we assumed that the hazard of the time-to-event is $\lambda(t|X, Z) = \lambda_s \exp(\beta_1^{\mathrm{T}} Z + \beta_2 X)$ for $t_{s-1} \leq t < t_s$, $s = 1, \ldots, q$, where the time axis is partitioned into $[t_0, t_1), \ldots, [t_{q-2}, t_{q-1})$, $[t_{q-1}, t_q)$, with $t_0 = 0$ and $t_q = \infty$. First we carried out a naive analysis (NVPE) where we fit the PE model by replacing X_i by $\bar{W}_i$. Then we conducted a RC analysis (RCPE) by fitting the PE model where X_i was replaced by $\hat{X}_i$ defined in Section 6.2.

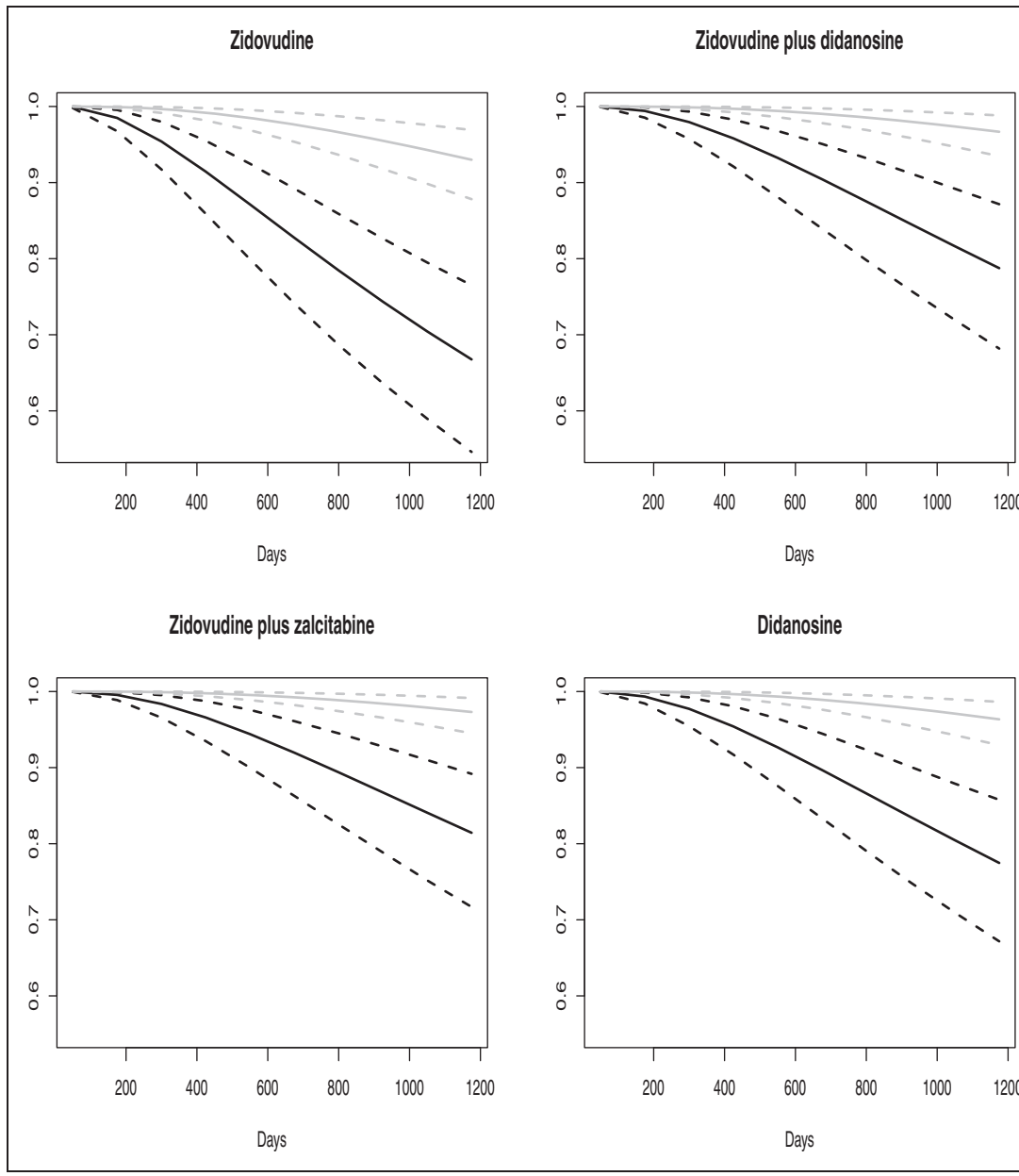

Figure 1. Survival probabilities and 95% pointwise credible intervals for baseline CD4 counts 232 (darker) and 539 (lighter).

Third, we fit the PE model with a parametric correction for the measurement errors (PCPE) where we assumed that X given Z followed a Normal$(\gamma_0 + \gamma_1^{\mathrm{T}} Z, \sigma_x^2)$ and $U \sim \text{Normal}(0, \sigma^2)$. The likelihood of the data is

$$L = \prod_{i=1}^{n} \int \left\{ \sum_{s=1}^{q} \lambda_s I(t_{s-1} \le V_i < t_s) \exp(\beta_1^{\mathrm{T}} Z_i + \beta_2 X_i) \right\}^{\Delta_i} \exp\left\{ - \sum_{s=1}^{q} \lambda_s I(t_{s-1} \le V_i) \exp(\beta_1^{\mathrm{T}} Z_i + \beta_2 X_i) \right\}$$
$$\times \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{ - \frac{(X_i - \gamma_0 - \gamma_1^{\mathrm{T}} Z_i)^2}{2\sigma_x^2} \right\} \times \frac{1}{(2\pi\sigma_u^2)^{m/2}} \exp\left\{ - \sum_{j=1}^{m} \frac{(W_{ij} - X_i)^2}{2\sigma_u^2} \right\} \mathrm{d}X_i$$

