The Annals of Applied Statistics
2017, Vol. 11, No. 1, 274–296
DOI: 10.1214/16-AOAS1003

PARTIALLY TIME-VARYING COEFFICIENT PROPORTIONAL HAZARDS MODELS WITH ERROR-PRONE TIME-DEPENDENT COVARIATES—AN APPLICATION TO THE AIDS CLINICAL TRIAL GROUP 175 DATA

BY XIAO SONG[1] AND LI WANG[2]

University of Georgia and Iowa State University

Due to cost and time considerations, interest has focused on identifying surrogate markers that could be substituted for the clinical endpoint, time to an event of interest, in evaluation of treatment efficacy. Joint models are often used to assess the effect of surrogate markers and treatment. Motivated by recent works studying the AIDS Clinical Trial Group (ACTG) 175 data, we propose a partially time-varying coefficient proportional hazards model for modeling the relationship between the hazard of failure and time-dependent and time-independent covariates. The time-varying coefficients are approximated by polynomial splines, and the corrected score and conditional score approaches are adopted to estimate the regression coefficients. The proposed estimators are consistent, and the asymptotic normality is established for the constant coefficients, which enables us to construct confidence intervals and permits joint inference. The finite-sample performance of the proposed method is assessed by Monte Carlo simulation studies. The proposed model is applied to ACTG 175 data to assess the temporal dynamics of the effect of treatment and CD4 count on time to AIDS or death.

1. Introduction. In biomedical studies, it is often of interest to characterize the relationship between survival time (time to an event of interest) and a set of covariates. Some of the covariates may vary over time and are measured intermittently through the studies. An example is the AIDS Clinical Trial Group (ACTG) 175 [Hammer et al. (1996)], a randomized clinical trial to compare four antiretroviral therapies, zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine or didanosine alone, in HIV infected subjects. During the study, 2467 subjects were recruited between December 1991 and October 1992 and followed until November 1994. CD4 count, a reflection of immune status, was measured about every 12 months. The survival time was time to progression to AIDS or death. The main objective of the study was to compare the treatments. A subsequent objective was to elucidate the relationship between prognosis and CD4 count and to investigate it as a potential surrogate marker for time to AIDS or death. Because of time

Received May 2016; revised November 2016.
[1]Supported in part by NSF award DMS 1106816 and NIH Grants HL121347 and CA201207.
[2]Supported in part by NSF awards DMS 0905730, 1106816 and 1542332.
Key words and phrases. Corrected score, conditional score, joint modeling, measurement error, polynomial spline, survival.

and cost issues, interest has focused on identifying surrogate markers that could be substituted for the clinical endpoint in evaluation of treatment efficacy. As the surrogate markers can be measured earlier than the clinical endpoints, it may reduce both time and cost for clinical trials. According to Prentice (1989), a surrogate marker should be prognostic for clinical outcome and the risk of progression given the marker should be independent of treatment. Thus, we need to assess the effect of treatment and CD4 count on time to AIDS or death. The (Cox) proportional hazards model may be used for this purpose. The partial likelihood approach is generally used to estimate the regression coefficients in the proportional hazards model, but it requires observations of the *true* covariate values at the event times for all at-risk subjects. However, since CD4 count was only measured intermittently, it was not available at the event times for all at-risk subjects; in addition, the measurements of CD4 count were subjected to potential measurement error. Naive approaches impute in the partial likelihood either the last observed values before the failure time (last value carried forward) or the least square estimates from the covariate profile (naive regression), which lead to biased estimation [Prentice (1982), Tsiatis and Davidian (2001)]. A popular approach is to use a joint model, which assumes that the longitudinal observations follow a mixed-effects model and the survival time depends on the random effects of the mixed-effects model through a proportional hazards model.

Various approaches have been proposed under the joint model framework. The regression calibration approach [Bycott and Taylor (1998), Dafni and Tsiatis (1998), Pawitan and Self (1993), Tsiatis, DeGruttola and Wulfsohn (1995)] reduces bias relative to naive approaches, but it may still give erroneous results [Tsiatis and Davidian (2001)]. The likelihood-based approaches [DeGruttola and Tu (1994), Faucett and Thomas (1996), Henderson, Diggle and Dobson (2000), Song, Davidian and Tsiatis (2002b), Wulfsohn and Tsiatis (1997), Xu and Zeger (2001)] are consistent, and they may be robust to misspecification of the random effect distribution when there is rich enough longitudinal information [Hsieh, Tseng and Wang (2006)]. However, they can be infeasible in the case of multiple time-dependent covariates and a large number of regression coefficients. In addition, they require the censoring time to be independent of the survival time, which may be too restricted in practice. Two attractive alternatives are the conditional score approach [Song, Davidian and Tsiatis (2002a), Tsiatis and Davidian (2001)] and the corrected score approach [Wang (2006)]. These estimating equation-based methods are easy to implement, and the estimators are consistent without distributional assumptions on the underlying true covariates. The censoring time is only assumed to be independent of the survival time given the covariates, as in the standard inference for the proportional hazards model.

However, the proportional hazards assumption may not hold for the ACTG 175 data. Song and Wang (2008) adopted the varying coefficient proportional hazards model, which allows the effect of coefficients to vary over time, and proposed a local conditional score approach; a related model is the additive coefficient model,

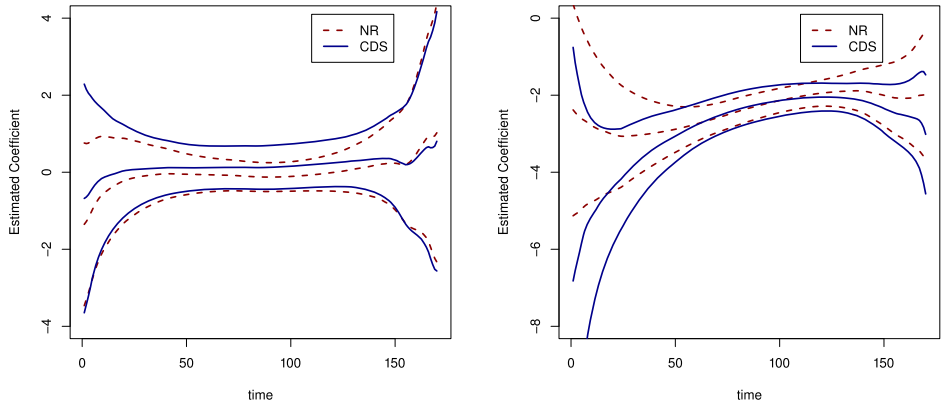


FIG. 1. *Local estimates* (*center curves*) *and* 95% *pointwise confidence intervals* (*outer curves*) *of regression coefficient for the ACTG* 175 *data under the varying coefficient model including* log(CD4) *and treatment with bandwidth* $h = 60$. *The left panel is for the coefficient of treatment, and the right panel is for the coefficient of* log(CD4). *NR, naive regression*; *CDS, conditional score.*

which is particularly helpful for studying nonlinear interaction effects of variables [Gu et al. (2014), Liu and Yang (2010), Xue and Liang (2010), Xue and Yang (2006)]. Figure 1 shows the local naive and conditional score estimates of the effect of treatment and log(CD4) [Hsieh, Tseng and Wang (2006)], where the logarithmic transformation is adopted to achieve approximate within-subject normality and constant variances. Although the effect of log(CD4) does vary over time, the effect of treatment seems stable after adjustment of log(CD4). This indicates that the coefficient of treatment may be a constant. Treating it as a time-varying coefficient may cause loss in efficiency; if log(CD4) is a surrogate marker, the treatment effect should be a constant zero after adjustment of log(CD4).

In addition, the kernel-based approach in Song and Wang (2008) is computationally intensive and has a number of undesirable features that severely impede its use in practice. First, time-varying coefficients may vary in different ways. It is desirable to allow a separate level of smoothing for each coefficient to achieve optimal efficiency. However, restricted by a single smoothing parameter (bandwidth), the kernel approach cannot provide different smoothing levels for the coefficients. Second, to obtain an estimate of the coefficient function, the estimating equations need to be solved at each time point of a dense time grid and can be computationally intensive. Third, to select the bandwidth, cross-validation is usually adopted, which imposes additional burden in computation. Fourth, when some coefficients are constant, backfitting may be adopted to estimate these coefficients. However, it will induce further complexity and intensity in computation.

This motivates an effort to develop a more flexible and efficient method to assess the effect of surrogate markers and treatment in biomedical studies. In this paper we propose using a partially time-varying coefficient proportional hazards model

which allows inclusion of both constant and time-varying covariates in the presence of measurement error. This hybrid model attempts to preserve the simplicity and efficiency when some covariates are time independent, and offers flexibility in assessing both constant and varying covariate effects on survival times. A BIC-type criterion is proposed to determine if the coefficients are constant or time varying.

