Biostatistics (2002), **3**, 4, *pp.* 511–528
Printed in Great Britain

An estimator for the proportional hazards model with multiple longitudinal covariates measured with error

XIAO SONG, MARIE DAVIDIAN, ANASTASIOS A. TSIATIS
Department of Statistics, Box 8203, North Carolina State University, Raleigh NC 27695-8203, USA
xsong@stat.ncsu.edu

SUMMARY

In many longitudinal studies, it is of interest to characterize the relationship between a time-to-event (e.g. survival) and several time-dependent and time-independent covariates. Time-dependent covariates are generally observed intermittently and with error. For a single time-dependent covariate, a popular approach is to assume a joint longitudinal data–survival model, where the time-dependent covariate follows a linear mixed effects model and the hazard of failure depends on random effects and time-independent covariates via a proportional hazards relationship. Regression calibration and likelihood or Bayesian methods have been advocated for implementation; however, generalization to more than one time-dependent covariate may become prohibitive. For a single time-dependent covariate, Tsiatis and Davidian (2001) have proposed an approach that is easily implemented and does not require an assumption on the distribution of the random effects. This technique may be generalized to multiple, possibly correlated, time-dependent covariates, as we demonstrate. We illustrate the approach via simulation and by application to data from an HIV clinical trial.

Keywords: Conditional score; Measurement error; Mixed effects model; Proportional hazards model; Semiparametric; Surrogate marker.

1. INTRODUCTION

In longitudinal studies, information is collected on a time-to-event (e.g. 'survival') and time-dependent and time-independent covariates. A routine objective is to model the association between these covariates and survival, usually in the framework of the proportional hazards model (Cox, 1972). True values of the time-dependent covariates at each unique failure time are required for implementation; however, these are collected intermittently and are subject to error.

Such modeling is often carried out to identify time-independent and -dependent covariates associated with prognosis. An additional goal may be to assess the value of time-dependent covariates as potential surrogate markers (Prentice, 1989), often accomplished by examining the interrelationship between time-dependent covariates and treatment effect. These objectives are common in the analysis of HIV clinical trials. An example is AIDS Clinical Trials Group (ACTG) 175, a randomized clinical trial to compare zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine, or didanosine alone, in HIV-infected subjects (Hammer *et al.*, 1996). Between December 1991 and October 1992, 2467 subjects were recruited and followed until November 1994. CD4 and CD8 counts, both reflections of immune status, were measured for each participant about every 12 weeks after randomization, and the time-to-event endpoint was time to progression to AIDS or death. The focus of the study was to carry out treatment

comparisons; subsequently, an objective was to elucidate the relationship between prognosis and CD4 and CD8 and to investigate these measures as potential surrogate markers.

Because of the complication of intermittent, error-prone time-dependent covariates, early approaches imputed their values at each failure time by what are now regarded as naive methods. These include 'last value carried forward', which substitutes the last covariate value prior to the failure time, and 'naive regression', where least-squares estimates based on each subject's entire longitudinal profile are substituted. Such methods may perform adequately in some situations (e.g. Raboud *et al.*, 1993), but can lead to biased and misleading inference on the Cox model parameters (Prentice, 1982; Tsiatis and Davidian, 2001) and biased estimates of treatment effects and thus erroneous conclusions on surrogacy (Dafni and Tsiatis, 1998).

A popular approach for a single time-dependent covariate assumes that the longitudinal covariate data follow a linear mixed-effects model and that the hazard depends both on the random effects and other time-independent covariates through a proportional hazards relationship. One strategy for implementation of these 'joint models' is a two-stage, 'regression calibration' technique (Carroll *et al.*, 1995), where the mixed model is fitted to data at each risk set under normality assumptions and ensuing best linear unbiased predictors are used to impute covariate values at each failure time (e.g. Pawitan and Self, 1993; Tsiatis *et al.*, 1995; Bycott and Taylor, 1998; Dafni and Tsiatis, 1998). This method reduces bias relative to naive approaches but may still give erroneous results (Tsiatis and Davidian, 2001). An alternative approach yielding sound inferences is to base estimation on the joint likelihood of the survival and longitudinal data under parametric (normal) assumptions on the random effects (e.g. DeGruttola and Tu, 1994; Wulfsohn and Tsiatis, 1997; Faucett and Thomas, 1996; Henderson *et al.*, 2000; Xu and Zeger, 2001a). Generalization to multiple time-dependent covariates is complicated by the need to model the joint relationship among all covariates and the computational burden of potentially high-dimensional integration for likelihood and Bayesian approaches. Moreover, these methods rely on normality or other parametric assumptions for the random effects. Hu *et al.* (1998) proposed likelihood methods under weakening of this assumption for a single error-prone time-independent covariate, but adaptation to multiple time-dependent covariates would suffer the same drawbacks as above.

Tsiatis and Davidian (2001) proposed for such joint models a semiparametric conditional score estimator for the parameters in the hazard relationship in the case of a single time-dependent covariate and time-independent covariates. The random effects are treated as nuisance parameters for which a sufficient statistic may be derived, and a set of estimating equations based on conditioning on the sufficient statistic may be deduced that remove the dependence on the random effects. The estimators are consistent and asymptotically normal and yield unbiased and reliable inferences on the parameters of the Cox model in finite samples. No assumption need be made on the distribution of the random effects, and the estimator is fast and simple to compute.

These features make generalization to multiple covariates an attractive alternative. This is not trivial, however, as different covariates may be recorded on different schedules by happenstance or design and may be subject to errors that are correlated. In Section 2, we define a joint longitudinal-survival model with multiple time-dependent and time-independent covariates. The generalization of the conditional score approach is derived in Section 3, and in Section 4 we remark on large-sample properties. Section 5 presents an analysis of the ACTG 175 data, showing the utility of this method in the evaluation of several potential surrogate markers. Section 6 demonstrates the performance of the method via simulation.

2. MODEL DEFINITION

For each subject i, $i = 1, \ldots, n$, let T_i denote failure time and C_i denote censoring time. The observed survival data are $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leqslant C_i)$, where $I(\cdot)$ is the indicator function; these and all

other variables are independent across i. Let $X_{ik}(u)$, $k = 1, \ldots, K$, denote K time-dependent covariates at time u, and let the r-dimensional vector Z_i denote r time-independent covariates for subject i. We assume each covariate process $X_{ik}(u)$ satisfies

$$X_{ik}(u) = \alpha_{ik}^T f_k(u), \tag{1}$$

where $f_k(u)$ is a $(q_k \times 1)$ vector of functions of u, α_{ik} is a $(q_k \times 1)$ random effect, and f_k and α_{ik} may be different for each k. This allows flexibility in representing the time trajectory of each covariate via polynomial or spline models, e.g. $\alpha_{ik}^T f_k(u) = \alpha_{ik0} + \alpha_{ik1}u + \cdots + \alpha_{ik(q_k-1)}u^{q_k-1}$. The covariate processes $X_{ik}(u)$ are not observed directly; rather, longitudinal measurements $W_{ik}(t_{ikj})$ on the kth covariate are taken at times t_{ikj}, $j = 1, 2, \ldots, m_{ik}$, for each i, where

$$W_{ik}(t_{ikj}) = X_{ik}(t_{ikj}) + e_{ikj}, \tag{2}$$

and e_{ikj} are normally distributed mean-zero 'errors' with variance σ_{kk} that may reflect both biological variation and measurement error. Thus, (2) with (1) is a linear mixed-effects model, and $X_{ik}(u)$ may be regarded as the 'inherent' trajectory for subject i and covariate k.

We assume that the available measurements are sufficiently separated in time that serial correlation associated with within-subject biological variation is negligible; however, for the hazard formulation given below, this could be relaxed. We allow measurements on different covariates at the same time to be correlated. More formally, noting that j indexes measurement times separately for each k, we may write for $k, k' = 1, \ldots, K$, $j = 1, \ldots, m_{ik}$, and $j' = 1, \ldots, m_{ik'}$, $\mathrm{cov}(e_{ikj}, e_{ik'j'}) = \sigma_{kk'} I(t_{ikj} = t_{ik'j'})$. Here $\sigma_{kk'}$ is the covariance between errors from covariates k and k' measured at the same time point, reflecting correlation of components of within-subject biological variation, the measurement error, or both; this formulation subsumes the case $\sigma_{kk'} = 0$ for all $k \neq k'$. Ordinarily, measurement errors across k might be thought of as uncorrelated; however, we allow the possibility that some covariate measurements may be derived from a common blood sample or measuring technique, e.g. CD4 and CD8.