In all three methods the parameters were estimated in a Bayesian framework using the MCMC method. Here we assumed a priori $\lambda_s \sim \text{Gamma}(a_\lambda, b_\lambda)$, $\beta_1 \sim \text{Normal}(\mu_{\beta_1}, \Sigma_{\beta_1})$, $\beta_2 \sim \text{Normal}(\mu_{\beta_2}, \sigma_{\beta_2}^2)$, $\gamma_0 \sim \text{Normal}(\mu_{\gamma_0}, \sigma_{\gamma_0}^2)$, $\gamma_1 \sim \text{Normal}(\mu_{\gamma_1}, \Sigma_{\gamma_1})$, $\sigma_x^2 \sim \text{IG}(a_{\sigma_x}, b_{\sigma_x})$, and $\sigma_u^2 \sim \text{IG}(a_{\sigma_u}, b_{\sigma_u})$. The MCMC details are given in Appendix A3. In particular, we took $q = 5$ and partitioned the time axis as $[0, t_{0.2}), [t_{0.2}, t_{0.4}), [t_{0.4}, t_{0.6}), [t_{0.6}, t_{0.8}), [t_{0.8}, \infty)$, where t_r denotes the rth quantile of the observed failure times. The prior means of β were the RC estimates of the CPH model and 5 was used as the prior variance for all parameters. We set $\mu_{\gamma_0} = \hat{\gamma}_0$, $\mu_{\gamma_1} = \hat{\gamma}_1$, where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ denote the estimated intercept and partial slopes for the linear regression of $\bar{W}_i$ on Z_i. Two times the square of the corresponding standard errors were used as the prior variance. Finally, we used $a_{\sigma_x} = a_{\sigma_u} = b_{\sigma_x} = b_{\sigma_u} = 1$. The results are given in the lower panel of Table 5. These results, like the SP analysis, also indicate that the treatments are significantly associated (based on the 95% credible interval) with the time-to-event. In particular, compared with zidovudine, the other treatments reduce the hazard of the event. In addition, log(CD4) is significantly negatively associated with the hazard of the time-to-event, a consistent finding with the SP analysis.

Since SP, SPPE, SPPU, SPPX, and PCPE are all Bayesian methods, we were able to compare these approaches using marginal probabilities pr$(\boldsymbol{D}|\mathcal{M})$ where $\mathcal{M}$ stands for a generic model. Figure 2 shows the boxplot of $l_{\mathcal{M}}^{(b)}$ based on the MCMC samples after discarding the first 10,000 burn-in samples. The range of its values clearly indicates that a straightforward estimate of pr$(\boldsymbol{D}|\mathcal{M})$ using the harmonic mean is not possible. Therefore, we adopted the approximation given in Section 5.2 and obtained log{pr$(\boldsymbol{D}|\mathcal{M})$} as 3306.63, 3292.64, 3022.74, 2937.14, and –171.93 for SP, SPPE, SPPU, SPPX, and PCPE, respectively. The above results along with equal prior probability for each of the model in Bayes factor calculations indicate that the SP model is the best and is closely followed by the SPPE model. This explains why the β_2 estimates in these two methods are so close.

8 Conclusions

In this paper we proposed a non-parametric Bayesian method for fitting the AFT model to a right censored data when a covariate is measured with error. We believe that our approach is the first attempt to solve this problem in a non-parametric framework. While we non-parametrically treat the three components (the stochastic noise of the model for the time-to-event, the distribution of the latent unobserved true covariate, and the distribution of the zero mean measurement errors), the computation is simple and easily programmable due to the novel application of the stick-breaking priors. Due to non-parametric modeling of all nuisance distributions the proposed method outperforms the naive, RC, SIMEX, and other partly semiparametric methods in our simulation studies.