We consider estimation of the coefficients using a polynomial spline approach [Nan et al. (2005)], where each time-varying coefficient is approximated by a B-spline function. To deal with the measurement error, we adopt the corrected score and conditional score approaches for their robustness and computational easiness. The polynomial spline-based estimators are global in terms of optimization, and thus it is enough to solve only one estimating equation to obtain the spline estimator. By taking advantage of spline approximation, our method is much more efficient in computation without solving the estimating equations at every time point as in kernel-based procedures, and allows different amounts of smoothing for different time-varying coefficients; see more advantages of the spline approach in Huang (1999), Huang and Liu (2006) and Wang and Yang (2007).

To our knowledge, our method represents the first attempt at investigation of the partially time-varying coefficient model for the analysis of intermittently measured longitudinal data and survival data under the joint model framework, although this model has been studied when all covariates are observed entirely at the failure times and without errors [Cai et al. (2008)]. Methodology proposed in this paper would have immediate applicability in a wide range of biomedical and epidemiological studies for dealing with survival data with longitudinal covariates measured with errors.

The rest of the article is organized as follows. We define the model in Section 2. The corrected score and conditional score estimators are proposed in Section 3 based on B-spline basis expansions. The performance of the estimators is assessed by simulations in Section 4, and illustrated by an application to the ACTG 175 data in Section 5. Section 6 investigates the misspecification issue of the hazard model. The paper concludes with a discussion in Section 7. The asymptotic properties of the proposed estimators and the corresponding proofs are provided in the Supplementary Material [Song and Wang (2017), Web Appendices A and B].

2. Model.

2.1. *Model definition.* Assume that the survival time T of an individual is subject to right censoring, and C is the underlying censoring time. Denote the observed survival or censoring time by $V = \min(T, C)$, and $\Delta = I\ (T \le C)$ is the indicator for failure, where $I(\cdot)$ is the indicator function. Suppose the survival time depends on K possible time-dependent covariates with the value $H(u) = (H_1(u), \ldots, H_K(u))^T$ at time u. The kth covariate process $H_k(u)$ $(k = 1, \ldots, K)$ is only observed at m_k time points $t_k = (t_{k1}, \ldots, t_{km_k})$ with the measures denoted

by $W_k = (W_{k1}, \ldots, W_{km_k})^T$. This general setting allows including error-free time-independent covariates with $m_k = 1$ and $W_k = X_k$. For example, in the ACTG 175, treatment was time-independent, and CD4 count was measured about every 12 weeks. Let $\boldsymbol{W} = (W_1^T, \ldots, W_K^T)^T$ denote the longitudinal observations of the K covariates, and let $\boldsymbol{t} = (t_1, \ldots, t_K)$, $\boldsymbol{m} = (m_1, \ldots, m_K)$ be the observation times. Using subscript i to denote variables for the ith subject, the observed dataset is $\{(V_i, \Delta_i, \boldsymbol{W}_i, \boldsymbol{t}_i, \boldsymbol{m}_i) : i = 1, \ldots, n\}$.

We assume the longitudinal covariate processes such as log(CD4) follow the linear mixed-effects models. Specifically, for $k = 1, \ldots, K$ and $j = 1, \ldots, m_{ik}$,

$$W_{ikj} = H_{ik}(t_{ikj}) + e_{ikj}, \qquad H_{ik}(u) = \zeta_{ik}^T f_k(u), \tag{2.1}$$

where $f_k(u)$ is a known q_k-dimensional function of u, and ζ_{ik} is a q_k-dimensional random effect. We allow f_k and ζ_{ik} to be different for each k, and no distributional assumption is placed on ζ_{ik}. This allows flexible modeling of the time trajectory of each covariate. For time-independent covariates such as treatment, $f_k(u) = 1$, and $e_{ikj} = 0$. The errors e_{ikj} are assumed to be normally distributed with mean zero and variance σ_{kk}. For simplicity, we assume the errors are independent across time; however, this assumption can be relaxed as discussed in Section 7. We allow measurements on different covariates at the same time to be correlated. Let $\sigma_{kk'}$ be the covariance between errors from covariates k and k' measured at the same time point. Let $e_i = (e_{i1}^T, \ldots, e_{iK}^T)^T$, where $e_{ik} = (e_{ik1}, \ldots, e_{ikm_{ik}})^T$, and define ζ_i and t_i similarly. We assume that e_i is independent of (T_i, C_i, ζ_i) given t_i.

Without loss of generality, suppose the first K_1 covariates $X(u)$ have constant effects on survival, and the rest of the K_2 covariates $Z(u)$ have time-varying effect on survival; that is, $H(u) = (X^T(u), Z^T(u))^T$, and $K = K_1 + K_2$. For example, in the ACTG 175 data, the covariate $X(u)$ may be treatment and $Z(u)$ may be log(CD4). We assume a partially time-varying coefficient proportional hazards model for the relationship between the hazard of failure and the covariates, under which the hazard for subject i:

$$\begin{aligned} \lambda_i(u|H) &= \lim_{du \to 0} du^{-1} \Pr\bigl(u \le T_i < u + du | T_i \ge u, \zeta_i, C_i, t_i(u), e_i(u)\bigr) \\ &= \lambda_0(u) \exp\bigl\{\beta_0^T X_i(u) + \alpha_0^T(u) Z_i(u)\bigr\}. \end{aligned} \tag{2.2}$$

Here $\lambda_0(u)$ is an unspecified baseline hazard; β_0 is a length-K_1 vector of regression parameters, and $\alpha_0(u)$ is a length-K_2 vector of smooth functions; $t_i(u) = \{t_{ikj} : t_{ikj} < u, k = 1, \ldots, K, j = 1, \ldots, m_{ik}(u)\}$ denotes the observation times by time u with $m_{ik}(u)$ being the number of observation times before u for the ith subject and kth covariate; and $e_i(u) = \{e_{ikj} : t_{ikj} < u, k = 1, \ldots, K, j = 1, \ldots, m_{ik}(u)\}$. Model (2.2) subsumes the standard proportional hazards model ($K_2 = 0$) and the varying-coefficient model ($K_1 = 0$). It implicitly assumes that censoring, timing of measurements, and covariate measurement errors are noninformative. Our interest focuses on estimation of $\eta_0(u) = (\beta_0^T, \alpha_0^T(u))^T$. Estimation of the cumulative hazard function is discussed in Section 7.

3. Approaches. For now, we assume the error variances and covariances $\sigma_{kk'} : k, k' = 1, \ldots, K$ are known. Let $\alpha_{0k}(u)$ be the kth component of $\alpha_0(u)$. To estimate $\alpha_0(u)$, a smoothing technique such as basis expansion is usually adopted. Various basis systems, including the Fourier bases, polynomial bases and B-spline bases, can be used in the basis expansion. For simplicity of implementation and fast computation, we use the B-spline basis expansion to do the approximation. Let n_k be the number of interior knots, and let p be the order of spline. Then $\alpha_{0k}(u)$ can be approximated well by a spline function so that $\alpha_{0k}(u) \approx \sum_{l=1}^{L_k} \gamma_{0kl} B_{kl}(u)$, where $\{B_{kl}(u)\}_{l=1}^{L_k}$ is a set of basis functions, and $L_k = n_k + p$ is the number of basis functions in approximating the function $\alpha_{0k}(u)$. Then model (2.2) becomes

$$\lambda_i(u) \approx \lambda_0(u) \exp\left\{\beta_0^T X_i(u) + \sum_{k=1}^{K_2} \sum_{l=1}^{L_k} \gamma_{0kl} B_{kl}(u) \mathbf{Z}_{ik}(u)\right\}. \tag{3.1}$$

The right side of (3.1) has a form of the standard proportional hazards model. Thus joint modeling approaches under the standard proportional hazards model could be applied to obtain an estimate $\hat{\theta}$ of $\theta_0 = (\beta_0^T, \gamma_0^T)^T$, where $\gamma_0 = (\gamma_{01}^T, \ldots, \gamma_{0K_2}^T)^T$, and $\gamma_{0k} = (\gamma_{0k1}, \ldots, \gamma_{0kL_k})$ for $k = 1, \ldots, K_2$. Then an estimate of $\alpha_{0k}(u)$ is $\hat{\alpha}_k(u) = \sum_{l=1}^{L_k} \hat{\gamma}_{kl} B_{kl}(u)$.

The polynomial splines were also used by Huang (1999) in the partially linear Cox models. The interior knots of the splines can be either equally spaced or placed on the sample quantiles of the failed events so that there are about the same numbers of failed events between any two adjacent knots.