Let $e_i = (e_{i1}^T, \ldots, e_{iK}^T)^T$, where $e_{ik} = (e_{ik1}, \ldots, e_{ikm_{ik}})^T$; $t_{ik} = (t_{ik1}, \ldots, t_{ikm_{ik}})^T$ be the ordered times for subject i, covariate k, and $t_i = (t_{i1}^T, \ldots, t_{iK}^T)^T$ be the set of time points where observations on all K covariates are available; $m_i = (m_{i1}, \ldots, m_{iK})^T$; and $\alpha_i = (\alpha_{i1}^T, \ldots, \alpha_{iK}^T)^T$ $(q \times 1)$, $q = \sum_k q_k$. We assume that the conditional distribution of e_i given $(T_i, C_i, \alpha_i, Z_i, t_i, m_i)$ is normal with covariance matrix depending only on m_i and the parameters $\sigma_{kk'}$. The α_{ik} may be correlated across k; however, this will be of no consequence in the sequel. Likewise, no distributional assumption is placed on the α_i, nor is one needed.

A proportional hazards regression model is assumed for the relationship between the hazard of failure and the covariates; that is, the hazard for subject i is

$$\begin{aligned}\lambda_i(u) &= \lim_{\mathrm{d}u \to 0} \mathrm{d}u^{-1} \mathrm{pr}\{u \leqslant T_i < u + \mathrm{d}u | T_i \geqslant u, \alpha_i, Z_i, C_i, e_i(u), t_i(u)\} \\ &= \lambda_0(u) \exp\{\gamma^T G(u, \alpha_i) + \eta^T Z_i\}. \end{aligned} \tag{3}$$

Here, $\lambda_0(u)$ is an unspecified baseline hazard function; $G(u, \alpha_i)$ is a $(s \times 1)$ vector whose elements are functions of u and α_i; γ and η are $(s \times 1)$ and $(r \times 1)$, respectively; $t_i(u) = (t_{ikj} \leqslant u;\ k = 1, \ldots, K)$ denotes the observation times up to and including u; and $e_i(u) = (e_{ikj} : t_{ikj} \leqslant u;\ k = 1, \ldots, K)$. Equation (3) makes explicit the assumption that censoring, timing of measurements, and covariate errors are noninformative. Interest focuses on estimation of γ and η.

The vector $G(u, \alpha_i)$ allows flexibility in modeling the hazard relationship. For example, for

$$X_{i1}(u) = \alpha_{i10} + \alpha_{i11}u + \alpha_{i12}u^2, \quad X_{i2}(u) = \alpha_{i20} + \alpha_{i21}u + \alpha_{i22}u^2, \quad K = 2, \tag{4}$$

if $G(u, \alpha_i) = G(u)\alpha_i$,

$$G(u) = \begin{bmatrix} 1 & u & u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & u & u^2 \end{bmatrix}, \ \alpha_i = (\alpha_{i10}, \alpha_{i11}, \alpha_{i12}, \alpha_{i20}, \alpha_{i21}, \alpha_{i22})^T, \ \gamma = (\gamma_1, \gamma_2)^T \ (s = 2),$$

then $\gamma^T G(u, \alpha_i) = \gamma_1 X_{i1}(u) + \gamma_2 X_{i2}(u)$. Here, dependence of the hazard on α_i is linear. Alternatively, the dependence could be nonlinear; for $X_{i1}(u)$ and $X_{i2}(u)$ in (4) and $\gamma = (\gamma_1, \gamma_2)^T$, $G(u, \alpha_i) = (\alpha_{i10} + \alpha_{i11}u + \alpha_{i12}u^2, \alpha_{i20}^2)^T$ yields $\gamma^T G(u, \alpha_i) = \gamma_1 X_{i1}(u) + \gamma_2 \alpha_{i20}^2$. The representation in (3) also accommodates models involving interactions between covariates and time, which are useful for assessing the relevance of the proportional hazards assumption as suggested by Fleming and Harrington (1991, p. 173) and demonstrated in Section 5. To illustrate, for the representations of covariates in (4), the hazard relationship linear in the α_i given by

$$\gamma_1 X_{i1}(u) + \gamma_2 X_{i2}(u) + \gamma_3 X_{i2}(u)u \tag{5}$$

may be expressed as $\gamma^T G(u)\alpha_i$, with $\gamma = (\gamma_1, \gamma_2, \gamma_3)^T$, $s = 3$ and

$$G(u) = \begin{bmatrix} 1 & u & u^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & u & u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & u & u^2 & u^3 \end{bmatrix}.$$

The hazard formulation in (3) assumes that the survival distribution depends on the time-dependent covariates only through random effects that characterize the individual 'inherent trajectories'. An alternative modeling strategy would be to decompose the 'error' in (2) into components associated with serially correlated biological variation and measurement error and to assume the hazard depends both on random effects and biological variation (e.g. Henderson *et al.*, 2000). Specification of the hazard would be guided by subject-matter considerations.

3. CONDITIONAL SCORE ESTIMATOR

Assume initially that $G(u, \alpha_i) = G(u)\alpha_i$. The derivation of the proposed estimator follows the same reasoning as in Tsiatis and Davidian (2001), motivated by the conditional score method of Stefanski and Carroll (1987), but is complicated by multiple covariates measured at possibly different times. Let $\hat{\alpha}_{ik}(u)$ be the ordinary least-squares estimator of α_{ik} based on all the longitudinal data measured before time u for the kth covariate for subject i, that is, based on $t_{ik}(u) = (t_{ikj} \leqslant u)$. In order to obtain $\hat{\alpha}_{ik}(u)$ for each k, i must have at least q_k measurements on the kth covariate; thus, letting $m_{ik}(u)$ denote the number of time points in $t_{ik}(u)$, to ensure estimation of α_{ik} at u is possible for all $k = 1, \ldots, K$, we require $m_{ik}(u) \geqslant q_k$ for each k. Let the counting process increment be $\mathrm{d}N_i(u) = I(u \leqslant V_i < u + \mathrm{d}u, \Delta_i = 1, m_{ik}(u) \geqslant q_k, k = 1, \ldots, K)$; let $Y_i(u) = I(V_i \geqslant u, m_{ik}(u) \geqslant q_k, k = 1, \ldots, K)$ denote the 'at risk' process; and define $\omega = \{\sigma_{kk'} : k \geqslant k'\}$, the distinct parameters characterizing the variances and covariances of the errors. For $k = 1, \ldots, K$, define $F_{ik}(u) = [\, f_k(t_{ik1}), \ldots, f_k(t_{ikm_{ik}(u)})]^T$, $\{m_{ik}(u) \times q_k\}$, and let $I_{ikk'}(u)$ be the $\{m_{ik}(u) \times m_{ik'}(u)\}$ matrix whose (j, j') entry is $I(t_{ikj} = t_{ik'j'})$, for $j = 1, \ldots, m_{ik}(u)$, $j' = 1, \ldots, m_{ik'}(u)$. Because $G(u, \alpha_i) = G(u)\alpha_i$ is linear in α_i, it follows that, conditional on $\{Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$, $G(u)\hat{\alpha}_i(u)$, $i = 1, \ldots, n$, are independently distributed as $\mathcal{N}\{G(u)\alpha_i, \Sigma_i(u, \omega)\}$, where $\Sigma_i(u, \omega) = G(u)\Gamma_i(u, \omega)G^T(u)$ $(s \times s)$. Here,

$$\Gamma_i(u, \omega) = \begin{bmatrix} \Gamma_{i11}(u, \omega) & \Gamma_{i12}(u, \omega) & \cdots & \Gamma_{i1K}(u, \omega) \\ \Gamma_{i21}(u, \omega) & \Gamma_{i22}(u, \omega) & \cdots & \Gamma_{i1K}(u, \omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{iK1}(u, \omega) & \Gamma_{iK2}(u, \omega) & \cdots & \Gamma_{iKK}(u, \omega) \end{bmatrix}, \quad (q \times q),$$

where $\Gamma_{ikk'}(u,\omega) = \sigma_{kk'}\{F_{ik}^T(u)F_{ik}(u)\}^{-1}F_{ik}^T(u)I_{ikk'}(u)F_{ik'}(u)\{F_{ik'}^T(u)F_{ik'}(u)\}^{-1}$ $(q_k \times q_{k'})$.