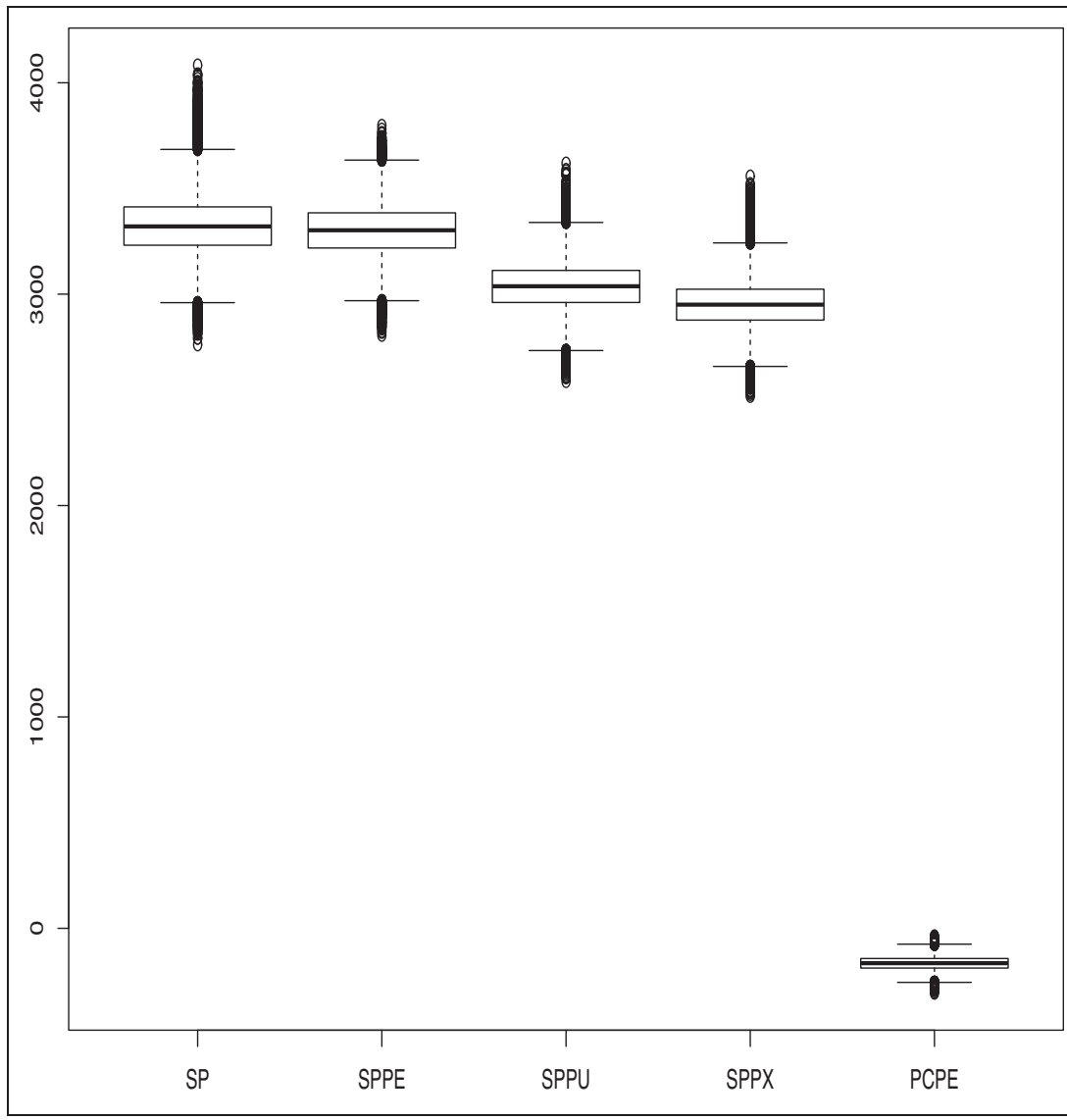

Figure 2. Boxplot of the logarithm of the complete data likelihood given the parameter values in the MCMC iterations for different models.

In this work, for notational simplicity we have assumed the number of the replications of the surrogate for the true covariate to be the same across all of the subjects (i.e. $m_i = m$). This assumption can be relaxed by using some more intense notations with some general regularity conditions on m_i. Furthermore, in principle, the proposed method can be extended to handle interval censored data.[30] For this purpose, in the posterior computations one should generate T_i^* from a normal distribution which is truncated on both sides. In this work we used the Monte Carlo estimates for estimating survival probabilities. We believe that using the importance sampling

method with proper importance weights one may improve the efficiency of the estimator. We have focused on time-invariant covariate. However, the measurement error issue may arise in a time-varying covariate; see, for example, Veronesi et al.[31] who considered the RC and SIMEX methods for handling such a covariate in a Cox regression model. Another interesting paper in this area is by Crowther et al.[32] who considered joint modeling of survival outcome using a semiparametric Cox model and longitudinally measured prognostic biomarkers using a linear mixed model. It is worth investigating how the nonparametric approaches considered in this paper could be implemented in these settings. The computational code can be obtained from the authors upon request. We are also creating an R-package for practitioners to use that will be available through our website.

Finally, we briefly discuss the efficiency property of the estimator. A semiparametrically efficient estimator is defined in the class of regular asymptotic linear (RAL) estimators that achieves the efficiency *bound* (Tsiatis,[33] p. 27). The *bound* is defined as the supremum of the most efficient RAL estimators of the parametric submodels that is a subset of the semiparametric class of models, and the true model belongs to the parametric submodels. For a parametric model where standard regularity conditions hold, the MLE produces the most efficient RAL estimator. By construction our estimator is not a RAL estimator. Therefore, it is generally difficult to compare it with the corresponding efficiency bound. However, like the Cramer–Rao lower bound, there exists a Bayesian Cramer–Rao lower bound.[34] It is worth exploring how a Bayesian minimax estimator can be constructed with that lower bound.

Acknowledgment

The authors gratefully acknowledge the constructive comments from the referees that led to a significant improvement of the manuscript.

Funding

This research was partially supported by the NIH (grant number R03CA176760).

References

1. Carroll RJ. *Measurement error in epidemiologic studies. Encyclopedia of Biostatistics*. New York: John Wiley & Sons, 1997.
2. Buckley J and James I. Linear regression with censored data. *Biometrika* 1979; **66**: 429–436.
3. Lai TL and Ying Z. Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann Statist* 1991; **19**: 1370–1402.
4. Zeng D and Lin DY. Efficient estimation for the accelerated failure time model. *J Am Statist Assoc* 2007; **102**: 1387–1396.
5. Jin Z, Lin DY, Wei LJ and Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika* 2003; **90**: 341–353.
6. Zhou M and Li G. Empirical likelihood analysis of the Buckley–James estimator. *J Multivar Anal* 2008; **99**: 649–664.
7. Cox C, Chu H, Schneider MF and Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statist Med* 2007; **26**: 4252–4374.
8. Christensen R and Johnson W. Modelling accelerated failure time with a Dirichlet process. *Biometrika* 1988; **75**: 693–704.
9. Kuo L and Mallick B. Bayesian semiparametric inference for the accelerated failure-time model. *Can J Statist* 1997; **25**: 457–472.
10. Walker S and Mallick BK. A Bayesian semiparametric accelerated failure time model. *Biometrics* 1999; **55**: 477–483.
11. He W, Yi GY and Xiong J. Accelerated failure time models with covariates subject to measurement Error. *Statist Med* 2007; **26**: 4817–4832.
12. Ma Y and Yin G. Censored quantile regression with covariate measurement errors. *Statistica Sinica* 2011; **21**: 949–971.
13. Ishwaran H and James LF. Gibbs sampling methods for stick breaking priors. *J Am Statist Assoc* 2001; **96**: 161–173.
14. Ghosh JK and Ramamoorthi RV. *Bayesian Nonparametrics*. New York: Springer, 2003.
15. Müller P and Roeder K. A Bayesian semiparametric model for case–control studies with errors in variables. *Biometrika* 1997; **84**: 523–537.
16. Gustafson P, Le ND and Vallee M. A Bayesian approach to case–control studies with errors in covariables. *Biostatistics* 2002; **3**: 229–243.
17. Sinha S, Mallick BK, Kipnis V and Carroll RJ. Semiparametric Bayesian analysis of nutritional