Using the basis expansion, the approximated model (3.1) has a standard proportional hazards form. Thus approaches under the standard proportional hazards model could be applied. Here we focus on the corrected score [Song and Wang (2008), Wang (2006)] and conditional score [Song, Davidian and Tsiatis (2002a), Tsiatis and Davidian (2001)] approaches, as they are easy to compute and require no distributional assumptions on the random effects.

Corrected score. Let $R(u) = (X^T(u), Z^T(u)B(u))^T$ be the vector of "covariates" in (3.1), where

$$B(u) = \begin{bmatrix} B_{11}(u) & \cdots & B_{1L_1}(u) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{21}(u) & \cdots & B_{2L_2}(u) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_{K_21}(u) & \cdots & B_{K_2L_{K_2}}(u) \end{bmatrix}$$

is a $K_2 \times L$ matrix and $L = \sum_{k=1}^{K_2} L_k$. Let $\hat{H}_{ik}(u)$ be the least square estimator of $H_{ik}(u)$ using the observations W_{ikj} by time u, $\hat{X}(u) = (\hat{H}_{i1}(u), \ldots, \hat{H}_{iK_1})$, $\hat{Z}(u) = (\hat{H}_{i(K_1+1)}(u), \ldots, \hat{H}_{iK})$, and $\hat{H}(u) = (\hat{X}^T(u), \hat{Z}^T(u))^T$. Define $N_i(u) = I(V_i \le u, \Delta_i = 1, m_{ik}(u) \ge q_k, k = 1, \ldots, K)$ as the counting process for the failure events, and define $Y_i(u) = I(V_i \ge u, m_{ik}(u) \ge q_k, k = 1, \ldots, K)$ the "at-risk"

process. Let $\eta(u) = (\beta^T, \gamma^T B^T(u))^T$, and let $\theta = (\beta^T, \gamma^T)^T$. The corrected score estimating equation for θ can be written as

$$U_n^c(\theta) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[\hat{R}_i(u) + \Sigma_{R_i}(u)\theta - \frac{S_n^c(u, \eta)[\hat{R}]}{S_n^c(u, \eta)[1]} \right] dN_i(u) = 0 \tag{3.2}$$

for a fixed time τ. Here $\hat{R}(u) = (\hat{X}^T(u), \hat{Z}^T(u)B(u))^T$ is the least square estimator of $R_i(u)$ using all the observations by time u; $\Sigma_{R_i}(u)$ is the variance of $\hat{R}(u)$ conditional on $t_i(u)$; and for a scalar, vector or matrix g, $S_n^c(t, \eta)[g] = n^{-1} \sum_{i=1}^n S_{ni}^c(t, \eta)[g]$ with

$$S_{ni}^c(t, \eta)[g] = Y_i(t) g_i \exp\{\eta^T(u)\hat{H}_i(u) - \eta^T(u)\Sigma_{H_i}(u)\eta(u)/2\}.$$

The corrected score estimating function can be viewed as a "correction" of the naive regression function with the correction term $\Sigma_{R_i}(u)\theta$. The variance $\Sigma_{R_i}(u) = B_*^T(u)\Sigma_{H_i}(u)B_*(u)$, where $\Sigma_{H_i}(u)$ is the variance of $\hat{H}_i(u)$ conditional on $t_i(u)$, and

$$B_*(u) = \begin{pmatrix} I_{K_1 \times K_1} & 0_{K_1 \times L} \\ 0_{K_2 \times K_1} & B(u) \end{pmatrix}$$

with $I_{K_1 \times K_1}$ denoting a $K_1 \times K_1$ identity matrix and $0_{r \times s}$ denoting an $(r \times s)$ zero matrix.

Conditional score. The conditional score estimating equation [Song, Davidian and Tsiatis (2002a)] for θ can be written as

$$U_n^d(\theta) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[\hat{R}_i^*(u) - \frac{S_n^d(u, \eta)[\hat{R}_i^*]}{S_n^d(u, \eta)[1]} \right] dN_i(u) = 0, \tag{3.3}$$

where $\hat{R}_i^*(u) = \hat{R}_i(u) + \Sigma_{R_i}(u)\theta\, dN_i(u)$ is a " sufficient statistic" for $R_i(u)$, for a scalar, vector or matrix g, $S_n^d(u, \eta)[g] = n^{-1} \sum_{i=1}^n S_{ni}^d(u, \eta)[g]$ with

$$S_{ni}^d(u, \eta)[g] = Y_i(u) g_i \exp\{\eta^T(u)\hat{H}_i^*(u) - \eta^T(u)\Sigma_{H_i}(u)\eta(u)/2\}$$

and $\hat{H}_i^*(u) = \hat{H}_i(u) + \Sigma_{H_i}(u)\eta(u)\, dN_i(u)$. Noting that $\hat{R}_i^*(u)\, dN_i(u) = \{\hat{R}_i(u) + \Sigma_{R_i}(u)\theta\}\, dN_i(u)$, the estimating function (3.3) differs from (3.2) only in the ratio term with $S_n^d(u, \eta)[\hat{R}_i^*]/S_n^d(u, \eta)[1]$ versus $S_n^c(u, \eta)[\hat{R}]/S_n^c(u, \eta)[1]$.

As the standard conditional score and corrected score approaches, once the spline basis functions are chosen, the proposed conditional score and corrected score estimators can be obtained by solving the corresponding estimating equation via efficient Newton–Raphson methods with an appropriate starting point such as the naive regression estimator.

Under some regularity conditions, we can show that the corrected score estimator $\hat{\eta}^c(u) = (\hat{\beta}^{cT}, \hat{\gamma}^{cT} B^T(u))^T$ solving (3.2) and the conditional score estimator

$\hat{\eta}^d(u) = (\hat{\beta}^{dT}, \hat{\gamma}^{dT} B^T(u))^T$ solving (3.3) are consistent. In addition, $\hat{\beta}^c$ and $\hat{\beta}^d$ are asymptotically normal. The details and the proofs are outlined in the Supplementary Material [Song and Wang (2017), Web Appendices A and B]. The asymptotic distribution result enables us to construct confidence intervals for the coefficients simultaneously.

When $\Sigma_{H_i}(u)$ is unknown, it can be estimated by the method of moments estimator $\hat{\Sigma}_{H_i}(u)$ as in Song, Davidian and Tsiatis (2002a). The corrected score and conditional score estimates can be obtained by substituting $\hat{\Sigma}_{H_i}(u)$ for $\Sigma_{H_i}(u)$ in (3.2) and (3.3). It can be easily shown that $\hat{\Sigma}_{H_i}(u)$ is a root-n consistent estimator of $\Sigma_{H_i}(u)$. Replacing $\Sigma_{H_i}(u)$ with $\hat{\Sigma}_{H_i}(u)$ does not affect the convergence rate of the corrected score and conditional score estimators and the asymptotic normality. The asymptotic variance can be estimated by stacking the estimating function for $\Sigma_{H_i}(u)$ and the estimating function for (3.2) or (3.3).

4. Simulation studies. Simulation studies were conducted to evaluate the performance of the approaches. We considered the case of three covariates $X_i(u) = \zeta_{i10}$, $Z_{i1}(u) = \zeta_{i20} + \zeta_{i21}u$, and $Z_{i2}(u) = \zeta_{i30} + \zeta_{i31}\sqrt{u}$, where $(\zeta_{i10}, \zeta_{i20}, \zeta_{i21}, \zeta_{i30}, \zeta_{i31})$ had a normal distribution or a mixture of two normal distributions [mixing proportion $p = 0.3$, and the distance between the means $(1, \mu, -0.01, 4, -0.01)$ and $(1, -\mu, -0.01, 4, -0.01)$ was 2 times the common standard deviation], with mean $(1, 5, -0.01, 4, -0.01)$, variance $(1, 1, 0.004, 1, 0.003)$ and a common correlation corr $= 0.25$ between any two components. The covariates $Z_{i1}(u)$ and $Z_{i2}(u)$ were measured longitudinally at times $0, 2, 4, 8, 12, 24, 36, \ldots, 180$ with normal measurement errors. The variances of the errors $\sigma_{11} = \sigma_{22}$ were 0.08 or 0.16. The true regression coefficients were $\beta_0 = -1.5$, $\alpha_{01}(u) = -0.5\sin(u\pi/90) - 1.8$, and $\alpha_{02}(u) = 0.3\log(u/5+1) - 1.5$. The baseline hazard was taken to be $\lambda_0(u) = I(u > 8)u^{0.8}\exp(-u^{0.01} + 3)$. Censoring time was generated from an exponential distribution with mean 800 and truncated at $u = 180$, leading to a censoring rate of 48%.