For now, assume that ω is known. Conditional on $\{Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$, $\mathrm{d}N_i(u)$ is distributed as Bernoulli with probability $\lambda_0(u)\mathrm{d}u \exp\{\gamma^T G(u)\alpha_i + \eta^T Z_i\}$, and $\mathrm{d}N_i(u)$ and $G(u)\hat{\alpha}_i(u)$ are conditionally independent, from whence it may be shown that the conditional likelihood of $\{\mathrm{d}N_i(u), G(u)\hat{\alpha}_i(u)\}$ given $\{Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$ up to order $\mathrm{d}u$ is equal to

$$\exp\left\{S_i^T(u,\gamma,\omega)\Sigma_i^{-1}(u,\omega)G(u)\alpha_i\right\}\frac{\{\lambda_0(u)\mathrm{d}u\exp(\eta^T Z_i)\}^{\mathrm{d}N_i(u)}}{(2\pi)^{s/2}|\Sigma_i(u,\omega)|^{1/2}}$$
$$\times\exp\left\{-\frac{\hat{\alpha}_i^T(u)G^T(u)\Sigma_i^{-1}(u,\omega)G(u)\hat{\alpha}_i(u) + \alpha_i^T G^T(u)\Sigma_i^{-1}(u,\omega)G(u)\alpha_i}{2}\right\}, \tag{6}$$

where $S_i(u,\gamma,\omega) = G(u)\hat{\alpha}_i(u) + \mathrm{d}N_i(u)\Sigma_i(u,\omega)\gamma$. From (6), conditional on $Y_i(u) = 1$, $S_i(u,\gamma,\omega)$ is a complete sufficient statistic for α_i, which suggests that conditioning on $S_i(u,\gamma,\omega)$ would remove the dependence on α_i. Hence, the conditional hazard $\lambda_i\{u|S_i(u,\gamma,\omega)\} = \lim_{\mathrm{d}u\to 0}\mathrm{d}u^{-1}\Pr\{\mathrm{d}N_i(u) = 1|S_i(u,\gamma,\omega), Z_i, t_i(u), Y_i(u)\}$ may be shown to be equal to

$$\lambda_0(u)\exp\left\{\gamma^T S_i(u,\gamma,\omega) - \gamma^T\Sigma_i(u,\omega)\gamma/2 + \eta^T Z_i\right\}Y_i(u).$$

From these developments, we may apply the same reasoning as in Tsiatis and Davidian (2001) to deduce the conditional score estimating equations for γ and η, given by

$$\sum_{i=1}^n\int\left[\{S_i^T(u,\gamma,\omega), Z_i^T\}^T - \frac{E_1(u,\gamma,\eta,\omega)}{E_0(u,\gamma,\eta,\omega)}\right]\mathrm{d}N_i(u) = 0, \tag{7}$$

where $E_0(u,\gamma,\eta,\omega) = \sum_{i=1}^n E_{0i}(u,\gamma,\eta,\omega)$, $E_{0i}(u,\gamma,\eta,\omega) = \exp\{\gamma^T S_i(u,\gamma,\omega) - \gamma^T\Sigma_i(u,\omega)\gamma/2 + \eta^T Z_i\}Y_i(u)$, $E_1(u,\gamma,\eta,\omega) = \sum_{i=1}^n E_{1i}(u,\gamma,\eta,\omega)$, $E_{1i}(u,\gamma,\eta,\omega) = \{S_i^T(u,\gamma,\omega), Z_i^T\}^T E_{0i}(u,\gamma,\eta,\omega)$. When none of the K covariates is error-prone, (7) reduces to the usual score equations for the maximum partial likelihood estimator of Cox (1975).

The above developments rely on $G(u,\alpha_i) = G(u)\alpha_i$, so that the predictor in the hazard depends linearly on α_i and $G(u)\hat{\alpha}_i(u)$ is exactly normally distributed. For $G(u,\alpha_i)$ that may be nonlinear in α_i, we propose invoking a linear approximation. By the delta method, $G\{u,\hat{\alpha}_i(u)\}$ is asymptotically normally distributed with mean $G(u,\alpha_i)$ and variance $\Sigma_i(u,\alpha_i,\omega) = \dot{G}(u,\alpha_i)\Gamma_i(u,\omega)\dot{G}^T(u,\alpha_i)$ $(s \times s)$, where $\dot{G}(u,\alpha_i) = \partial G(u,\alpha_i)/\partial{\alpha_i}^T$. This suggests substituting $\dot{G}\{u,\hat{\alpha}_i(u)\}$ for $G(u)$ and $\Sigma_i\{u,\hat{\alpha}_i(u),\omega\}$ for $\Sigma_i(u,\omega)$ in (7) to obtain estimating equations for general $G(u,\alpha_i)$. We evaluate the performance of this approximation in Section 6.

4. Large-Sample Properties

Generally ω is unknown; however, under our assumptions, it may be estimated based on least-squares fits to all the data on each covariate for each subject when possible (i.e. $m_{ik} > q_k$). It is shown in the Appendix that an unbiased estimator for ω is $\hat{\omega}$, with element $\sigma_{kk'}$ estimated by

$$\hat{\sigma}_{kk'} = \frac{\sum_{i=1}^n I(m_{ik} > q_k, m_{ik'} > q_{k'}, m_{ikk'} > 0)R_{ik}^T A_{ikk'}^* R_{ik'}}{\sum_{i=1}^n I(m_{ik} > q_k, m_{ik'} > q_{k'}, m_{ikk'} > 0)\mathrm{tr}\{P_{ik}A_{ikk'}^* P_{ik'}A_{ikk'}^{*T}\}}, \tag{8}$$

where R_{ik}, $A_{ikk'}^*$, P_{ik}, and $m_{ikk'}$ are defined in the Appendix.

Let $S_i^*(u,\gamma,\omega) = \{S_i^T(u,\gamma,\omega), Z_i^T\}^T$ and $\bar{S}^*(u,\gamma,\eta,\omega) = E_1(u,\gamma,\eta,\omega)/E_0(u,\gamma,\eta,\omega)$. The combined centered estimating equations for $\tau = (\gamma^T, \eta^T, \omega^T)^T$ are $\sum_{i=1}^n \phi_i(\gamma,\eta,\omega) = 0$, where

$$\phi_i(\gamma,\eta,\omega) = \begin{bmatrix} \int \{S_i^*(u,\gamma,\omega) - \bar{S}^*(u,\gamma,\eta,\omega)\} \left\{ dN_i(u) - \dfrac{E_{0i}(u,\gamma,\eta,\omega)}{E_0(u,\gamma,\eta,\omega)} dN(u) \right\} \\ \psi_i(\omega) \end{bmatrix},$$

and $\psi_i(\omega)$ is a vector with the elements $I(m_{ik} > q_k, m_{ik'} > q_{k'}, m_{ikk'} > 0)[R_{ik}^T A_{ikk'}^* R_{ik'} - \sigma_{kk'} \mathrm{tr}\{P_{ik} A_{ikk'}^* P_{ik'} A_{ikk'}^{*T}\}]$, $k \geqslant k'$. By arguments similar to those in Tsiatis and Davidian (2001), the estimator $\hat{\tau} = (\hat{\gamma}^T, \hat{\eta}^T, \hat{\omega}^T)^T$ solving these equations is consistent and asymptotically normal. An estimator for the variance of $\hat{\tau}$ is the sandwich matrix $n^{-1}A^{-1}B(A^{-1})^T$, where $A = n^{-1}\sum_{i=1}^n \partial/\partial\tau^T\{\phi_i(\hat{\gamma},\hat{\eta},\hat{\omega})\}$, and $B = n^{-1}\sum_{i=1}^n \phi_i(\hat{\gamma},\hat{\eta},\hat{\omega})\phi_i^T(\hat{\gamma},\hat{\eta},\hat{\omega})$.

