UW Biostatistics Working Paper Series

11-13-2003

A Corrected Pseudo-score Approach for Additive Hazards Model With Longitudinal Covariates Measured With Error

Xiao Song
University of Washington, songx@u.washington.edu

Yijian Huang
Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, yhuang5@sph.emory.edu

Suggested Citation
Song, Xiao and Huang, Yijian, "A Corrected Pseudo-score Approach for Additive Hazards Model With Longitudinal Covariates Measured With Error" (November 2003). *UW Biostatistics Working Paper Series.* Working Paper 217.
http://biostats.bepress.com/uwbiostat/paper217

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors

1. Introduction

In medical studies, information is often collected on several time-dependent and time-independent covariates, along with a possible censored time-to-event. It is of interest to characterize the relationship between the covariates and the time-to-event. However, time-dependent covaritates are usually collected intermittently and with error. An example is the AIDS Clinical Trial Group (ACTG) 175 study, a randomized clinical trial to compare zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine and didanosine alone in treating HIV-infected subjects (Hammer *et al.*, 1996). As a time-dependent covariate, CD4 count is measured about every 12 weeks. It is well known that CD4 measurements are subject to considerable intra-subject variation and measurement error. An objective is to characterize the relationship between progression to AIDS or death and CD4.

To model such data, recent interests focus on the proportional hazards model framework, with longitudinal data jointly modeled using a mixed effects model. Various approaches were proposed for such joint models, including the regression calibration (e.g. Pawitan and Self, 1993; Tsiatis *et al.*, 1995) and likelihood based approaches (e.g. Schluchter, 1992; DeGruttola and Tu, 1994; Wulfsohn and Tsiatis, 1997; Faucett and Thomas, 1996; Henderson *et al.*, 2000; Xu and Zeger, 2001). However, these approaches depend on parametric (normal) assumptions of the random effects and the error. In addition, the regression calibration might give biased estimation (Tsiatis and Davidian, 2001). Some recent approaches were devoted to relaxing the assumptions on the distribution of the random effects. The semiparametric likelihood approach (Song *et al.*, 2002b) assumes the random effects have a "smooth" density and the conditional score approach (Tsiatis and Davidian, 2001; Song *et al.*, 2002a) requires no distributional assumption on the random effects. The corrected score approach (Nakamura, 1992) was originally proposed to deal with error contaminated time-independent covariates, but can be extended to time-dependent covariates directly. However, these approaches are still based on the normality assumption of the errors. In the ACTG data case, our preliminary study indicates that the joint distribution of the random effects is not normal (Song *et al.*, 2002b) and the error might deviate from the normal (Song *et al.*, 2002a). In such cases, approaches based on normality assumption on either the random effects or the error might lead to biased estimation. Furthermore, all these approaches might encounter intractable numerical difficulties in practice. The regression calibration and the likelihood based approaches can be computationally intensive and infeasible to be implemented when multiple time-dependent covariates are measured with error. The conditional score approach might have multiple solutions; in addition, like the corrected score approach, it might break down in the situation of small sample and large measurement error.

The restrictions of these error dealing approaches under the proportional hazards model motivate us to consider an alternative framework, namely the additive hazards model. Like the proportional hazards model, the additive hazards model has sound biological and empirical bases (Breslow and Day 1980, pp. 53–59; 1987, pp. 122–131). When the covariates are not measured with error, Lin and Ying (1994) proposed a pseudo-score approach for the additive hazards model. In the case that time-independent covariates are measured with error, Kulich and Lin (2000) proposed a corrected pseudo-score estimator based on a validation data set. However, little has been done for the additive hazards model with longitudinal covariates measured with error.

In this paper, we propose a simple corrected pseudo-score approach for estimating the regression parameters of the additive hazards model with time-dependent covariates measured with error. In contrast to the approaches under the proportional hazards model, this approach requires only assumptions on the variance structure of the error process and no

Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

assumption at all on the random effects. The resulting estimator has an explicit form and easy to compute. We describe the model in Section 2 and derive the estimator in Section 3. The asymptotic properties are given in Section 4. We illustrate the finite sample performance in Section 5 and apply it to the ACTG 175 data in Section 6.

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 $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Let $\mathbf{X}_i(u) = \{X_{i1}(u), \ldots, X_{iK}(u)\}^T$ denote K underlying time-dependent covariates at time u, and let the $r \times 1$ vector $\mathbf{Z}_i$ denote r time-independent covariates, which are not subject to measurement error. We assume each of the covariate process $X_{ik}(u)$, $k = 1, \ldots, K$, satisfies