We used quadratic splines and equally spaced knots in the simulations. The number of knots was selected by a BIC-type criterion, specifically, by minimizing $-2L_n(\theta) + n_p\log(n)$, where n_p is the number of parameters, and

$$L_n(\theta) = n^{-1}\int_0^\tau \{\theta^T(u)R_i(u) - \log S_n(u,\eta)[1]\}\,dN_i(u)$$

is the log partial likelihood function, where $S_n(t,\eta)[1] = n^{-1}\sum_{i=1}^n S_{ni}(t,\eta)[1]$ with $S_{ni}(t,\eta)[1] = Y_i(t)\exp\{\eta^T(u)H_i(u)\}$. For corrected score and conditional score approaches, we used the corrected log partial likelihood function

$$L_n^c(\theta) = n^{-1}\int_0^\tau \{\theta^T(u)\hat{R}_i(u) - \log S_n^c(u,\eta)[1]\}\,dN_i(u). \tag{4.1}$$

For each scenario, we generated 500 simulated datasets with sample size $n =$ 500 or 1000. For each dataset, the regression coefficients $\alpha_{01}(u)$ and $\alpha_{02}(u)$ were

approximated by the B-spline and estimated in five ways: (i) using the "ideal" approach (Ideal) where the true values of $X_i(u)$ were used; (ii) using the naive regression (NR) approach; (iii) using the risk set regression calibration (RC) approach [Tsiatis and Davidian (2001)]; (iv) using the conditional score (CDS) approach; and (v) using the corrected score (CRS) approach. Ideally, implementation of the RC approach requires one to fit a linear mixed-effects model at each event time. This is prohibitive in simulations as the number of linear mixed-effects model fitting will be overwhelming. Instead, we used a strategy that has been used in Tsiatis and Davidian (2001) and only fitted the linear mixed-effects model using available data up to the 10th, 20th, ..., and 100th percentiles. Our preliminary studies indicate that the difference between the full risk set regression calibration and the simplified modification is negligible.

To evaluate the performance of the approaches, for each time-varying coefficient, we calculated the average of the mean absolute bias (bias) at equally spaced grids between the 5th and 95th percentiles of the observed survival times with increment 1, where the absolute bias at time u equals $|\hat{\alpha}_k(u) - \alpha_0(u)|$ $(k = 1, 2)$, across the simulated datasets. Similarly, we calculated the average of the mean standard deviation (SD), mean standard error (SE) and mean coverage probability (CP) of 95% Wald confidence intervals based on the approximated model (3.1). For the constant coefficient, we gave the same statistics except replacing the mean absolute bias by the bias. We also report the nonconvergence (NC) rate for each estimation approach.

The results for the normal covariates are shown in Table 1, and the results for the mixture of normal covariates are given in Table 2. The CDS estimates have bias and coverage probabilities close to the ideal estimates, and the performance improves with the sample size increasing. In contrast, the NR and RC estimates are biased, and the coverage probabilities are way below the nominal level, which worsen with increased sample size. This is more dramatic for the NR approach. The CRS approach does not converge or has outlier estimates for some simulated datasets, especially when the measurement error is relatively large and the sample size is relatively small. The nonconvergence rates are more than 70% for $n = 500$ and $\sigma_{kk} = 0.16$. The estimates have larger bias and standard deviations than the CDS estimates, although the discrepancy decreases with the sample size increasing. In Figures 2 and 3, the true time-varying coefficients $\alpha_1(u)$ and $\alpha_2(u)$ are overlaid by the estimates and the 95% pointwise Wald confidence intervals, which conform to the results in Tables 1 and 2.

To assess the impact of the correlations between the covariates, we also ran simulations under different correlations between the random effects. Table 3 shows the results for corr $= 0, 0.25, 0.5, 0.75$ in the case of normal random effects with sample size $n = 1000$ and $\sigma_{kk} = 0.16$, which indicate that the empirical standard deviations increase with the increase of correlations. The results for the CRS are not presented due to the high rate of nonconvergence.

TABLE 1
Simulation results in the case of normal random effects

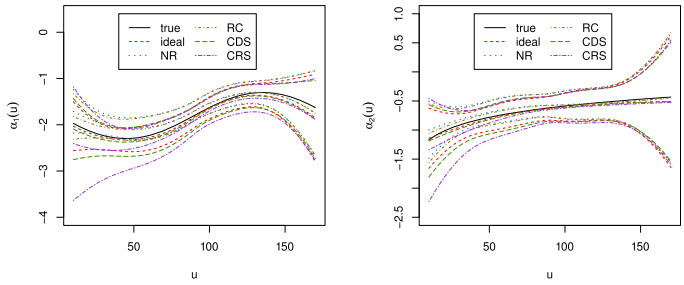
		β				α_1				α_2				
		bias	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**NC (%)**
							$n = 500$							
	ideal	−0.021	0.097	0.102	0.960	0.036	0.136	0.132	0.940	0.021	0.130	0.123	0.942	0
$\sigma_{kk} = 0.08$	NR	0.109	0.096	0.097	0.784	0.141	0.131	0.120	0.706	0.058	0.128	0.118	0.878	0
	RC	0.096	0.097	0.097	0.820	0.117	0.132	0.123	0.771	0.045	0.131	0.121	0.904	0
	CDS	−0.045	0.127	0.119	0.926	0.069	0.179	0.152	0.900	0.036	0.159	0.135	0.908	0
	CRS	−0.128	0.160	0.152	0.901	0.176	0.229	0.209	0.905	0.077	0.180	0.159	0.918	4.8
$\sigma_{kk} = 0.16$	NR	0.199	0.097	0.094	0.428	0.265	0.126	0.113	0.407	0.110	0.127	0.115	0.739	0
	RC	0.178	0.097	0.094	0.518	0.224	0.130	0.117	0.487	0.083	0.131	0.120	0.824	0
	CDS	−0.064	0.155	0.136	0.906	0.097	0.219	0.176	0.882	0.048	0.188	0.154	0.899	0
	CRS	−0.172	0.209	0.375	0.974	0.214	0.287	0.509	0.960	0.086	0.209	0.270	0.944	77.0
							$n = 1000$							
	ideal	−0.011	0.072	0.070	0.946	0.018	0.093	0.089	0.939	0.012	0.085	0.083	0.945	0
$\sigma_{kk} = 0.08$	NR	0.130	0.068	0.066	0.488	0.172	0.088	0.081	0.469	0.071	0.085	0.080	0.779	0
	RC	0.115	0.069	0.067	0.580	0.148	0.089	0.082	0.552	0.056	0.087	0.081	0.829	0
	CDS	−0.017	0.091	0.084	0.926	0.028	0.124	0.107	0.910	0.015	0.107	0.094	0.921	0
	CRS	−0.056	0.102	0.092	0.904	0.078	0.145	0.122	0.892	0.035	0.115	0.100	0.912	0
$\sigma_{kk} = 0.16$	NR	0.222	0.067	0.064	0.076	0.299	0.085	0.075	0.196	0.126	0.084	0.078	0.558	0
	RC	0.199	0.067	0.064	0.136	0.256	0.086	0.078	0.240	0.098	0.087	0.081	0.671	0
	CDS	−0.028	0.113	0.098	0.910	0.044	0.157	0.127	0.891	0.021	0.128	0.109	0.910	0
	CRS	−0.115	0.145	0.148	0.911	0.142	0.195	0.207	0.926	0.055	0.147	0.142	0.934	36.8

NR, naive regression; RC, regression calibration; CDS, conditional score; CRS, corrected score; bias, average of estimated bias or absolute bias for β or $\alpha(u)$; SD, empirical standard deviation across simulated datasets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval; NC, nonconvergence rate.

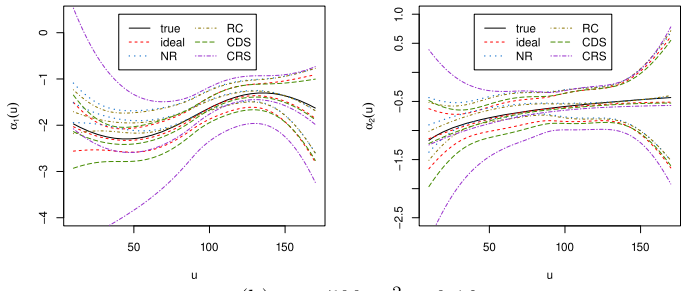
TABLE 2
Simulation results in the case of mixture of normal random effects