5. APPLICATION TO AIDS CLINICAL TRIALS GROUP 175

To demonstrate the utility of the methods for investigating associations among multiple time-dependent covariates and clinical endpoint and for elucidating the joint role of multiple covariates as potential surrogate markers, we apply the methods to the ACTG 175 data. As discussed by Xu and Zeger (2001b), there is scientific rationale for considering several time-dependent covariates simultaneously. The association between covariates and endpoint may reflect underlying mechanisms whose elucidation may enhance understanding of the disease and allow better prediction. Moreover, it may be fruitful to consider several covariates in evaluating surrogacy. Because of time and cost issues, interest has focused on identifying surrogate markers that could be substituted for the clinical endpoint in evaluation of treatment efficacy. According to Prentice (1989), (i) a surrogate marker should be prognostic for clinical outcome and (ii) the risk of progression given the marker should be independent of treatment. As Xu and Zeger (2001b) point out, associations with several covariates may reflect multiple biological pathways of treatment action. Even in the case where there is a single, predominant pathway, it may be characterized more effectively with more than one covariate.

These considerations suggest it would be advantageous to be able to entertain different models for the relationship between several time-dependent covariates and prognosis. An appealing feature of the conditional score approach is computational ease and speed, even for a large data set such as that for ACTG 175. The fit of each model in the analyses below took only a few minutes, demonstrating the feasibility of screening numerous, complex models in practice.

For the ACTG 175 data, we focus on features of the time trajectories of CD4 and CD8 (count per cubic millimeter) in combination and their association with the clinical endpoint (progression to AIDS or death) and their potential combined surrogacy. There were 308 events, with on average, 8.2 CD4 and 8.1 CD8 measurements per subject. CD4 was available with CD8 in all but two cases. The original analysis (Hammer *et al.*, 1996) found zidovudine alone to be inferior to the other three therapies; thus, for simplicity, we consider two treatment groups, zidovudine alone and the combination of the other three therapies, and let $Z_i = I$ (treatment $\neq$ zidovudine). Figure 1 presents CD4 and CD8 profiles for 10 randomly selected subjects and shows an apparent initial increase in both measures, with a peak at week 12, followed by a decline. Because only nine events occurred before week 12, for simplicity, we consider the data including and after week 12.

To achieve approximate within-subject normality and constant variance of CD4 and CD8 measurements, it is customary to transform these measures. We investigated several transformations, including square-, cube-, fourth-root, and logarithmic. Here, we report on results using base-10 logarithmic transformation for CD4 and CD8; use of base-10 logarithm is standard in the medical literature. The results for other transformations are qualitatively similar to those reported below.

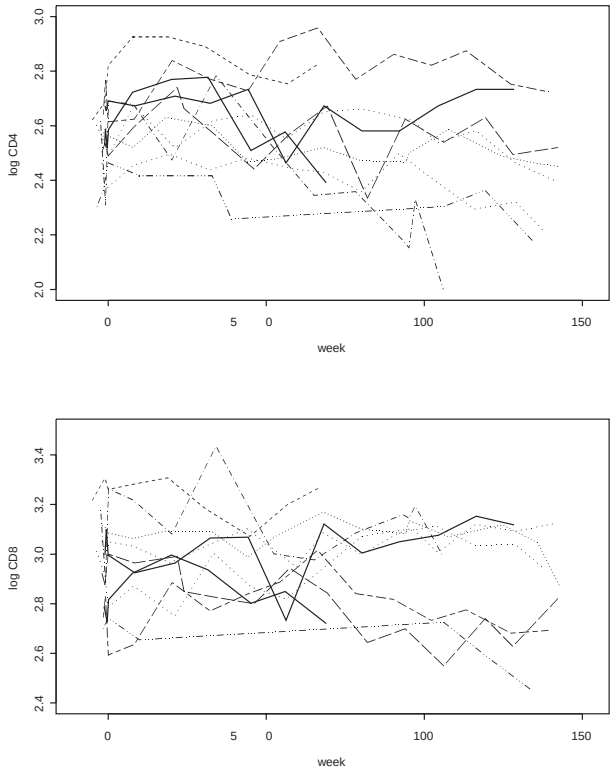

Fig. 1. Trajectories of log CD4 and log CD8 for 10 randomly selected subjects. The same type line is used for each subject in each panel.

Here, $K = 2$. From Figure 1, both log CD4, with inherent trajectory $X_{i1}(u)$ and log CD8, with inherent trajectory $X_{i2}(u)$, seem to follow approximate straight-line relationships $X_{ik}(u) = \alpha_{ik0} + \alpha_{ik1}u$, $k = 1, 2$, after week 12. To assess if such simple linear mixed-effects models suffice to represent inherent log CD4 and log CD8, we investigated formally whether a quadratic relationship as in (4) provides a better characterization using a conditional F-test, which is in the spirit of leaving the distribution of the α_{ik}, $k = 1, 2$, unspecified. That is, for each k, letting W_{ik} denote the vector of log-transformed measurements for subject i, we regarded the α_{ik}, $i = 1, \ldots, n$, as fixed parameters in an overall model for W_{ik}, $i = 1, \ldots, n$, and fitted the linear ('reduced') and quadratic ('full') versions of the model by ordinary least squares using the data for subjects with $m_{ik} > 3$ observations. We then constructed the usual F-statistic based on the error sums of squares for the 'full' and 'reduced' models. As discussed in the Appendix, under our assumption on the conditional distribution of e_i given $(T_i, C_i, \alpha_i, t_i, m_i)$, the estimators for the variance σ_{kk} for each k will be unbiased. Thus, because the usual F-statistic is a function of this estimator for each k, the resulting test should have the usual properties. In particular, letting N_k and n_k denote the total number of observations and total number of subjects with $m_{ik} > 3$, the test statistic under the null hypothesis ('reduced' model) and normal errors has an F distribution with $(n_k, N_k - 3n_k)$ degrees of freedom. Because N_k and n_k are large, the appropriate critical value is virtually 1. For log CD4 ($k = 1$) and log CD8 ($k = 2$), the statistics are 2.576 and 1.854, respectively, suggesting that the quadratic model (4) provides a better characterization for each, leading us to adopt this model for both log CD4 and log CD8.

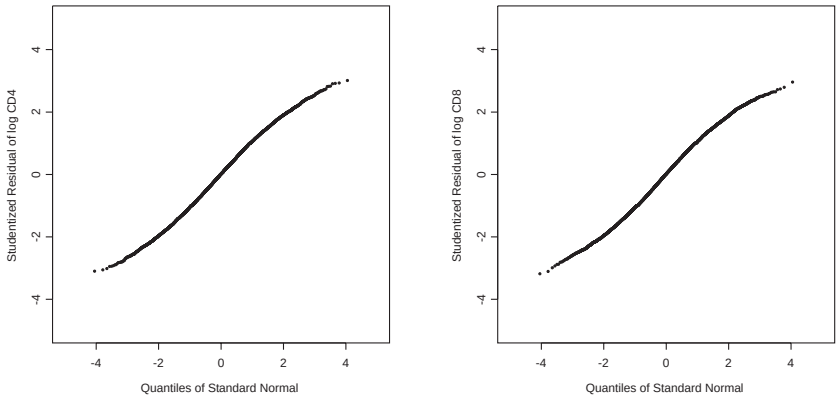

Fig. 2. Q–Q plots for studentized residuals for log CD4 and log CD8.

Figure 2 shows Q–Q plots of studentized residuals for each quadratic model fit. Under our assumptions on the conditional distribution of e_i, these residuals should have the usual linear model properties for each k, so should reflect deviation from normality in the conventional manner. The plots suggest that the error distribution may be short-tailed relative to the normal but symmetric. Although the derivation of Section 3 requires the errors e_{ikj} to be normally distributed, evidence in Section 6 shows that the conditional score method is not affected by short-tailed error distributions. The sample correlation coefficient for studentized residuals of log CD4 and log CD8 is 0.58, indicating that it would be inappropriate to assume that the covariance $\sigma_{12} = 0$.

To assess the association of log CD4 and log CD8 with prognosis and their role as surrogate markers, we carried out several analyses.