$$X_{ik}(u) = \boldsymbol{\alpha}_{ik}^T \mathbf{f}_k(u), \tag{1}$$

where $\mathbf{f}_k(u)$ is a known $(q_k \times 1)$ vector of functions of u, $\boldsymbol{\alpha}_{ik}$ is a $(q_k \times 1)$ random effect, and $\mathbf{f}_k(u)$ and $\boldsymbol{\alpha}_{ik}$ may be different for each k. This allows flexibility in representing the time trajectory of each covariate via polynomial or spline models. 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 independently distributed, mean-zero "errors" with variance σ_k^2 that may reflect both biological variation and measurement error. Thus, (2) along 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 longitudinal measurements are sufficiently separated in time such that the errors are independent across time; however, for the hazard formulation given below, this could be relaxed and other correlation structures could be considered. We allow measurements on different covariates at the same time to be correlated to represent possible correlation between the biological variation and measurement error for different covariates. More formally, we may write $\text{cov}(e_{ikj}, e_{ik'j'}) = \sigma_{kk'} I(t_{ikj} = t_{ik'j'})$, for $k, k' = 1, \ldots, K$, $j = 1, \ldots, m_{ik}$, and $j' = 1, \ldots, m_{ik'}$. This formulation subsumes the case that $\sigma_{kk'} = 0$ for all $k \neq k'$. Let $\mathbf{e}_i = (\mathbf{e}_{i1}^T, \ldots, \mathbf{e}_{iK}^T)^T$, where $\mathbf{e}_{ik} = (e_{ik1}, \ldots, e_{ikm_{ik}})^T$. And $\boldsymbol{\alpha}_i$, $\mathbf{t}_i$, $\mathbf{t}_{ik}$, $\mathbf{W}_i$, $\mathbf{W}_{ik}$ and $\mathbf{m}_i$ are defined similarly. We assume that $(T_i, C_i, \boldsymbol{\alpha}_i, \mathbf{Z}_i, \mathbf{t}_i, \mathbf{m}_i, \mathbf{e}_i)$ are independent and identically distributed across i.

The hazard of failure depends on the random effects and other covariates through the following additive hazards model

$$\begin{aligned} \lambda_i(u) &= lim_{du \to 0} du^{-1} Pr\{u \leq T_i < u + du | T_i \geq u, \boldsymbol{\alpha}_i, C_i, \mathbf{e}_i, \mathbf{t}_i\} \\ &= \lambda_0(u) + \boldsymbol{\gamma}_0^T \mathbf{X}_i(u) + \boldsymbol{\eta}_0^T \mathbf{Z}_i, \end{aligned} \tag{3}$$

where $\lambda_0(u)$ is an unspecified function; $\boldsymbol{\gamma}_0$ and $\boldsymbol{\eta}_0$ are $K \times 1$ and $r \times 1$ vectors of parameters, respectively. This formulation implies that censoring, measurement error and timing of measurements are noninformative. Interest focuses on estimation of $\boldsymbol{\beta}_0 = (\boldsymbol{\gamma}_0^T, \boldsymbol{\eta}_0^T)^T$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

3. Corrected Pseudo-score Estimator

Let τ be a fixed time. When the true covariate processes $\mathbf{X}_i(u)$ were known, we might consider the following estimating function

$$\tilde{\mathbf{U}}(\boldsymbol{\beta};\mathbf{X}) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{S}_i(u)-\overline{\mathbf{S}}(u)\right\}\left[dN_i(u)-Y_i(u)\mathbf{S}_i^T(u)\boldsymbol{\beta}du\right], \tag{4}$$

where $\boldsymbol{\beta} = (\boldsymbol{\gamma}^T,\boldsymbol{\eta}^T)^T$, $\mathbf{S}(u) = \{\mathbf{X}^T(u),\mathbf{Z}^T\}^T$, $\overline{\mathbf{S}}(u) = \frac{\sum_{i=1}^n Y_i(u)\mathbf{S}_i(u)}{\sum_{i=1}^n Y_i(u)}$, $N_i(u) = I(V_i \le u, \Delta_i = 1, m_{ik}(u) \ge q_k, k = 1,\ldots,K)$ is the counting process, $Y_i(u) = I(V_i \ge u, m_{ik}(u) \ge q_k, k = 1,\ldots,K)$ is the at risk process, and $m_{ik}(u)$ is the number of the observations up to time u for the kth covariate. Estimating function (4) is similar to Lin and Ying's pseudo-score function (Lin and Ying, 1994), only with the replacements of $N_i^*(u) = I(V_i \le u, \Delta_i = 1)$ by $N_i(u)$ and $Y_i^*(u) = I(V_i \ge u)$ by $Y_i(u)$. Noting that $N_i(u) = N_i^*(u)I\{m_{ik}(u) \ge q_k, k = 1,\ldots,K\}$ and $Y_i(u) = Y_i^*(u)I\{m_{ik}(u) \ge q_k, k = 1,\ldots,K\}$, it is easy to see that (4) is a weighted pseudo-score function with predictable weights $I\{m_{ik}(u) \ge q_k, k = 1,\ldots,K\}$. The reason to use these weights will become clear below.