		β				α_1				α_2				
		bias	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**NC (%)**
							$n = 500$							
	ideal	−0.021	0.107	0.103	0.946	0.035	0.164	0.154	0.936	0.021	0.146	0.141	0.941	0
$\sigma_{kk} = 0.08$	NR	0.115	0.100	0.097	0.740	0.181	0.152	0.139	0.672	0.093	0.153	0.135	0.852	0
	RC	0.104	0.100	0.097	0.786	0.144	0.156	0.143	0.745	0.075	0.153	0.138	0.879	0
	CDS	−0.034	0.130	0.118	0.932	0.064	0.221	0.183	0.908	0.027	0.198	0.161	0.899	0
	CRS	−0.111	0.162	0.152	0.909	0.155	0.226	0.214	0.916	0.061	0.179	0.160	0.917	5.4
$\sigma_{kk} = 0.16$	NR	0.204	0.099	0.094	0.424	0.323	0.147	0.128	0.348	0.156	0.151	0.131	0.700	0
	RC	0.186	0.100	0.094	0.488	0.260	0.154	0.136	0.461	0.120	0.154	0.136	0.790	0
	CDS	−0.050	0.158	0.134	0.912	0.095	0.281	0.216	0.891	0.038	0.242	0.186	0.887	0
	CRS	−0.175	0.203	0.266	0.971	0.204	0.312	0.510	0.971	0.067	0.251	0.308	0.934	72.6
							$n = 1000$							
	ideal	−0.017	0.071	0.071	0.948	0.024	0.092	0.090	0.939	0.009	0.086	0.083	0.941	0
$\sigma_{kk} = 0.08$	NR	0.120	0.071	0.067	0.540	0.167	0.088	0.081	0.476	0.073	0.086	0.080	0.788	0
	RC	0.107	0.071	0.067	0.622	0.142	0.089	0.083	0.560	0.059	0.087	0.082	0.842	0
	CDS	0.033	0.096	0.085	0.912	0.039	0.124	0.108	0.910	0.014	0.107	0.095	0.920	0
	CRS	−0.075	0.108	0.093	0.874	0.092	0.144	0.122	0.876	0.034	0.116	0.101	0.914	0.2
$\sigma_{kk} = 0.16$	NR	0.211	0.071	0.064	0.136	0.294	0.085	0.076	0.204	0.126	0.086	0.078	0.560	0
	RC	0.191	0.070	0.065	0.200	0.252	0.087	0.079	0.253	0.100	0.088	0.081	0.675	0
	CDS	0.048	0.119	0.099	0.886	0.057	0.156	0.127	0.892	0.022	0.128	0.110	0.913	0
	CRS	0.154	0.156	0.167	0.915	0.171	0.203	0.226	0.917	0.060	0.155	0.148	0.932	41.0

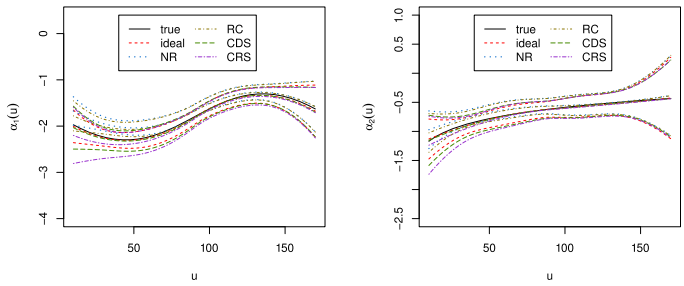
NR, naive regression; RC, regression calibration; CDS, conditional score; CRS, corrected score; bias, average of estimated bias or absolute bias for β or $\alpha(u)$; SD, empirical standard deviation across simulated datasets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval; NC, nonconvergence rate.



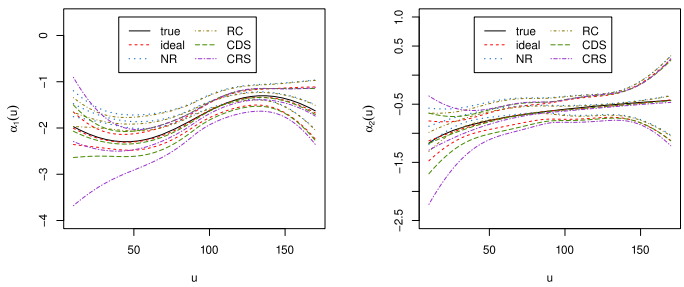
(a) $n = 500$, $\sigma^2 = 0.08$



(b) $n = 500$, $\sigma^2 = 0.16$

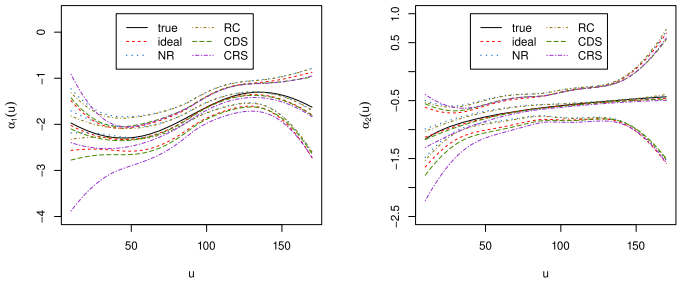


(c) $n = 1000$, $\sigma^2 = 0.08$

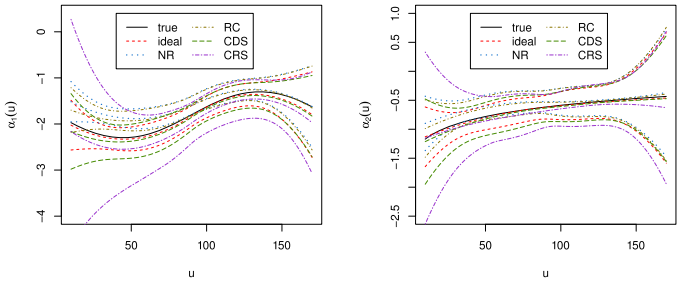


(d) $n = 1000$, $\sigma^2 = 0.16$

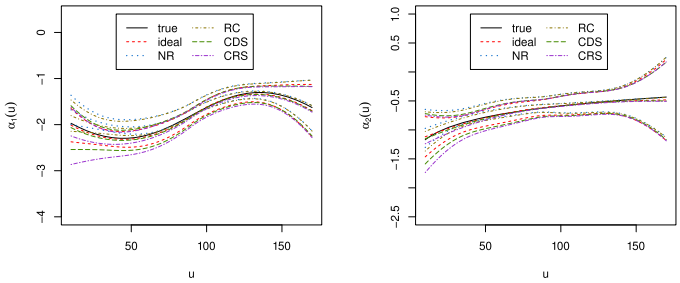
FIG. 2. *Estimates* (*center curves*) *of* $\alpha_1(u)$ *and* $\alpha_2(u)$ *with* 95% *pointwise confidence intervals* (*outer curves*) *in the case of normal random effects*. *NR*, *naive regression*; *RC*, *regression calibration*; *CDS*, *conditional score*; *CRS*, *corrected score*.



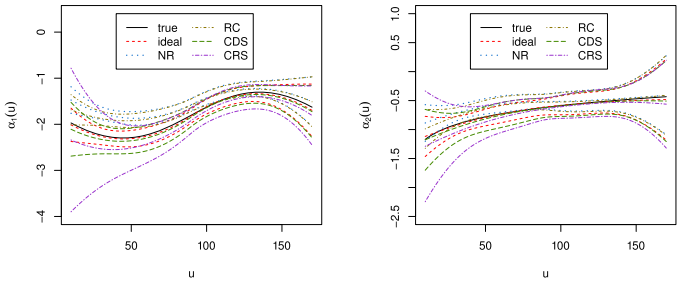
(a) $n = 500$, $\sigma^2 = 0.08$



(b) $n = 500$, $\sigma^2 = 0.16$



(c) $n = 1000$, $\sigma^2 = 0.08$



(d) $n = 1000$, $\sigma^2 = 0.16$

FIG. 3. *Estimates* (*center curves*) *of* $\alpha_1(u)$ *and* $\alpha_2(u)$ *with* 95% *pointwise confidence intervals* (*outer curves*) *in the case of mixture of normal random effects*. *NR*, *naive regression*; *RC*, *regression calibration*; *CDS*, *conditional score*; *CRS*, *corrected score*.