1. We fitted the model with treatment only, $\lambda_i(u) = \lambda_0(u)\exp\{\eta Z_i\}$, by usual partial likelihood methods (model 1).
2. We investigated various combinations of features of the inherent trajectories of log CD4 and log CD8 to determine which combination(s) are most strongly associated with the endpoint. In particular, we considered a series of models of the general form $\lambda_i(u) = \lambda_0(u)\exp[\gamma^T\{G(u)\alpha_i^T\}]$, where $G(u)\alpha_i^T$ was chosen to represent different linear combinations of current value and intercept of log CD4 and log CD8, and fitted them using the conditional score method.
   Because the conditional score approach is not likelihood-based, we used the Wald statistic for the null hypothesis $\gamma = 0$ for each model, which has approximately a $\chi^2_{\dim(\gamma)}$ distribution, to reflect strength of association. The estimates and Wald statistics for several models are shown in Table 1 (models 2.1–2.8). The table excludes models that resulted in unstable estimation, gauged by a very large ratio of largest to smallest eigenvalues of the estimated covariance matrix for $\hat{\gamma}$. With the exception of model 2.8, this occurred only for models with $\dim(\gamma) > 2$. Among models with $\dim(\gamma) > 2$, none showed appreciable improvement in the Wald statistic over those in the table. In all models except 2.8, the sign of the coefficients for current value and intercept of log CD4 and log CD8 is consistent with scientific understanding; as both measures reflect status of the body's immune response, the higher the better, decrease of the hazard with increased CD4/CD8 is biologically plausible.
   Among all the models, model 2.5 with $\lambda_i(u) = \lambda_0(u)\exp\{\gamma_1 X_{i1}(u) + \gamma_2 X_{i2}(u)\}$ involving the combination of current values of log CD4 and log CD8 leads to the largest Wald statistic. From Table 1, model 2.1 with log CD4 alone yields the second-largest Wald statistic; adding intercept of

log CD8 to this model (model 2.6) results in virtually no improvement, while adding current value as in model 2.5 yields a considerably larger statistic. Models 2.3 and 2.4 indicate that features of log CD8 may be associated with prognosis on their own.

All of the models considered embody the assumption of proportional hazards. We assessed the adequacy of this assumption using the following approach, which we illustrate using model 2.5. Following a suggestion of Fleming and Harrington (1991, p. 173), we re-fitted the model separately for the data in two time intervals, corresponding roughly to the events occurring before and after the median event time. According to these authors, under proportional hazards, we would expect the estimates for γ to be comparable in each interval. The results in the first part of Table 2 indicate that the proportional hazards assumption may be reasonable, but there is some suggestion of possible deviation from it. To investigate further, we followed Fleming and Harrington (1991, p. 173) and considered two additional models, model 2.5*, which includes log CD4, log CD8, and the interaction of time and log CD4, and model 2.5**, including the terms in model 2.5 plus an additional interaction of time and log CD8. Evidence that the coefficient of an interaction term is non-zero may reflect a deviation from the proportional hazards assumption; thus, unlike the diagnostic above, this approach allows formal significance testing. As noted in Section 2, fitting these models may be implemented straightforwardly in our framework, and the interaction parameters may be estimated and tested using the conditional score method with no additional difficulty, as in (5). The second part of Table 2 shows the results. The fit of model 2.5** is entirely similar to that of model 2.5, with no evidence of an interaction of log CD8 with time (p-value = 0.908), indicating no deviation from proportional hazards. The fit of model 2.5* suggests a decreasing hazard relationship to log CD4 over time (p-value for the interaction term = 0.016). Although the interaction effect of log CD4 and time is significant at level 0.05, the practical implication is modest. From the fit of model 2.5*, predicted log hazard ratios for log CD4 at the 10th, 50th, and 90th percentiles of the observed distribution of failure times are -2.620, -1.972, and -1.514, respectively, compared with the constant estimate of -1.838 from model 2.5. Thus, whether the interaction term is included or not, the relationship with log CD4 is qualitatively similar. Results for other models in Table 2 resemble those here.

Different data analysts may make different subjective judgments over whether ease of interpretation of a proportional hazards outweighs the importance of the possible, practically modest departure from this assumption in gauging the association of the covariates with prognosis. In the sequel, we focus for illustration on proportional hazards models but also indicate results for models including interaction terms.

3. We fitted several of the models showing evidence of strong associations in step 2 augmented by an additive effect of treatment in the hazard. For example, for model 2.5, we considered $\lambda_i(u) = \lambda_0(u) \exp\{\gamma_1 X_{i1}(u) + \gamma_2 X_{i2}(u) + \eta Z_i\}$. (These models are denoted as model 3.X below.)

Models for step 2 were used to deduce the most productive combination for describing the association between covariates and prognosis. The results for models 2.5 suggest that, although current value of log CD4 is strongly prognostic on its own, consideration of current value of log CD8 offers additional explanatory value.

Models from step 2 also were used to verify Prentice's condition (i), which would be indicated by $\gamma = (\gamma_1, \gamma_2)^T \neq 0$. Model 1 and those in step 3 were used to investigate condition (ii); $\eta \neq 0$ in model 1 but $\eta = 0$ in models in step 3 would suggest that the treatment effect is mediated through the relevant combination of features of log CD4 and log CD8. The estimate of the parameters of the error covariance matrix, $\omega^T = (\sigma_{11}, \sigma_{12}, \sigma_{22})$, is (0.009 28, 0.005 58, 0.010 09). For comparison, models from step 2 and 3 were also fitted using the naive regression method.

Table 1. *Results for several models fitted to the ACTG 175 data. Entries are the estimate and its estimated standard error in parentheses. The fit of model 2.8 is suspect, as the Wald statistic decreases relative to that for model 2.2 and the coefficient of intercept of log CD8 is positive. This model yielded a large eigenvalue ratio, suggesting instability*

	Current log CD4	Intercept of log CD4	Current log CD8	Intercept of log CD8	Wald statistic
Model 2.1	−2.033 (0.146)				193.577
Model 2.2		−1.568 (0.405)			22.095
Model 2.3			−1.722 (0.593)		8.440
Model 2.4				−0.994 (0.342)	8.442
Model 2.5	−1.838 (0.147)		−0.982 (0.215)		312.370
Model 2.6	−2.034 (0.150)			0.037 (0.668)	193.598
Model 2.7		−1.842 (0.324)	−1.532 (0.321)		43.490
Model 2.8		−1.819 (0.452)		0.209 (0.055)	18.529*

Table 2. *Diagnostic checks of the proportional hazards assumption in model 2.5*

Fits of model 2.5 using data from the first and second time intervals					
	$\hat{\gamma}_1$	SE($\hat{\gamma}_1$)	$\hat{\gamma}_2$	SE($\hat{\gamma}_2$)	
Using first half events	−2.072	0.233	−0.897	0.318	
Using second half events	−1.533	0.179	−1.362	0.275	
Fits of models 2.5* and 2.5**					
	log CD4	log CD8	log CD4×time	log CD8×time	Wald statistic
Model 2.5*	−3.030 (0.552)	−1.073 (0.250)	0.012 (0.005)		327.178
Model 2.5**	−1.844 (0.169)	−1.032 (0.435)		0.001 (0.006)	312.133

To illustrate, results for models 1, 2.5, and 3.5, and involving current values of log CD4 and log CD8 in combination, are shown in Table 3; *p*-values are for the corresponding Wald statistic for the null hypothesis that each entry is equal to zero. Treatment alone (model 1) yields a *p*-value of 0.003, indicating a strong effect, as deduced by Hammer *et al.* (1996). For model 2.5, the results from the conditional score and naive regression methods differ considerably; the naive regression analysis suggests no association with current log CD8, while the conditional score shows strong association with both covariates. The model 3.5 conditional score results show that the treatment effect is not significant after adjustment for log CD4 and log CD8, which supports Prentice's condition (ii) that the treatment effect is mediated through a combination of log CD4 and log CD8. These results are in marked contrast to the conclusion that would be reached from the naive regression fit, which shows a significant treatment effect even after adjustment. Table 3 also includes results for models 2.5* and 3.5*, which yield the same qualitative conclusions.