epidemiology data in the presence of measurement error. *Biometrics* 2010; **66**: 444–454.
18. Delaigle A, Hall P and Meister A. On deconvolution with repeated measurements. *Ann Statist* 2008; **36**: 665–685.
19. Carroll RJ, Ruppert D, Stefanski LA and Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edn. Boca Raton, FL: Chapman and Hall/CRC Press, 2006.
20. Ma Y and Li R. Variable selection in measurement error models. *Bernoulli* 2010; **16**: 274–300.
21. Ma Y and Carroll RJ. Locally efficient estimators for semiparametric models with measurement error. *J Am Statist Assoc* 2006; **101**: 1465–1474.
22. Ishwaran H and Zarepour M. Exact and approximate sum representations for the Dirichlet process. *Can J Statist* 2002; **30**: 269–283.
23. Hall P and Ma Y. Measurement error models with unknown error structure. *J R Statist Soc Ser B* 2007; **69**: 429–446.
24. Li T and Vuong Q. Nonparametric estimation of the measurement error model using multiple indicators. *J Multivar Anal* 1998; **65**: 139–165.
25. Yang M, Dunson DB and Baird D. Semiparametric Bayes hierarchical models with mean and variance constraints. *Computat Statist Data Anal* 2010; **54**: 2172–2186.
26. Escobar MD and West M. Bayesian density estimation and inference using mixtures. *J Am Statist Assoc* 1995; **90**: 577–588.
27. Cowles MK and Carlin BP. Markov chain monte carlo convergence diagnostics: A comparative review. *J Am Statist Assoc* 1996; **91**: 883–904.
28. Newton MA and Raftery AE. Approximate Bayesian inference with the weighted likelihood bootstrap. *J R Statist Soc Ser B* 1994; **56**: 3–26.
29. Hammer SM, Katzenstein DA, Hughes MD, et al. 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* 1996; **335**: 1081–1090.
30. Huang J and Wellner J. Interval censored survival data: a review of recent progress. In *Proceedings of the first Seattle symposium in biostatistics: survival analysis* (*Lecture Notes in Statistics*, vol. 123). Berlin: Springer, 1997; 123–169.
31. Veronesi G, Ferrario MM and Chambless LE. Comparing measurement error correction methods for rate-of-change exposure variables in survival analysis. *Statist Meth Med Res* 2011; **22**: 583–597.
32. Crowther MJ, Lambert PC and Abrams KR. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. *BMC Med Res Methodol* 2013; **13**: 146.
33. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York: Springer, 2006.
34. Gill RD and Levit BY. Applications of the van Trees inequality: a Bayesian Cramér–Rao bound. *Bernoulli* 1995; **1**: 59–79.
35. Gelfand A, Hills SE, Racine-Poon A and Smith AM. Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Statist Assoc* 1990; **85**: 972–985.

Appendix

A1 A discussion on density estimation referenced in Section 3

Along the lines of Escobar and West,[26] we discuss the connection between the nonparametric kernel density estimation and the Dirichlet process mixture of the Gaussian kernels. Had we observed $e_1, \ldots, e_n$, the density at a new value e_{n+1} based on the normal kernel would be $f_e(e_{n+1}|e_1, \ldots, e_n) = (nh\sqrt{2\pi})^{-1} \sum_{i=1}^{n} \exp\{-(e_{n+1} - e_i)^2/2h^2\}$, for a bandwidth $h > 0$. This is a mixture of normal kernels centered around each observed value of the variable. In our nonparametric Bayesian context, conditional on the hyperparameters, the predictive density at e_{n+1} is

$$\begin{aligned} f(e_{n+1}|e_1, \ldots, e_n) &= \int \frac{1}{\sqrt{2\pi\theta_{(n+1)e,2}}} \exp\left\{ -\frac{(e_{n+1} - \theta_{(n+1)e,1})^2}{2\theta_{(n+1)e,2}} \right\} \mathrm{d}\pi(\theta_{(n+1)e}, \alpha_e|e_1, \ldots, e_n) \\ &= \int \left[\sum_{k=1}^{N_e} \frac{n_k^e + N_e^{-1}\alpha_e}{\alpha_e + n} \times \frac{I(n_k^e > 0)}{\sqrt{2\pi\phi_{ke,2}}} \exp\left\{ -\frac{(e_{n+1} - \phi_{ke,1})^2}{2\phi_{ke,2}} \right\} \right. \\ &\quad \left. + \frac{\alpha_e(1 - N_e^{-1}\sum_{k=1}^{N_e} I(n_k^e > 0))}{\alpha_e + n} t_{2a_{\sigma_e}}(m_e, \nu) \right] \mathrm{d}\pi(\phi_e, \alpha_e|e_1, \ldots, e_n) \end{aligned} \tag{A1}$$

where $t_{2a_{\sigma_e}}(m_e, \nu)$ denotes a t distribution with mode m_e, scale $\nu = \sqrt{(1 + \tau_e)/a_{\sigma_e} b_{\sigma_e}}$, and degrees of freedom $2a_{\sigma_e}$, and $\pi(\theta_{(n+1)e}, \alpha_e|e_1, \ldots, e_n)$ denotes the posterior distribution of the parameters. Some essential distinctions are to be made. First, unlike fixed h that helps to determine the weight of each observation to the density at e_{n+1} in the kernel density estimation, α_e in (A1) varies and it plays the role of a smoothing parameter. In particular, α_e has a dual role, one in the weights of the mixing kernels, and two in determining the number of non-empty clusters (where $n_k^e > 0$) also known as

effective number of mixing components. Unlike kernel density estimation, the effective number of mixing components varies in our predictive density. Usually a large α_e assigns more weight on $t_{2a_{\sigma_e}}(m_e, v)$, the distribution derived from the base probability measure, and tends to create more non-empty clusters. Of course the number of non-empty clusters is also governed by the observed data, resulting in some traded values of α_e where the posterior probabilities become large. We use a prior distribution on α_e and numerically obtain samples from its posterior distribution by sampling α_e from its conditional density given all other parameters and data in every MCMC iteration. Alternatively, in a semi-Bayesian approach, one may determine α_e by solving an estimating equation characterizing a relation between α_e and the number of non-empty clusters in every iteration of the MCMC sample.[17] Second, unlike kernel density, the centers of the normal distributions in (A1) are random rather than fixed at the observed values of the variable. Third, the overall density estimate in (A1) is shrunk towards the baseline prior density $t_{2a_{\sigma_e}}(m_e, v)$.