TABLE 3

Simulation results under various correlation values (corr) *between the random effects in the case of normal covariates with* $n = 1000$ *and* $\sigma_{kk} = 0.16$

		β				α_1				α_2			
		bias	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**
corr = 0.00	ideal	−0.013	0.069	0.067	0.952	0.021	0.085	0.084	0.945	0.006	0.080	0.077	0.948
	NR	0.225	0.067	0.061	0.076	0.274	0.078	0.072	0.194	0.127	0.080	0.074	0.534
	RC	0.211	0.067	0.061	0.086	0.246	0.079	0.074	0.228	0.096	0.082	0.076	0.695
	CDS	−0.028	0.107	0.094	0.908	0.041	0.133	0.117	0.916	0.015	0.115	0.100	0.910
corr = 0.25	ideal	−0.011	0.072	0.070	0.946	0.018	0.093	0.089	0.939	0.012	0.085	0.083	0.945
	NR	0.222	0.067	0.064	0.076	0.299	0.085	0.075	0.196	0.126	0.084	0.078	0.558
	RC	0.199	0.067	0.064	0.136	0.256	0.086	0.078	0.240	0.098	0.087	0.081	0.671
	CDS	−0.028	0.113	0.098	0.910	0.044	0.157	0.127	0.891	0.021	0.128	0.109	0.910
corr = 0.50	ideal	−0.010	0.081	0.077	0.944	0.017	0.099	0.096	0.943	0.011	0.102	0.096	0.937
	NR	0.203	0.080	0.072	0.234	0.315	0.092	0.081	0.237	0.129	0.098	0.089	0.584
	RC	0.174	0.080	0.072	0.348	0.259	0.093	0.085	0.285	0.103	0.102	0.092	0.691
	CDS	−0.028	0.129	0.108	0.888	0.047	0.176	0.141	0.895	0.025	0.151	0.128	0.911
corr = 0.75	ideal	0.006	0.099	0.095	0.930	0.037	0.150	0.135	0.929	0.022	0.130	0.125	0.940
	NR	0.137	0.098	0.091	0.670	0.248	0.140	0.119	0.451	0.135	0.127	0.114	0.662
	RC	0.120	0.098	0.091	0.716	0.194	0.142	0.123	0.497	0.109	0.130	0.119	0.761
	CDS	−0.015	0.149	0.130	0.916	0.075	0.234	0.182	0.888	0.035	0.209	0.177	0.915

NR, naive regression; RC, regression calibration; CDS, conditional score; bias, average of estimated bias or absolute bias for β or $\alpha(u)$; SD, empirical standard deviation across simulated datasets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

In the simulation studies, for each method and each time-varying coefficient, three knots were selected for more than 90% of the simulated datasets. To evaluate the effect on the estimation of the variance of the estimators due to ignoring of the variability induced by the choice of the number of knots, we compared the results with those when the number of knots are fixed at three. There is only slight inflation in the empirical standard deviations when the number of knots are selected by BIC.

In the ACTG 175 study, the error seems short tailed compared to the normal and close to a scaled t-distribution with degrees of freedom 4 (see Section 5). Thus we conducted simulations under the same scenario as for Table 1 except that the error for $Z_{i1}(u)$ was generated from a scaled t-distribution with degree 4. We show the results for $n = 500$ in Table 4. The results are mostly similar to those in Table 1. The CDS approach has slightly lower coverage probability when the error variances equal 0.16.

We also conducted simulations to compare the proposed spline-based approaches and the kernel-based approaches in Song and Wang (2008). Since the latter cannot be directly applied to the partially time-varying coefficient model with constant coefficients, we only considered the simpler time-varying coefficient model without constant coefficients. The simulation setting was taken to be the same as in Song and Wang (2008) with a single time-dependent covariate $Z_i(u) = \zeta_{i0} + \zeta_{i1}u$, error variance 0.4, the true regression coefficient $\alpha_0(u) = 0.3\log(u/5 + 1) - 1.5$, and the observed survival time between $[0, 80]$. Different from the spline-based approaches, the kernel-based approaches require separate runs to get the estimates of $\alpha_0(u)$ at different time points. In addition, the threefold cross-validation was used to select the bandwidth which requires estimation of $\alpha_0(u)$ at each failure time for each "fold." We calculated the "ideal," NR and CDS estimates with either spline or kernel approximation at a grid of points $u = 20, 30, \ldots, 60$. The results from 100 simulated datasets with $n = 600$ are shown in Table 5, which indicates that the spline approaches are comparable to the kernel-based approaches in estimation of $\alpha_0(u)$. However, the spline-based approaches are much more computationally efficient. For each dataset, it took on average about 22 seconds to obtain the spline-based estimates on a PC with Intel Xeon CPU X5355 @ 2.66GHz, while about 41 minutes to obtain the kernel-based estimates. All approaches were implemented in C++.

5. Application. We applied the approaches to the ACTG 175 data. We were interested in evaluating CD4 count as a potential surrogate marker, and thus needed to assess the effect of treatment on time to AIDS or death adjusted for CD4 count. According to Prentice (1989), a surrogate marker should satisfy two conditions: (i) the marker should be prognostic for clinical outcome; (ii) the risk of progression given the marker should be independent of treatment.

In the ACTG 175 study, there were 308 events with an average of 8.2 CD4 measurements per subject. Figure 4 presents $\log_{10}$ transformed CD4 profiles for 10 randomly selected subjects and shows an initial increase, with a peak at week 12,

TABLE 4
Simulation results in the case of nonnormal error

		$\boldsymbol{\beta}$				$\boldsymbol{\alpha_1}$				$\boldsymbol{\alpha_2}$				
		bias	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**bias**	**SD**	**SE**	**CP**	**NC (%)**
								$n = 500$						
	ideal	−0.023	0.100	0.103	0.962	0.031	0.129	0.131	0.954	0.012	0.124	0.121	0.944	0
$\sigma_{kk} = 0.08$	NR	0.112	0.105	0.097	0.744	0.155	0.133	0.119	0.685	0.068	0.124	0.117	0.865	0
	RC	0.097	0.106	0.097	0.778	0.130	0.134	0.121	0.753	0.053	0.126	0.119	0.894	0
	CDS	−0.031	0.139	0.119	0.918	0.037	0.191	0.152	0.903	0.016	0.154	0.134	0.910	0.2
	CRS	−0.113	0.170	0.159	0.914	0.140	0.249	0.224	0.918	0.057	0.178	0.163	0.921	4.2
$\sigma_{kk} = 0.16$	NR	0.203	0.110	0.094	0.408	0.282	0.136	0.111	0.374	0.123	0.124	0.114	0.729	0
	RC	0.181	0.111	0.094	0.512	0.240	0.139	0.115	0.465	0.094	0.128	0.118	0.805	0
	CDS	−0.036	0.180	0.136	0.876	0.046	0.275	0.182	0.858	0.026	0.204	0.155	0.889	0.2
	CRS	−0.127	0.230	0.254	0.942	0.119	0.326	0.342	0.912	0.055	0.218	0.220	0.923	58.8

NR, naive regression; RC, regression calibration; CDS, conditional score; CRS, corrected score; bias, average of estimated bias or absolute bias for β or $\alpha(u)$; SD, empirical standard deviation across simulated datasets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval; NC, nonconvergence rate.

TABLE 5
Simulation results comparing the spline- and kernel-based approaches

		α			
		bias	**SD**	**SE**	**CP**
Kernel	ideal	0.016	0.060	0.059	0.958
	NR	0.074	0.059	0.056	0.658
	CDS	0.030	0.099	0.085	0.934
Spline	ideal	0.007	0.069	0.072	0.976
	NR	0.061	0.069	0.066	0.796
	CDS	0.012	0.104	0.091	0.934

NR, naive regression; CDS, conditional score; bias, absolute bias; SD, empirical standard deviation across simulated datasets; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

followed by an approximate linear decline. Because only nine events occurred before week 12, for simplicity, we considered the data including and after week 12. These included 2266 subjects with at least one CD4 observation, among which there were 286 events with 273 events occurring after two CD4 measurements.

To achieve approximate within-subject normality and constant variance, we applied base-10 logarithmic transformation to CD4 measurements. The trajectory of log(CD4) seemed approximately linear after week 12. Thus we assumed that $Z_i(u) = \alpha_{i10} + \alpha_{i11}u$ represented the inherent log(CD4) count for subject i at time u. Figure 5 shows the residual plots from the least square estimates and the corresponding normal Q–Q plot and t Q–Q plot with degrees of freedom 4. It seems reasonable to assume constant error variance, and the error distribution may be short tailed relative to the normal and close to the scaled t-distribution with

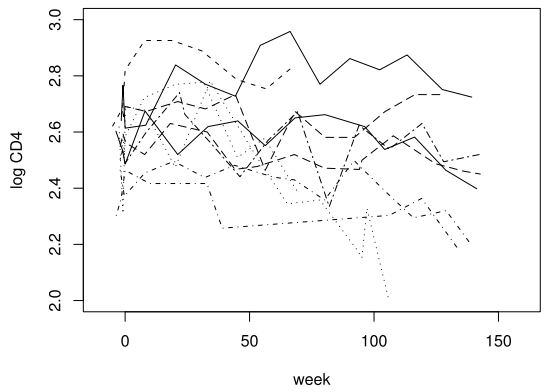


FIG. 4. *Trajectories of* log(CD4) *for* 10 *randomly selected subjects*.