The conditional score results suggest that some combination of current values of log CD4 and log CD8 is a surrogate for the treatment considered in ACTG 175. We carried out similar analyses (not shown) for models 2.1 (current log CD4 alone) and 2.3 (current log CD8 alone) and the augmented versions of these models including a treatment effect (models 3.1 and 3.3, respectively). In model 3.1, the estimate of the treatment effect is insignificant, indicating that current value of log CD4 on its own represents a potential surrogate marker. As with model 3.5, the naive regression analysis of this model yields a highly significant treatment effect. In contrast, the conditional score estimate of treatment effect in model 3.3 is

Table 3. *Fits of models 1, 2.5 and 2.5*, and 3.5 and 3.5* for the ACTG 175 data*

Model 1						
Terms included	Est.	SE	p-value			
Treatment	−0.375	0.128	0.003			
Models 2.5 and 2.5*						
			Conditional score			
Terms included	Est.	SE	p-value	Est.	SE	p-value
Current log CD4	−1.838	0.147	0.000	−3.030	0.552	0.000
Current log CD8	−0.982	0.215	0.000	−1.073	0.250	0.000
log CD4 × time				0.012	0.005	0.016
			Naive regression			
Terms included	Est.	SE	p-value	Est.	SE	p-value
Current log CD4	−0.116	0.013	0.000	−0.801	0.107	0.000
Current log CD8	−0.008	0.008	0.308	0.006	0.009	0.490
log CD4 × time				0.006	0.001	0.000
Models 3.5 and 3.5*						
			Conditional score			
Terms included	Est.	SE	p-value	Est.	SE	p-value
Current log CD4	−1.831	0.150	0.000	−3.003	0.551	0.000
Current log CD8	−0.999	0.212	0.000	−1.077	0.247	0.000
log CD4 × time				0.012	0.005	0.019
Treatment	−0.190	0.194	0.326	−0.066	0.207	0.749
			Naive regression			
Terms included	Est.	SE	p-value	Est.	SE	p-value
Current log CD4	−0.119	0.014	0.000	−0.816	0.110	0.000
Current log CD8	−0.009	0.008	0.274	0.006	0.009	0.518
log CD4 × time				0.006	0.001	0.000
Treatment	−0.408	0.138	0.003	−0.416	0.139	0.003

highly significant (p-value $= 0.006$), indicating that current log CD8 on its own may not be a plausible surrogate.

Given these results, we investigated whether there is added benefit to considering both measures as a surrogate relative to current log CD4 only. Following a procedure similar to that of Xu and Zeger (2001b, p. 83), we inspected point estimates and 95% Wald confidence intervals for the treatment effect in models 3.1 and 3.5, −0.074 (−0.514, 0.366) and −0.194 (−0.570, 0.189), respectively. The intervals overlap considerably, and the estimated treatment effect is greater in magnitude for model 3.5, suggesting that adding current log CD8 offers little benefit. Analyses using models incorporating interaction terms yield similar qualitative conclusions.

Summarizing, the results of competing model fits imply that while the combination of current log CD4 and log CD8 is highly associated with prognosis, the strength of the combination of current log CD4 and log CD8 as a surrogate is derived primarily through log CD4. Use of the naive approach to carry out these analyses would potentially result in erroneous inferences on both association and surrogacy. Although in this example there was no benefit to considering more than one covariate as a surrogate, the availability of accessible methods for fitting multiple-covariate models allowed this possibility to be considered. Had log CD4 and log CD8 considered separately not shown evidence of surrogacy, consideration of their joint association would permit evaluation of their potential surrogacy in combination.

6. SIMULATION STUDIES

We conducted several simulation experiments, based roughly on the situation of the ACTG 175 data. For each of n subjects, the treatment indicator Z_i was generated according to Bernoulli distribution with probability 0.5. For simplicity, we took log CD4 and log CD8 to follow a linear mixed-effects model, i.e. $X_{i1}(u) = \alpha_{i10} + \alpha_{i11}u$ for log CD4 and $X_{i2}(u) = \alpha_{i20} + \alpha_{i21}u$ for log CD8, with $t_{ijk} = 0, 12, 24, 36, \ldots, 168$ weeks. The true hazard relationship was $\lambda_i(u) = \lambda_0(u)\exp\{\gamma_1 X_{i1}(u) + \gamma_2 X_{i2}(u) + \eta Z_i\}$, with $\gamma_1 = -2.0$, $\gamma_2 = -1.0$, and $\eta = 0$. The baseline hazard $\lambda_0(u) = 2$ for $u > 12$ and 0 otherwise. To mimic the censoring of the ACTG 175 data, censoring was generated from an exponential distribution with mean 250, leading to 86.3% and 87.1% censoring, respectively, for the normal and bimodal random effects distributions described below. To represent missed visits that are often seen in practice, after week 12, (CD4,CD8) pairs were missing with probability 0.1 at each time point for all subjects.

We considered several scenarios representing different magnitudes of error and distributions for the random effects as follows. True values for the variances of the errors e_{i1j} and e_{i2j} and covariance matrices of α_{i1} and α_{i2} were specified from separate linear mixed-effects model fits to the ACTG 175 data for log CD4 and log CD8. We added correlation between errors at the same time using the value observed for the correlation among least-squares residuals, and chose arbitrary but non-negligible values for cross-correlations among elements of α_{i1} and α_{i2}. The errors $e_{ij} = (e_{i1j}, e_{i2j})^T$ were initially taken to be independent and identically normally distributed with covariance matrix D such that $\text{vech}(D) = (0.010\,89, 0.006\,28, 0.011\,14)^T$. The random effects α_i were initially taken to be jointly normal with mean $(2.5915, -0.003\,15, 2.9329, -0.000\,48)^T$ for $Z_i = 0$ and mean $(2.5915, -0.001\,98, 2.9329, -0.000\,38)^T$ for $Z_i = 1$ and covariance matrix

$$\begin{bmatrix} 0.024\,08 & -0.000\,08 & 0.010\,36 & -0.000\,06 \\ -0.000\,08 & 0.0000\,14 & -0.000\,21 & 6.113 \times 10^{-6} \\ 0.010\,36 & -0.000\,21 & 0.035\,40 & -0.000\,11 \\ -0.000\,06 & 6.113 \times 10^{-6} & -0.000\,11 & 3.451 \times 10^{-6} \end{bmatrix}. \tag{9}$$

We carried out simulations for the scenario assuming D and (9) for $n = 250$ and 500; results for larger sample sizes are similar to those for $n = 500$. For each n, we also considered the cases of a fourfold increase in measurement error, i.e. taking the covariance matrix of e_{ij} to be four times D, replacing the normal specification for α_i by a bimodal mixture of normals described below, and both of these together. For the bimodal specification, α_{i1} and α_{i2} were taken to be independently distributed, each following a bimodal mixture of normals as described in Davidian and Gallant (1993) with mixing proportion $p = 0.5$ and sep $= 4$ for α_{i1} and $p = 0.3$ and sep $= 3$ for α_{i2} with same means and covariance matrices as in the normal case above.

For each scenario, 1000 Monte Carlo data sets were simulated. For each data set, we fitted the model above in four ways: (i) using the 'ideal' approach in which the true values of $X_{i1}(u)$ and $X_{i2}(u)$ at each failure time were used (fitting using the usual partial likelihood method); (ii) using the conditional score estimator, (iii) using naive regression; (iv) using LVCF. For (i), (iii), and (iv), we estimated standard errors as for the usual time-dependent Cox model and used the approach in Section 4 for the conditional score. For all methods, 95% Wald confidence intervals for γ_1, γ_2, and η were constructed.

The results are given in Tables 4 and 5. In all cases and for both sample sizes, the conditional score estimator shows negligible bias close to that of the unachievable 'ideal' estimator. For $n = 500$, coverage probabilities are close to or achieve the nominal level. For $n = 250$ and fourfold error, the coverage is a little below the nominal level. This is not surprising; with such large censoring rates, $n = 250$ corresponds to a case with roughly 33 events. That the conditional score estimator yields reasonable results under these difficult conditions is encouraging. For all sample sizes, the conditional score estimator provides unbiased

Table 4. *Simulation results for sample size* $n = 250$. *'1-fold' and '4-fold' refer to error according to* D *and* $4D$*, as described in the text*