A2 Details of the MCMC steps referenced in Section 4

Step 0:

Initialize α_e, α_u, and α_x. Then initialize the values of $\beta_1, \beta_2, \gamma_1, T_i^*, s_{ie}, \phi_e, X_i, s_{ix}, \phi_x, s_{iu}$ and ϕ_u^*, for $j = 1 \ldots, m,\ i = 1, \ldots, n$ and draw $(p_{1u}, \ldots, p_{N_u u})$ from Dirichlet$(\alpha_u N_u^{-1} + n_1^u, \ldots, \alpha_u N_u^{-1} + n_{N_u}^u)$. Then set $\phi_{ku,1} = \phi_{ku,1}^* - \mu_u^*$ and $\phi_{ku,2} = \phi_{ku,2}^*$, where $\mu_u^* = \sum_{l=1}^{N_u} p_{lu}\phi_{lu,1}^*$.

Step 1:

(a) Draw $\beta_1|\text{rest} \sim \text{Normal}[V_1\{\sum_{i=1}^n Z_i(T_i^* - \beta_2 X_i - \phi_{s_{ie}e,1})/\phi_{s_{ie}e,2} + \mu_{\beta_1}\Sigma_{\beta_1}^{-1}\}, V_1]$, where $V_1 = (\sum_{i=1}^n Z_i Z_i^{\mathrm{T}}/\phi_{s_{ie}e,2} + \Sigma_{\beta_1}^{-1})^{-1}$.

(b) Draw $\beta_2|\text{rest} \sim \text{Normal}[V_2\{\sum_{i=1}^n X_i(T_i^* - \beta_1^{\mathrm{T}} Z_i - \phi_{s_{ie}e,1})/\phi_{s_{ie}e,2} + \mu_{\beta_2}/\sigma_{\beta_2}^2\}, V_2]$, where $V_2 = (\sum_{i=1}^n X_i^2/\phi_{s_{ie}e,2} + 1/\sigma_{\beta_2}^2)^{-1}$.

Step 2:

(a) For $\Delta_i = 0$ we resample T_i^* from a truncated normal distribution Normal$(\beta_1 Z_i + \beta_2 X_i + \phi_{s_{ie}e,1}, \phi_{s_{ie}e,2})I\{T^* > \log(V_i)\}$, where $I(\cdot)$ denotes an indicator function, and following Gelfand *et al.*[35] we write the resampled value of T_i^* as

$$T_i^* = \beta_1 Z_i + \beta_2 X_i + \phi_{s_{ie}e,1} + \sqrt{\phi_{s_{ie}e,2}}\,\Phi^{-1}\left\{(1-R)\Phi\left(\frac{\log(V_i) - \beta_1^{\mathrm{T}} Z_i - \beta_2 X_i - \phi_{s_{ie}e,1}}{\sqrt{\phi_{s_{ie}e,2}}}\right) + R\right\}$$

where $R \sim \text{Uniform}(0,1)$ and Φ is the standard normal cumulative distribution function.

(b) Resample X_i from the following conditional distribution $[X_i|\text{rest}] \sim \text{Normal}(m_x, V_x)$, where

$$m_x = V_x\left\{\frac{(T_i^* - Z_i^{\mathrm{T}}\beta_1 - \phi_{s_{ie}e,1})}{\phi_{s_{ie}e,2}} + \frac{(\phi_{s_{ix}x,1} + Z_i^{\mathrm{T}}\gamma_1)}{\phi_{s_{ix}x,2}} + \sum_{j=1}^{m}\frac{(W_{ij} - \phi_{s_{lu}u,1})}{\phi_{s_{lu}u,2}}\right\}$$

$$V_x = \left(\frac{\beta_2^2}{\phi_{s_{ie}e,2}} + \frac{1}{\phi_{s_{ix}x,2}} + \sum_{j=1}^{m}\frac{1}{\phi_{s_{lu}u,2}}\right)^{-1}$$

Step 3:

(a) Do this step for each $i = 1, \ldots, n$. Sample s_{ie} from a Multinomial$(p^*_{1e}, \ldots, p^*_{N_e e})$, where $(p^*_{1e}, \ldots, p^*_{N_e e}) \propto \{p_{1e} f_e(e_i|\phi_{1e}), \ldots, p_{N_e e} f_e(e_i|\phi_{N_e e})\}$ and $f_e(e_i|\phi_{je}) = \exp\{-0.5(T^*_i - \beta_1 Z_i - \beta_2 X_i - \phi_{je,1})^2/\phi_{je,2}\}/\sqrt{2\pi\phi_{je,2}}$. Finally update $n^e_1, \ldots, n^e_{N_e}$.
(b) Draw $(p_{1e}, \ldots, p_{N_e e})$ from Dirichlet$(\alpha_e N_e^{-1} + n^e_1, \ldots, \alpha_e N_e^{-1} + n^e_{N_e})$.
(c) Do this step for $j = 1, \ldots, N_e$. If $n^e_j > 0$, sample ϕ_{je} from

$$\pi(\phi_{je}|\text{rest}) \propto \prod_{s_{ie}=j} \frac{1}{\sqrt{2\pi\phi_{je,2}}} \exp\left\{ -\frac{(T^*_i - \beta_1 Z_i - \beta_2 X_i - \phi_{je,1})^2}{2\phi_{je,2}} \right\} H_{0e}(\phi_{je})$$

otherwise sample ϕ_{je} from $\pi(\phi_{je}|\text{rest}) \propto H_{0e}(\phi_{je})$.
(d) Draw τ_e from the conditional distribution

$$\text{IG}\left[0.5 \sum_{j=1}^{N_e} I(n^e_j > 0) + g_e, \left\{ h_e^{-1} + 0.5 \sum_{j=1}^{N_e} \frac{\phi^2_{je,1}}{\phi_{je,2}} I(n^e_j > 0) \right\}^{-1} \right]$$