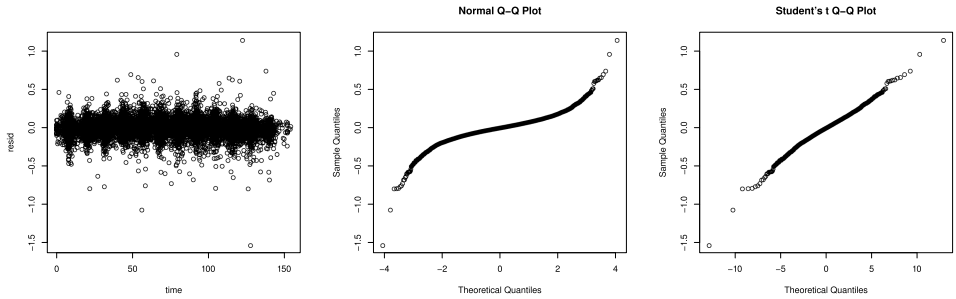


FIG. 5. *Left*: *residual plot*; *right*, *Q–Q plot of the residuals*.

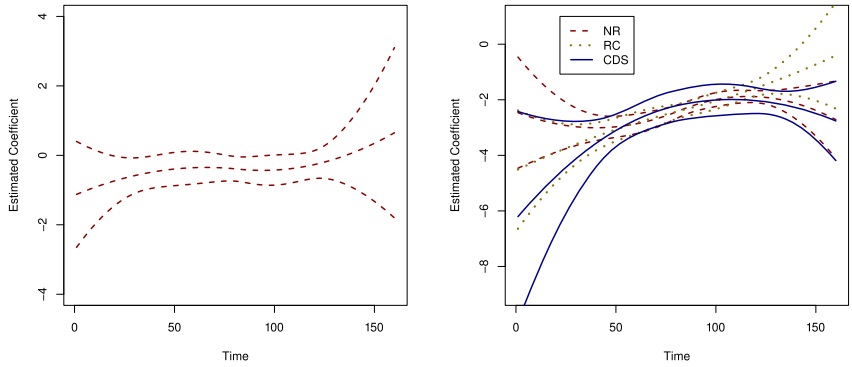
degrees of freedom 4. The estimated error variance was 0.011, which was about 42% of the estimated baseline CD4 variance.

As the original analysis found that zidovudine alone was inferior to the other three therapies [Hammer et al. (1996)], we focused on two treatment groups, zidovudine alone and the combination of the other three. Let $X_i = I(\text{treatment} \neq \text{zidovudine alone})$. We proceeded to the main analysis below.

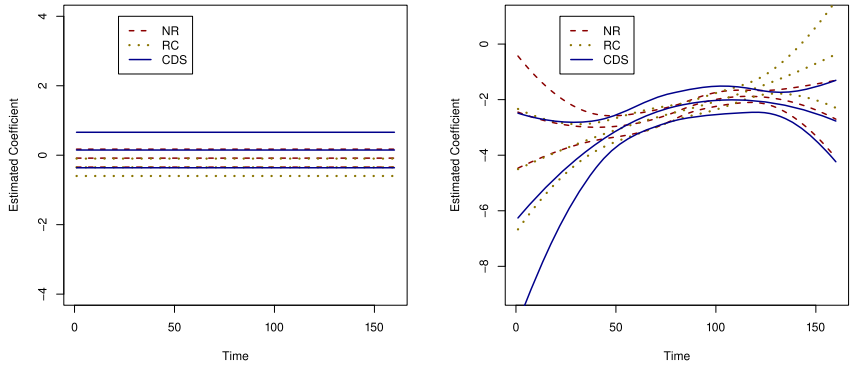
To assess the association of log(CD4) as a surrogate marker, we fitted three models: (1) the hazard model with the treatment X_i only; (2) the hazard model with the log(CD4) $Z_i(u)$ only; (3) the hazard model with both the treatment and log(CD4). For each model, we used the BIC criterion described in Section 4 to determine if the coefficients are constant or time varying and to select the number of knots for spline smoothing.

Model (1) included only a time-independent covariate (treatment), and it was fitted via the standard partial likelihood approach after the spline approximation. Models (2) and (3) included a time-dependent covariate log(CD4), and we fitted them using the NR, RC and CDS approaches. BIC is used for model selection with the number of deaths used as effective sample size. For models (1) and (2), BIC is smallest when the coefficient is time varying with one interior knot. For model (3), BIC is smallest when the coefficient for treatment is constant and the coefficient for log(CD4) is time varying with one interior knot; this conforms to what we have observed in Figure 1. The CRS estimates were unstable and are not shown.

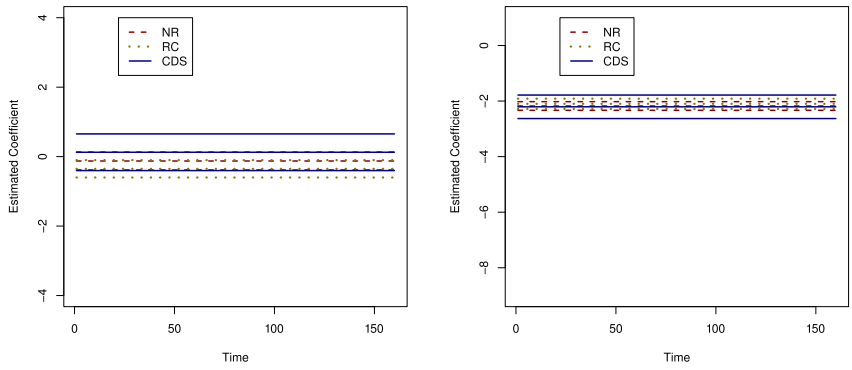
The estimated regression coefficients and the 95% pointwise confidence intervals are shown in Figure 6. Treatment alone [model (1)] shows a significant effect, and the magnitude of effect seems attenuated toward zero over time and may disappear eventually around week 130. For model (2), the results from the CDS, NR and RC approaches all show that log(CD4) alone has a significant effect over time, indicating larger log(CD4) is associated with longer time to AIDS or death, which confirms Prentice's condition (i) that log(CD4) is prognostic for survival time. However, the results differ considerably; the CDS estimate suggests the effect of log(CD4) decreases obviously before 80 weeks and levels off afterward; the NR shows a similar trend but the decrease is more modest before 80 weeks; in contrast,



(a) treatment only (b) log(CD4) only



(c) treatment and log(CD4): partially time-varying



(d) treatment and log(CD4): constant

FIG. 6. (a) *Estimation of regression coefficients* (*center curves*) *and* 95% *pointwise confidence intervals* (*outer curves*) *for the ACTG* 175 *data under the partially varying coefficient proportional hazards models including* (a) *treatment only*; (b) log(CD4) *only*; (c) *treatment and* log(CD4); *and* (d) *under the standard proportional hazard model including treatment and* log(CD4). *The left panel is for the coefficient of treatment, and the right panel is for the coefficient of* log(CD4). *NR, naive regression*; *RC, regression calibration*; *CDS, conditional score.*

the RC suggests an almost linear trend in decreasing with the rate intermediate between those of the CDS and NR estimates before week 80. For model (3), both the NR and CDS results show that the treatment effect is no longer significant after adjustment for log(CD4), which supports Prentice's condition (ii) that the effect of treatment is mediated through log(CD4). The confidence bands are narrower than those in Figure 1, which implies possible improvement of efficiency, although the kernel estimates lead to the same inference. These results are in marked contrast to the conclusion that would be reached from the RC approach, which shows a significant treatment effect even after adjustment.

As for comparison, we also show the estimated coefficients under the standard proportional hazards model; see Figure 6(d). For each approach, the effect of log(CD4) looks like some kind of average of the effect under model (3), and the effect of treatment is similar. Thus we reach the same conclusion in assessing the surrogacy of log(CD4) even though the hazard model may be misspecified. A possible explanation is given in Section 6.

The proposed spline approach and the kernel approach in Song and Wang (2008) were both implemented in C++. The computing time to get the NR and CDS estimates in Figure 6(c) was 27 seconds on a PC with Intel Xeon CPU X5355 @ 2.66GHz. In contrast, it took 24 minutes to get the kernel estimates in Figure 1 at time points $1, 2, \ldots, 170$ weeks, which would be longer if the time points were denser. This indicates that the spline approach is much more efficient in computation.