Error		1-fold			4-fold		
Method	I	CS	N	LV	CS	N	LV
				Normal α_i			
$\hat{\gamma}_1$	−2.047	−2.055	−2.029	−2.060	−2.080	−1.971	−1.748
SD($\hat{\gamma}_1$)	0.588	0.615	0.592	0.573	0.712	0.599	0.494
SE($\hat{\gamma}_1$)	0.565	0.581	0.560	0.558	0.674	0.543	0.496
CP($\hat{\gamma}_1$)	0.954	0.944	0.946	0.954	0.942	0.932	0.920
$\hat{\gamma}_2$	−1.056	−1.065	−0.972	−0.667	−1.136	−0.753	−0.193
SD($\hat{\gamma}_2$)	0.958	1.081	0.994	0.831	1.717	1.035	0.653
SE($\hat{\gamma}_2$)	0.940	1.025	0.917	0.848	1.335	0.854	0.673
CP($\hat{\gamma}_2$)	0.944	0.934	0.919	0.941	0.928	0.876	0.782
$\hat{\eta}$	0.004	0.005	0.002	−0.012	0.004	−0.004	−0.041
SD($\hat{\eta}$)	0.377	0.381	0.377	0.371	0.397	0.376	0.366
SE($\hat{\eta}$)	0.362	0.363	0.362	0.361	0.378	0.362	0.358
CP($\hat{\eta}$)	0.949	0.947	0.953	0.958	0.953	0.953	0.953
				Bimodal α_i			
$\hat{\gamma}_1$	−2.070	−2.074	−2.021	−1.956	−2.108	−1.904	−1.580
SD($\hat{\gamma}_1$)	0.496	0.524	0.498	0.483	0.630	0.507	0.435
SE($\hat{\gamma}_1$)	0.492	0.509	0.487	0.491	0.597	0.472	0.441
CP($\hat{\gamma}_1$)	0.962	0.945	0.954	0.961	0.935	0.929	0.833
$\hat{\gamma}_2$	−1.016	−1.029	−0.910	−0.604	−1.089	−0.663	−0.077
SD($\hat{\gamma}_2$)	0.852	0.958	0.884	0.766	1.271	0.918	0.617
SE($\hat{\gamma}_2$)	0.824	0.910	0.801	0.748	1.170	0.744	0.598
CP($\hat{\gamma}_2$)	0.951	0.940	0.926	0.910	0.921	0.869	0.652
$\hat{\eta}$	−0.008	−0.008	−0.013	−0.037	−0.004	−0.026	−0.076
SD($\hat{\eta}$)	0.387	0.390	0.386	0.381	0.403	0.385	0.376
SE($\hat{\eta}$)	0.373	0.374	0.373	0.371	0.387	0.372	0.370
CP($\hat{\eta}$)	0.955	0.952	0.957	0.953	0.954	0.957	0.953

I, 'ideal' method; CS, Conditional Score; N, naive regression; LV, last value carried forward; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, coverage probability of 95% Wald confidence interval.

estimation regardless of the distribution of α_i. In contrast, the naive regression and LVCF approaches can yield biased estimates and coverage probabilities well below the nominal level. For onefold error, the results for conditional score and naive regression are similar to the ideal, but LVCF performance is degraded. For fourfold error, differences among the methods are more dramatic.

To assess the performance of conditional score method when the error distribution is short-tailed compared with the normal, as in the ACTG 175 analysis, we conducted simulations taking the error

Table 5. *Simulation results for sample size $n = 500$. '1-fold' and '4-fold' refer to error according to D and $4D$, as described in the text*

Error		1-fold			4-fold		
Method	I	CS	N	LV	CS	N	LV
				Normal α_i			
$\hat{\gamma}_1$	−2.052	−2.054	−2.034	−2.061	−2.065	−1.978	−1.756
SD($\hat{\gamma}_1$)	0.406	0.429	0.413	0.389	0.492	0.422	0.339
SE($\hat{\gamma}_1$)	0.388	0.398	0.384	0.384	0.452	0.373	0.343
CP($\hat{\gamma}_1$)	0.945	0.936	0.934	0.944	0.944	0.921	0.891
$\hat{\gamma}_2$	−0.978	−0.991	−0.900	−0.630	−1.042	−0.696	−0.181
SD($\hat{\gamma}_2$)	0.684	0.757	0.711	0.611	0.994	0.748	0.483
SE($\hat{\gamma}_2$)	0.653	0.717	0.638	0.592	0.929	0.597	0.472
CP($\hat{\gamma}_2$)	0.939	0.938	0.923	0.895	0.941	0.850	0.596
$\hat{\eta}$	0.005	0.006	0.003	−0.012	0.008	−0.003	−0.042
SD($\hat{\eta}$)	0.256	0.258	0.256	0.253	0.264	0.255	0.250
SE($\hat{\eta}$)	0.252	0.252	0.252	0.251	0.259	0.251	0.250
CP($\hat{\eta}$)	0.946	0.947	0.948	0.946	0.945	0.951	0.949
				Bimodal α_i			
$\hat{\gamma}_1$	−2.029	−2.032	−1.987	−1.929	−2.058	−1.878	−1.567
SD($\hat{\gamma}_1$)	0.348	0.365	0.348	0.338	0.432	0.349	0.302
SE($\hat{\gamma}_1$)	0.336	0.347	0.333	0.338	0.408	0.323	0.306
CP($\hat{\gamma}_1$)	0.950	0.942	0.937	0.938	0.939	0.906	0.684
$\hat{\gamma}_2$	−1.050	−1.068	−0.964	−0.651	−1.142	−0.738	−0.119
SD($\hat{\gamma}_2$)	0.592	0.660	0.607	0.519	0.876	0.622	0.410
SE($\hat{\gamma}_2$)	0.570	0.625	0.555	0.520	0.825	0.518	0.417
CP($\hat{\gamma}_2$)	0.944	0.944	0.930	0.900	0.930	0.871	0.426
$\hat{\eta}$	0.004	0.003	−0.001	−0.024	0.004	−0.013	−0.061
SD($\hat{\eta}$)	0.261	0.263	0.261	0.258	0.270	0.260	0.255
SE($\hat{\eta}$)	0.258	0.259	0.258	0.257	0.265	0.258	0.257
CP($\hat{\eta}$)	0.950	0.947	0.951	0.951	0.947	0.954	0.945

I, 'ideal' method; CS, Conditional Score; N, naive regression; LV, last value carried forward; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, coverage probability of 95% Wald confidence interval.

distribution as the normal truncated at three standard deviations. The results for $n = 500$ and fourfold error are shown in Table 6. The conditional score performs well, while the naive methods are subject to bias. For $G(u, \alpha_i)$ nonlinear in α_i, we also carried out simulations to investigate the performance of the linear approximation in Section 3, using the same scenarios as above except that the true hazard was taken to be $\lambda_i(u) = \lambda_0(u) \exp\{\gamma_1 X_{i1}(u) + \gamma_2 \alpha_{i20}^2 + \eta Z_i\}$, with $\gamma_1 = -2.0$, $\gamma_2 = -0.2$ and $\eta = 0$. Results

Table 6. *Simulation results with $n = 500$ and '4-fold' error according to 4D as described in the text: short-tailed (truncated normal) error*

α_i	Normal				Bimodal			
Method	I	CS	N	LV	I	CS	N	LV
$\hat{\gamma}_1$	−2.039	−2.057	−1.971	−1.762	−2.039	−2.073	−1.890	−1.578
SD($\hat{\gamma}_1$)	0.408	0.490	0.416	0.336	0.357	0.452	0.362	0.302
SE($\hat{\gamma}_1$)	0.388	0.450	0.374	0.345	0.337	0.410	0.324	0.306
CP($\hat{\gamma}_1$)	0.946	0.937	0.920	0.883	0.934	0.936	0.902	0.706
$\hat{\gamma}_2$	−0.976	−1.028	−0.698	−0.192	−1.008	−1.080	−0.695	−0.106
SD($\hat{\gamma}_2$)	0.687	0.943	0.722	0.472	0.568	0.839	0.601	0.405
SE($\hat{\gamma}_2$)	0.651	0.914	0.597	0.474	0.566	0.815	0.516	0.419
CP($\hat{\gamma}_2$)	0.934	0.940	0.857	0.599	0.948	0.950	0.848	0.411
$\hat{\eta}$	0.013	0.017	0.008	−0.029	0.009	0.014	−0.008	−0.058
SD($\hat{\eta}$)	0.248	0.258	0.248	0.242	0.249	0.261	0.250	0.244
SE($\hat{\eta}$)	0.251	0.258	0.251	0.250	0.257	0.265	0.257	0.256
CP($\hat{\eta}$)	0.956	0.955	0.959	0.958	0.959	0.958	0.955	0.955

I, 'ideal' method; CS, Conditional Score; N, naive regression; LV, last value carried forward; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, coverage probability of 95% Wald confidence interval.

for the 'ideal', conditional score and naive regression methods with $n = 500$ and fourfold error are in Table 7, and show that the linear approximation yields good performance. The estimates of the parameter ω containing the error variances σ_{11} and σ_{22} and the covariance σ_{12} for all simulations are given in Table 8 and demonstrate that the proposed estimator for this parameter given in Section 4 is unbiased.

Overall, the simulation evidence suggests that the conditional score estimator yields reliable inferences under a broad range of conditions.