Step 4:

(a) Do this step for each $l = 1, \ldots, nm$. Sample s_{lu} from a Multinomial$(p^*_{1u}, \ldots, p^*_{N_u u})$, where $(p^*_{1u}, \ldots, p^*_{N_u u}) \propto \{p_{1u} f_U(W_{ij} - X_i|\phi_{1u}), \ldots, p_{N_u u} f_U(W_{ij} - X_i|\phi_{N_u u})\}$ and $f_U(W_{ij} - X_i|\phi_{lu}) = \exp\{-0.5(W_{ij} - X_i - \phi_{lu,1})^2/\phi_{lu,2}\}/\sqrt{2\pi\phi_{lu,2}}$. In the end update $n^u_1, \ldots, n^u_{N_u}$.
(b) Draw $(p_{1u}, \ldots, p_{N_u u})$ from Dirichlet$(\alpha_u N_u^{-1} + n^u_1, \ldots, \alpha_u N_u^{-1} + n^u_{N_u})$.
(c) Do this step for each $k = 1, \ldots, N_u$. We determine ϕ_{ku} as follows. If $n^u_k > 0$, sample $\phi^*_{ku} = (\phi^*_{ku,1}, \phi^*_{ku,2})^{\text{T}}$ from

$$\pi(\phi^*_{ku}|\text{rest}) \propto \prod_{(ij):s_{lu}=k} \frac{1}{\sqrt{2\pi\phi^*_{ku,2}}} \exp\left[-\frac{\{W_{ij} - X_i - (\phi^*_{ku,1} - \mu^*_u)\}^2}{2\phi^*_{ku,2}} \right] H_{0u}(\phi^*_{ku})$$

Thus, we first draw $\phi^*_{ku,2}$ from IG$((a_{\sigma_u} + n^u_k/2 + 0.5), [b^{-1}_{\sigma_u} + 0.5 \sum_{(ij):s_{lu}=k} \{W_{ij} - X_i - (\phi^*_{ku,1} - \mu^*_u)\}^2 + (\phi^*_{ku,1} - m_u)^2/\tau_u]^{-1})$, and then conditional on $\phi^*_{ku,2}$ draw $\phi^*_{ku,1}$ from

$$\text{Normal}\left[\left(\frac{n^u_k(1-p_{ku})^2}{\phi^*_{ku,2}} + \frac{1}{\tau_u \phi^*_{ku,2}} \right)^{-1} \left(\frac{\sum_{(ij):s_{lu}=k} (W_{ij} - X_i + \sum_{l=1, l\neq k}^{N_u} \phi^*_{lu,1} p_{lu})(1-p_{ku})}{\phi^*_{ku,2}} + \frac{m_u}{\tau_u \phi^*_{ku,2}} \right), \left(\frac{n^u_k}{\phi^*_{ku,2}} + \frac{1}{\tau_u \phi^*_{ku,2}} \right)^{-1} \right]$$

If $n^u_k = 0$, sample $\phi^*_{ku,2}$ from IG$(a_{\sigma_u}, b_{\sigma_u})$ and sample $\phi^*_{ku,1}$ from Normal$(m_u, \tau_u \phi^*_{ku,2})$. In either case, let $\mu^*_u = \sum_{l=1}^{N_u} p_{lu} \phi^*_{lu,1}$. We then set $\phi_{ku,1} = \phi^*_{ku,1} - \mu^*_u$ and $\phi_{ku,2} = \phi^*_{ku,2}$.
(d) Draw τ_u from the conditional distribution IG$(0.5 \sum_{j=1}^{N_u} I(n^u_j > 0) + g_u, [h_u^{-1} + 0.5 \sum_{j=1}^{N_u} I(n^u_j > 0)\{(\phi^*_{ju,1})^2/\phi^*_{ju,2}\}]^{-1})$.

Step 5:

(a) Do this step for $i = 1, \ldots, n$. Sample s_{ix} from a Multinomial$(p^*_{1x}, \ldots, p^*_{N_x x})$, where $(p^*_{1x}, \ldots, p^*_{N_x x}) \propto \{p_{1e} f_x(X_i|Z_i, \gamma_1, \phi_{1x}), \ldots, p_{N_x x} f_x(X_i|Z_i, \gamma_1, \phi_{N_x x})\}$ and $f_x(X_i|Z_i, \gamma_1, \phi_{jx}) = \exp\{-0.5(X_i - Z_i^{\text{T}}\gamma_1 - \phi_{jx,1})^2/\phi_{jx,2}\}/\sqrt{2\pi\phi_{jx,2}}$.

(b) Draw $(p_{1x}, \ldots, p_{N_x x})$ from Dirichlet$(\alpha_x N_x^{-1} + n_1^x, \ldots, \alpha_x N_x^{-1} + n_{N_x}^x)$.
(c) Do this step for $j = 1, \ldots, N_x$. If $n_j^x > 0$, sample ϕ_{jx} from

$$\pi(\phi_{jx}|\text{rest}) \propto \prod_{s_{ix}=j} \frac{1}{\sqrt{2\pi\phi_{jx,2}}} \exp\left\{ -0.5 \frac{(X_i - Z_i^{\mathrm{T}}\gamma_1 - \phi_{jx,1})^2}{\phi_{jx,2}} \right\} H_{0x}(\phi_{jx})$$

otherwise sample ϕ_{jx} from $\pi(\phi_{jx}|\text{rest}) \propto H_{0e}(\phi_{jx})$.
(d) Draw τ_x from the conditional distribution IG$[0.5\sum_{j=1}^{N_x} I(n_j^x > 0) + g_x,\ \{h_x^{-1} + 0.5\sum_{j=1}^{N_x} I(n_j^x > 0)\phi_{jx,1}^2/\phi_{jx,2}\}^{-1}]$.