6. Misspecification of standard proportional hazards model. Consider the ideal case that all covariates are observed without error through the time. If the time-varying coefficients in model (2.2) are misspecified as constants, which reduce to the standard proportional hazards model, it can be shown that there is a unique solution $\hat{\eta}_{\mathrm{mis}}$ to the partial likelihood estimating equation based on the misspecified model, and $\hat{\eta}_{\mathrm{mis}}$ converges to a constant vector η_{mis}; see Web Appendix C for the convergence result. However, the limit may be far off from the truth. Consider a hypothetical example of a clinical trial comparing an intervention with a control. Suppose that $X_i \sim \text{binomial}(0.5)$ is an indicator of treatment group, which equals 0 if i is in the control group and 1 otherwise, and a surrogate marker $Z_i(u) = \zeta_{i0} + \zeta_{i1}u$, where $\zeta_i = (\zeta_{i0}, \zeta_{i1})$ is normally distributed. The hazard of event is $\lambda_i(u) = \lambda_0(u)\exp\{\beta_0 X_i(u) + \alpha_0(u)Z_i(u)\}$, where $\lambda_0(u) = 0.4I(u \geq 5)$ represents the case that the hazard can be ignored at the beginning of the study, $\beta_0 = 0$ indicates that the treatment effect is mediated through the surrogate marker, and $\alpha_0(u) = 0.5\log\{(uI(u < 90) + 90I(u \geq 90))/5 + 1\} - 1.5$ shows a marker effect that diminishes before $u = 90$ and becomes stable afterward (Figure 7).

We consider two cases: (1) The mean of the random effect ζ_i equals $(2.5915, -0.00315I(X_i = 0) - 0.00198I(X_i = 1))$, and the variance of ζ_i equals $(0.02408, 0.000014)$ with covariance -0.00008 [obtained by fitting a mixed-effects model

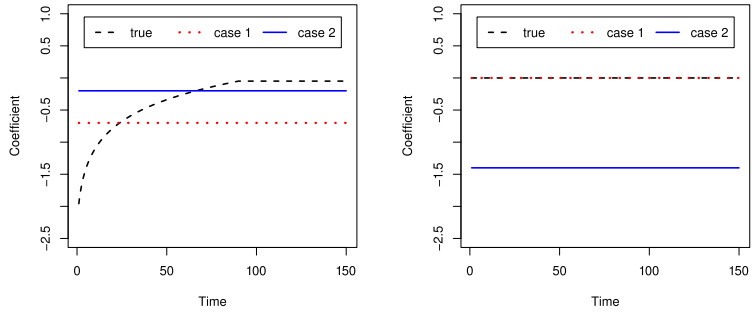


FIG. 7. *Misspecification of partially time-varying proportional hazards model. Left*: *time-varying coefficient*; *right*, *constant coefficient*.

to log(CD4) in the ACTG175 data]; that is, the means are the same for the control and intervention due to the randomization and the slopes are different. Under the misspecified model assuming constant coefficients for both covariates, $\hat{\eta}_{\text{mis}}$ converges to $\eta_{\text{mis}} \approx (-0.7, 0.0)$, which may still reveal the marker surrogacy although the estimation of the marker effect is quite off. (2) The variance of ζ_i is the same and the mean equals $(2.5915, -0.00315I(X_i = 0) + 0.15I(X_i = 1))$, which only differs from case (1) in the slope for the intervention. The estimator $\hat{\eta}_{\text{mis}}$ converges to $\eta_{\text{mis}} \approx (-0.2, -1.4)$, which may mask the surrogacy of $Z_i(u)$.

7. Discussion. Due to time and cost considerations, it is of interest to identify surrogate markers that could be substituted for the clinical endpoint, time to an event of interest, in evaluation of treatment efficacy. In practice, the effect of the marker may be time varying, and the effect of treatment after adjustment of the marker is a constant. Using a standard proportional hazards model with constant covariate effects may lead to erroneous inference. On the other hand, adopting a time-varying coefficient model with all time-varying covariate effects may cause loss of efficiency. A partially varying coefficient model strikes a delicate balance between the simplicity of the standard proportional hazard model and the flexibility of the time-varying coefficient model.

In this article we have adopted polynomial spline approximation for the partially varying-coefficient proportional hazards model with intermittently measured time-dependent covariates. Although both corrected score and conditional score approaches are asymptotically equivalent, we recommend the conditional score approach for its better finite sample performance. The ratio term in (3.3) is a weighted average of the pseudo "observations" $\hat{R}_i^*(u)$, while the ratio term in (3.2) is a weighted average of the unadjusted estimated covariates $\hat{R}_i(u)$. In fact, as shown in Song and Huang (2005), the conditional score estimating equation is unbiased, which might account for the superiority of this estimator.

The model as presented can be extended to more general cases. For simplicity, we have focused on the case when the errors are independent across time. This

can be generalized to other error correlation structure, such as the exponential correlation [Diggle et al. (2002)]. In addition, the proportional hazards model may be generalized to include functions of the random effects by analogy to Song, Davidian and Tsiatis (2002a).

In summary, survival data with error-prone time-dependent covariates are commonly encountered in biomedical studies, econometrics, environmental studies and epidemiology. The method developed in this paper provides a mechanism for analyzing this type of data and closes multiple gaps among the areas of survival analysis, joint modeling and measurement error.

SUPPLEMENTARY MATERIAL

Supplementary materials for Partially time-varying coefficient proportional hazards models with error-prone time-dependent covariates—an application to the AIDS Clinical Trial Group 175 data (DOI: 10.1214/16-AOAS1003SUPP; .pdf). This supplement consists of four web appendices. Web Appendix A gives the asymptotic properties, the regularity conditions and the proofs of the asymptotic properties. Web Appendix B lists the lemmas used in the proofs. Web Appendix C derives the convergence result when the Hazard model is misspecified. Web Appendix D shows the analysis results of the ACTG 175 data including both log(CD4) and log(CD8).

REFERENCES

BYCOTT, P. and TAYLOR, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Stat. Med.* **17** 2061–2077.

CAI, J., FAN, J., JIANG, J. and ZHOU, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 141–158. MR2412635

DAFNI, U. G. and TSIATIS, A. A. (1998). Evaluating surrogate markers of clinical outcome measured with error. *Biometrics* **54** 1445–1462.

DEGRUTTOLA, V. and TU, X. M. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* **50** 1003–1014.

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. MR2049007

FAUCETT, C. L. and THOMAS, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.* **15** 1663–1685.

GU, L., WANG, L., HÄRDLE, W. K. and YANG, L. (2014). A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *TEST* **23** 806–843. MR3274476

HAMMER, S. M., KATEZSTEIN, D. A., HUGHES, M. D., GUNDAKER, H., SCHOOLEY, R. T., HAUBRICH, R. H., HENRY, W. K., LEDERMAN, M. M., PHAIR, J. P., NIU, M., HIRSCH, M. S. and MERIGAN, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *N. Engl. J. Med.* **335** 1081–1090.

HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostat.* **4** 465–480.

HSIEH, F., TSENG, Y.-K. and WANG, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62** 1037–1043. MR2297674

HUANG, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563. MR1742499

HUANG, J. Z. and LIU, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics* **62** 793–802. MR2247208

LIU, R. and YANG, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory* **26** 29–59.

NAN, B., LIN, X., LISABETH, L. D. and HARLOW, S. D. (2005). A varying-coefficient Cox model for the effect of age at a marker event on age at menopause. *Biometrics* **61** 576–583. MR2140931

PAWITAN, Y. and SELF, S. (1993). Modeling disease marker processes in AIDS. *J. Amer. Statist. Assoc.* **88** 719–726.

PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika* **69** 331–342.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operation criteria. *Stat. Med.* **8** 431–440.

SONG, X., DAVIDIAN, M. and TSIATIS, A. A. (2002a). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostat.* **3** 511–528.

SONG, X., DAVIDIAN, M. and TSIATIS, A. A. (2002b). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58** 742–753. MR1945011

SONG, X. and HUANG, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* **61** 702–714. MR2196158

SONG, X. and WANG, C. Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics* **64** 557–566, 669. MR2432426

SONG, X. and WANG, L. (2017). Supplement to "Partially time-varying coefficient proportional hazards models with error-prone time-dependent covariates—an application to the AIDS Clinical Trial Group 175 data." DOI:10.1214/16-AOAS1003SUPP.

TSIATIS, A. A. and DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88** 447–458. MR1844844

TSIATIS, A. A., DEGRUTTOLA, V. and WULFSOHN, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90** 27–37.

WANG, C. Y. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statist. Sinica* **16** 235–253. MR2256090

WANG, L. and YANG, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35** 2474–2503. MR2382655

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186

XU, J. and ZEGER, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *J. Roy. Statist. Soc. Ser. C* **50** 375–387. MR1856332

XUE, L. and LIANG, H. (2010). Polynomial spline estimation for a generalized additive coefficient model. *Scand. J. Stat.* **37** 26–46. MR2675938

XUE, L. and YANG, L. (2006). Additive coefficient modeling via polynomial spline. *Statist. Sinica* **16** 1423–1446. MR2327498

DEPARTMENT OF EPIDEMIOLOGY
AND BIOSTATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602
USA
E-MAIL: xsong@u.uga.edu

DEPARTMENT OF STATISTICS
AND STATISTICAL LABORATORY
IOWA STATE UNIVERSITY
AMES, IOWA 50011
USA
E-MAIL: lilywang@iastate.edu