7. Discussion

We have proposed an estimator for a joint model for survival and longitudinal data for which the hazard relationship may depend on multiple time-dependent covariates measured at potentially different time points whose observation is subject to error. The method offers a feasible approach to a problem that entails considerable computational complexity otherwise. The estimator is easy to compute and does not require an assumption on the distribution of underlying random effects nor the need to model or identify this distribution. Application of the method and simulation evidence show that it is accessible for routine use and leads to reliable inference.

We implemented the conditional score method in the C++ language using Newton–Raphson to solve the conditional score estimating equations; code is available upon request.

Table 7. *Simulation results with $n = 500$ and '4-fold' error according to 4D as described in the text: nonlinear $G(u, \alpha_i)$*

α_i		Normal			Bimodal	
Method	I	CS	N	I	CS	N
$\hat{\gamma}_1$	−2.020	−2.033	−1.816	−2.023	−2.050	−1.850
SD($\hat{\gamma}_1$)	0.240	0.315	0.250	0.242	0.311	0.242
SE($\hat{\gamma}_1$)	0.243	0.303	0.230	0.243	0.305	0.230
CP($\hat{\gamma}_1$)	0.951	0.941	0.849	0.946	0.949	0.871
$\hat{\gamma}_2$	−0.203	−0.211	−0.130	−0.206	−0.213	−0.134
SD($\hat{\gamma}_2$)	0.076	0.103	0.065	0.078	0.111	0.069
SE($\hat{\gamma}_2$)	0.076	0.103	0.063	0.077	0.106	0.063
CP($\hat{\gamma}_2$)	0.950	0.954	0.791	0.946	0.953	0.780
$\hat{\eta}$	−0.002	−0.001	−0.018	0.006	0.010	−0.006
SD($\hat{\eta}$)	0.165	0.169	0.163	0.166	0.170	0.165
SE($\hat{\eta}$)	0.165	0.169	0.164	0.164	0.169	0.164
CP($\hat{\eta}$)	0.956	0.962	0.964	0.946	0.943	0.948

I, 'ideal' method; CS, Conditional Score; N, naive regression; SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, coverage probability of 95% Wald confidence interval.

ACKNOWLEDGEMENTS

This work was supported by NIH grants R01-AI31789, R01-CA85848, and R01-CA51962. The authors are grateful to Michael Hughes, Heather Gorski, and the AIDS Clinical Trials Group for providing the data from ACTG 175.

APPENDIX. DERIVATION OF (8)

For subject i, let $W_{ik} = \{W_{ik}(t_{ik1}), \ldots, W_{ik}(t_{ikm_{ik}})\}^T$ be the full set of observations on the kth covariate. For each k, $W_{ik} = F_{ik}\alpha_{ik} + e_{ik}$, where $F_{ik} = [\, f_k(t_{ik1}), \ldots, f_k(t_{ikm_{ik}})\,]^T$, $(m_{ik} \times q_k)$. Assume that $\text{rank}(F_{ik}) = q_k$ if $m_{ik} > q_k$, as would ordinarily be the case with longitudinal modeling. The residuals from a least-squares fit of this model to all the data on covariate k for subject i may be written as $R_{ik} = P_{ik}W_{ik} = P_{ik}e_{ik}$, where $P_{ik} = I_{m_{ik}} - F_{ik}(F_{ik}^T F_{ik})^{-1}F_{ik}^T$, and $I_{m_{ik}}$ is a m_{ik}-dimensional identity matrix. Suppose covariates k and k' are observed in common at $m_{ikk'} > 0$ time points. Let A_{ik} be the $(m_{ikk'} \times m_{ik})$ matrix of zeros and ones that identifies the residuals for covariate k at the common time points, so that $A_{ik}R_{ik}$ is the $(m_{ikk'} \times 1)$ vector of least-squares residuals at the common time points, and define $A_{ik'}$ similarly. Then the sum of the product of residuals for covariates k and k' over the common times is $R_{ik}^T A_{ik}^T A_{ik'} R_{ik'} = e_{ik}^T P_{ik} A_{ikk'}^* P_{ik'} e_{ik'}$, where $A_{ikk'}^* = A_{ik}^T A_{ik'}$ $(m_{ik} \times m_{ik'})$, which has expectation equal to $\text{tr}\{P_{ik}A_{ikk'}^* P_{ik'} E(e_{ik'}e_{ik}^T)\} = \sigma_{kk'}\text{tr}\{P_{ik}A_{ikk'}^* P_{ik'} A_{ikk'}^{*T}\}$; note that $A_{ikk'}^* = I_{ikk'}(\infty)$. The estimator in (8) follows.

Table 8. *Simulation results for the estimation of* $\omega^T = (\sigma_{11}, \sigma_{12}, \sigma_{22})$

n	α_i	Error	Mean($\hat{\omega}^T \times 10^3$)	SD($\hat{\omega}^T \times 10^3$)
250	normal	1-fold	(10.877, 6.268, 11.121)	(0.012, 0.010, 0.012)
250	bimodal	1-fold	(10.897, 6.285, 11.145)	(0.011, 0.009, 0.011)
250	normal	4-fold	(43.509, 25.071, 44.484)	(0.047, 0.039, 0.047)
250	bimodal	4-fold	(43.587, 25.141, 44.581)	(0.044, 0.037, 0.046)
500	normal	1-fold	(10.891, 6.273, 11.135)	(0.008, 0.007, 0.008)
500	bimodal	1-fold	(10.894, 6.280, 11.146)	(0.008, 0.006, 0.008)
500	normal	4-fold	(43.565, 25.093, 44.541)	(0.033, 0.027, 0.033)
500	bimodal	4-fold	(43.575, 25.120, 44.584)	(0.032, 0.026, 0.033)
500	normal	truncated 4-fold	(42.186, 23.865, 43.035)	(0.031, 0.026, 0.031)
500	bimodal	truncated 4-fold	(42.135, 23.863, 43.060)	(0.030, 0.024, 0.030)
			Nonlinear	
500	normal	4-fold	(43.608, 25.161, 44.597)	(0.034, 0.028, 0.036)
500	bimodal	4-fold	(43.631, 25.132, 44.583)	(0.033, 0.028, 0.035)

For 1-fold normal error, the true value of $\omega^T = (\sigma_{11}, \sigma_{12}, \sigma_{22}) = (10.89, 6.28, 11.14) \times 10^{-3}$. For 4-fold normal error, the true value of $\omega^T = (43.56, 25.12, 44.56) \times 10^{-3}$. For 4-fold normal error truncated at 3 standard deviations, the true value of $\omega^T = (42.13, 23.89, 43.09) \times 10^{-3}$. Mean($\hat{\omega}^T \times 10^3$) and SD($\hat{\omega}^T \times 10^3$) are the Monte Carlo mean and standard deviation of $\hat{\omega}^T \times 10^3$, respectively.

REFERENCES

BYCOTT, P. AND TAYLOR, J. (1998). A comparison of smoothing techniques for CD4 data measured witherror in a time-dependent Cox proportional hazards model. *Statistics in Medicine* **17**, 2061–2077.

CARROLL, R. J., RUPPERT, D. AND STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

COX, D. R. (1972). Regression models and lifetables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–200.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

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

DAVIDIAN, M. AND GALLANT, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80**, 475–488.

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

FAUCETT, C. J. AND THOMAS, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

FLEMING, T. R. AND HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

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). For the AIDS Clinical Trials Group Study 175 Study Team. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* **335**, 1081–1089.

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

HU, P., TSIATIS, A. A. AND DAVIDIAN, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–1419.

PAWITAN, Y. AND SELF, S. (1993). Modeling disease marker processes in AIDS. *Journal of the American Statistical Association* **83**, 719–726.

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

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

RABOUD, J., REID, N., COATES, R. A. AND FAREWELL, V. T. (1993). Estimating risks of progressing to AIDS when covariates are measured with error. *Journal of the Royal Statistical Society, Series A* **156**, 396–406.

STEFANSKI, L. A. AND CARROLL, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika* **74**, 703–716.

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. *Journal of the American Statistical Association* **90**, 27–37.

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.

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

XU, J. AND ZEGER, S. L. (2001a). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

XU, J. AND ZEGER, S. L. (2001b). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81–87.

[*Received May 29, 2001; revised November 2, 2001; accepted for publication November 19, 2001*]