Step 6:

Draw γ_1 from the conditional distribution Normal$(V_{\gamma_1}\{\sum_{i=1}^n Z_i(X_i - \phi_{s_{ix}x,1})/\phi_{s_{ix}x,2} + \mu_{\gamma_1}\Sigma_{\gamma_1}^{-1}\}, V_{\gamma_1})$, where $V_{\gamma_1} = (\sum_{i=1}^n Z_i Z_i^{\mathrm{T}}/\phi_{s_{ix}x,2} + \Sigma_{\gamma_1}^{-1})^{-1}$.

Step 7

(a) The full conditional distribution of α_e is

$$\pi(\alpha_e|\text{rest}) \propto \frac{\Gamma(\alpha_e)}{\{\Gamma(\alpha_e/N_e)\}^{N_e}} (p_{1e})^{\alpha_e/N_e - 1} \times \cdots \times (p_{N_e e})^{\alpha_e/N_e - 1} \pi(\alpha_e)$$

To draw α_e from the above conditional distribution, we shall use a Metropolis–Hastings algorithm with $\pi(\alpha_e)$ as the proposal density. Suppose that at the tth iteration we draw $\alpha_e^{(new)}$ from $\pi(\alpha)$. Then

$$\alpha_e^{(t+1)} = \begin{cases} \alpha_e^{(new)} \text{ with probability } \rho(\alpha_{\mathrm{e}}^{(\mathrm{new})}, \alpha_{\mathrm{e}}^{(\mathrm{t})}) \\ \alpha_e^{(t)} \quad \text{otherwise} \end{cases}$$

where

$$\rho(\alpha_e^{(new)}, \alpha_e^{(t)}) = \frac{(p_{1e})^{\alpha_e^{(new)}/N_e - 1} \times \cdots \times (p_{N_e e})^{\alpha^{(new)_e}/N_e - 1} \Gamma(\alpha_e^{(new)})/\{\Gamma(\alpha_e^{(new)}/N_e)\}^{N_e}}{(p_{1e})^{\alpha_e^{(t)}/N_e - 1} \times \cdots \times (p_{N_e e})^{\alpha_e^{(t)}/N_e - 1} \Gamma(\alpha_e^{(t)})/\{\Gamma(\alpha_e^{(t)}/N_e)\}^{N_e}}$$

(b) Draw α_u from $\pi(\alpha_u|\text{rest}) \propto \Gamma(\alpha_u)\{\Gamma(\alpha_u/N_u)\}^{-N_u}(p_{1u})^{\alpha_u/N_u - 1} \times \cdots \times (p_{N_u u})^{\alpha_u/N_u - 1}\pi(\alpha_u)$.
(c) Draw α_x from $\pi(\alpha_x|\text{rest}) \propto \Gamma(\alpha_x)\{\Gamma(\alpha_x/N_x)\}^{-N_x}(p_{1x})^{\alpha_x/N_x - 1} \times \cdots \times (p_{N_x x})^{\alpha_x/N_x - 1}\pi(\alpha_x)$.

Step 8:

(a) Draw b_{σ_e} from IG$[\eta_e + a_{\sigma_e}\sum_{j=1}^{N_e} I(n_j^e > 0), \{\zeta_e^{-1} + \sum_{j=1}^{N_e} I(n_j^e > 0)/\phi_{je,2}\}^{-1}]$.

(b) Draw b_{σ_u} from IG$[\eta_u + a_{\sigma_u}\sum_{j=1}^{N_u} I(n_j^u > 0), \{\zeta_u^{-1} + \sum_{j=1}^{N_u} I(n_j^u > 0)/\phi_{ju,2}\}^{-1}]$.

(c) Draw b_{σ_x} from IG$[\eta_x + a_{\sigma_x}\sum_{j=1}^{N_x} I(n_j^x > 0), \{\zeta_x^{-1} + \sum_{j=1}^{N_x} I(n_j^x > 0)/\phi_{jx,2}\}^{-1}]$.

A3 Details of the MCMC steps referenced in Section 7

Here we describe the MCMC steps used for the PCPE method.

Step 1:
Draw σ_u^2 from $\mathrm{IG}\left[a_{\sigma_u} + 0.5nm, \left\{0.5\sum_{i=1}^{n}\sum_{j=1}^{m}(W_{ij} - X_i)^2 + 1/b_{\sigma_u}\right\}^{-1}\right]$.

Step 2:
Draw σ_x^2 from $\mathrm{IG}\left[a_{\sigma_x} + 0.5n, \left\{0.5\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \gamma_0 - \gamma_1^{\mathrm{T}}Z_i)^2 + 1/b_{\sigma_x}\right\}^{-1}\right]$.

Step 3:
Draw γ_0 from $\mathrm{Normal}\left[(n/\sigma_x^2 + 1/\sigma_{\gamma_0}^2)^{-1}\left\{\mu_{\gamma_0}/\sigma_{\gamma_0}^2 + \sum_{i=1}^{n}(X_i - \gamma_1^{\mathrm{T}}Z_i)/\sigma_x^2\right\}; (n/\sigma_x^2 + 1/\sigma_{\gamma_0}^2)^{-1}\right]$.

Step 4:
Draw γ_1 from

$$\mathrm{Normal}\left[\left(\sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}}/\sigma_x^2 + I_p/\sigma_{\gamma_1}^2\right)^{-1}\left\{\mu_{\gamma_1}/\sigma_{\gamma_1}^2 + \sum_{i=1}^{n}(X_i - \gamma_0)Z_i/\sigma_x^2\right\}; \left(\sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}}/\sigma_x^2 + I_p/\sigma_{\gamma_1}^2\right)^{-1}\right].$$

Step 5:
To draw $\lambda_1, \ldots \lambda_q$ we shall use the Metropolis–Hasting's algorithm. Repeat the following steps for each $s = 1, \ldots, q$:

(a) sample a proposal value $\lambda_s^{(p)}$ from $\mathrm{Gamma}(a_\lambda, b_\lambda)$.
(b) sample r_1 from $\mathrm{Uniform}(0, 1)$.
(c) if $r_1 < \rho_1$ we accept $\lambda_s = \lambda_s^{(p)}$ otherwise λ_s remains unchanged, where

$$\rho_1 = \prod_{i=1}^{n}\left\{\frac{\sum_{j=1, j\neq s}^{q}\lambda_j I(t_{j-1} \le V_i < t_j) + \lambda_s I(t_{s-1} \le V_i < t_s)}{\sum_{j=1}^{q}\lambda_j I(t_{j-1} \le V_i < t_j)}\right\}^{\Delta_i} \times \exp\{-(\lambda_s^{(p)} - \lambda_s)I(t_{s-1} \le V_i)\exp(\beta_1^{\mathrm{T}}Z_i + \beta_2 X_i)\}$$

Step 6:

We update β_1 by the Metropolis-Hastings algorithm:

(a) draw a proposal $\beta_1^{(p)}$ from $\mathrm{Normal}(\mu_{\beta_1}, \Sigma_{\beta_1})$.
(b) draw r_2 from $\mathrm{Uniform}(0, 1)$.
(c) if $r_2 < \rho_2$ accept $\beta_1 = \beta_1^{(p)}$, otherwise β_1 remains unchanged, where

$$\rho_2 = \prod_{i=1}^{n}\exp\{\Delta_i(Z_i^{\mathrm{T}}(\beta_1^{(p)} - \beta_1)\}\exp\left[-\sum_{s=1}^{q}\lambda_s I(t_{s-1} \le V_i)\exp(\beta_2 X_i)\left\{\exp(Z_i^{\mathrm{T}}\beta_1^{(p)}) - \exp(Z_i^{\mathrm{T}}\beta_1)\right\}\right]$$

Step 7:

We update β_2 by the Metropolis–Hastings algorithm:

(a) draw a proposal $\beta_2^{(p)}$ from $\mathrm{Normal}(\mu_{\beta_2}, \sigma_{\beta_2})$.

(b) draw r_3 from Uniform(0, 1).
(c) if $r_3 < \rho_3$ accept $\beta_2 = \beta_2^{(p)}$, otherwise β_2 remains unchanged, where

$$\rho_3 = \prod_{i=1}^{n} \exp\{\Delta_i(X_i(\beta_2^{(p)} - \beta_2)\} \exp\left[- \sum_{s=1}^{q} \lambda_s I(t_{s-1} \leq V_i) \exp(\beta_1^{\mathrm{T}} Z_i)\{ \exp(X_i\beta_2^{(p)}) - \exp(X_i\beta_2)\}\right]$$

Step 8:

For $i = 1, \ldots, n$, X_i is drawn from the following conditional distribution

$$\pi(X_i|\text{rest}) \propto \exp\left\{\Delta_i\beta_2 X_i - \sum_{s=1}^{q} \lambda_s I(t_{s-1} \leq V_i) \exp(\beta_1^{\mathrm{T}} Z_i + \beta_2 X_i) - \frac{(X_i - \gamma_0 - Z_i^{\mathrm{T}}\gamma_1)^2}{2\sigma_x^2} - \sum_{j=1}^{m} \frac{(W_{ij} - X_i)^2}{2\sigma_u^2}\right\}$$

A4 Further analyses of the real data using some alternative approaches

In the Buckley–James method e is treated non-parametrically. In the data analysis section we have adopted the naive and RC approaches in the Buckley–James estimates setting. We now adopt the SIMEX approach in the same setting. Furthermore, we adopt a flexible parametric model for e, and use the three-parameter generalized gamma distribution that includes gamma, Weibull, and lognormal models as special cases. Under the generalized gamma model we conducted the naive,

Table 6. Results for the ACTG AIDS clinical trial data. Here Z, Z + D, Z + Z, and D stand for zidovudine, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine, respectively, and AFT stands for accelerated failure time. The 95% Wald-type confidence intervals are given in parentheses right beneath the estimates. The bootstrap method was used to compute the standard error of the regression calibration and the SIMEX methods.

Method	Z + D (Ref: Z)	Z + Z (Ref: Z)	D (Ref: Z)	log(CD4)
Semiparametric AFT model, e is nonparametric				
SIMEX1	0.322	0.417	0.252	1.092
	(–0.044, 0.708)	(–0.004, 0.838)	(–0.122, 0.626)	(0.384, 1.799)
SIMEX2	0.306	0.396	0.248	1.133
	(–0.058, 0.671)	(–0.025, 0.817)	(–0.141, 0.636)	(0.484, 1.782)
Parametric AFT model, e is Generalized Gamma				
NV	0.339	0.443	0.338	0.972
	(0.052, 0.625)	(0.135, 0.750)	(0.064, 0.612)	(0.594, 1.350)
RC	0.339	0.440	0.341	1.181
	(0.053, 0.625)	(0.132, 0.748)	(0.063, 0.619)	(0.740, 1.622)
SIMEX1	0.349	0.435	0.334	1.094
	(0.026, 0.672)	(0.139, 0.731)	(0.024, 0.643)	(0.578, 1.609)
SIMEX2	0.333	0.435	0.332	1.121
	(0.015, 0.656)	(0.141, 0.729)	(0.026, 0.637)	(0.623, 1.618)

RC, and the two SIMEX analyses, SIMEX1 and SIMEX2. The details of these two are described in the third last paragraph in the simulation section. In SIMEX, λ values were taken between 0 and 2 with 0.2 increment. Furthermore, we have used a quadratic extrapolation function. The results are given in Table 6. The top panel (SIMEX1 and SIMEX2) of Table 6 is a continuation of the top panel of Table 5 as we have used the same setting of semiparametric AFT model with the distribution of e being left unspecified. The results indicate that the estimates under the generalized gamma model are somewhat close to that in the semiparametric AFT model. All of the methods show statistically significant association between the CD4 count and the time-to-event. Furthermore, all four approaches under the parametric AFT model indicate statistically significant association (at the 5% level) between the time-to-AIDS/death and the treatments. Note that the Wald-type confidence interval for the semiparametric AFT model is always slightly larger than that for the parametric AFT model.