Using the empirical mean operator $\hat{\mathcal{E}}(\cdot)$ with $\hat{\mathcal{E}}(\mathbf{a}) = n^{-1}\sum_{i=1}^n \mathbf{a}_i$, (4) can be rewritten as

$$\begin{aligned}\tilde{\mathbf{U}}(\boldsymbol{\beta};\mathbf{X}) \;=\; \int_0^{\tau}\Bigg(&d\hat{\mathcal{E}}\left\{\mathbf{S}(u)N(u)\right\} - \hat{\mathcal{E}}\left\{\frac{d\mathbf{S}(u)}{du}N(u)\right\}du - \hat{\mathcal{E}}\left\{Y(u)\mathbf{S}(u)\mathbf{S}^T(u)\boldsymbol{\beta}\right\}du \\ &-\frac{\hat{\mathcal{E}}\left\{Y(u)\mathbf{S}(u)\right\}}{\hat{\mathcal{E}}\left\{Y(u)\right\}}\left[d\hat{\mathcal{E}}\left\{N(u)\right\} - \hat{\mathcal{E}}\left\{Y(u)\mathbf{S}^T(u)\boldsymbol{\beta}\right\}du\right]\Bigg).\end{aligned} \tag{5}$$

Here $\frac{d\mathbf{S}(u)}{du} = (\frac{d\mathbf{X}^T(u)}{du}, \mathbf{0}_r^T)^T$, and $\mathbf{0}_r$ is $r \times 1$ zero vector. The empirical processes in (5) converge almost surely by the extended strong law of large numbers in Appendix III of Andersen and Gill (1982). Thus, $\tilde{\mathbf{U}}(\boldsymbol{\beta};\mathbf{X})$ converges almost surely to

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta};\mathbf{X}) \;=\; \int_0^{\tau}\Bigg(&d\mathrm{E}\left\{\mathbf{S}(u)N(u)\right\} - \mathrm{E}\left\{\frac{d\mathbf{S}(u)}{du}N(u)\right\}du - \mathrm{E}\left\{Y(u)\mathbf{S}(u)\mathbf{S}^T(u)\boldsymbol{\beta}\right\}du \\ &-\frac{\mathrm{E}\left\{Y(u)\mathbf{S}(u)\right\}}{\mathrm{E}\left\{Y(u)\right\}}\left[d\mathrm{E}\left\{N(u)\right\} - \mathrm{E}\left\{Y(u)\mathbf{S}^T(u)\boldsymbol{\beta}\right\}du\right]\Bigg),\end{aligned}$$

which, by some simple algebra, reduces to

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta};\mathbf{X}) \;=\; \int_0^{\tau}\Bigg(&\mathrm{E}\left[\mathbf{S}(u)\left\{dN(u) - Y(u)\mathbf{S}^T(u)\boldsymbol{\beta}du\right\}\right] \\ &-\frac{\mathrm{E}\left\{Y(u)\mathbf{S}(u)\right\}}{\mathrm{E}\left\{Y(u)\right\}}\left[\mathrm{E}\left\{dN(u) - Y(u)\mathbf{S}^T(u)\boldsymbol{\beta}du\right\}\right]\Bigg).\end{aligned}$$

In practice, the true covariate processes $\mathbf{X}_i(u)$ are unknown. One major technique to deal with covariate measurement error is the corrected score method, see Carroll, Ruppert and Stefanski (Chapter 6, 1995) for a detailed description. The idea of this approach is to remove bias induced by measurement error by correction of the "score" function. To use

COBRA
A BEPRESS
Collection of Biostatistics
Research Archive

this technique, we replace $\mathbf{X}_i(u)$ in (4) by ordinary least square estimates. We consider two ways for computation of the least square estimate. A natural consideration along the line of counting process and martingale theory is to compute the ordinary least square estimates based on observations prior to u, as utilized by Tsiatis and Davidian (2001) for the conditional score approach for the proportional hazards model. This might not be efficient since only part of the longitudinal information are used, which motivate us to consider a more efficient method, that is, to compute the estimates based on all longitudinal observations. The derivation in the sequel utilizes this second method; The derivation using the first one is similar. We compare these two approaches in Section 5.

Thus, let $\hat{X}_{ik}(u)$ be the ordinary least square estimate of $X_{ik}(u)$ based on all the longitudinal data for the ith subject, that is, based on $\mathbf{t}_{ik}$. The estimation of $\hat{X}_{ik}(u)$ is possible only if $m_{ik} \geq q_k$, which explains the usage of the weights $I\{m_{ik}(u) \geq q_k, k = 1, \ldots, K\}$ in (4). Let $\boldsymbol{\xi}_0 = \{\sigma_{kk'} : k \geq k'\}$ denote the distinct parameters characterizing the error covariance. For now, we assume $\boldsymbol{\xi}_0$ is known. Let $\hat{\mathbf{X}}_i(u) = \{\hat{X}_{i1}^T(u), \ldots, \hat{X}_{iK}^T(u)\}^T$. For $k = 1, \ldots, K$, define $\mathbf{F}_{ik} = [\mathbf{f}_k(t_{ik1}), \ldots, \mathbf{f}_k(t_{ikm_{ik}})]^T$, $(m_{ik} \times q_k)$, and let $\mathbf{I}_{ikk'}$ be the $(m_{ik} \times m_{ik'})$ matrix whose (j, j') entry is $I(t_{ikj} = t_{ik'j'})$, for $j = 1, \ldots, m_{ik}$, $j' = 1, \ldots, m_{ik'}$. Let $q = \sum_{k=1}^K q_k$ and

$$\mathbf{F}(u) = \begin{bmatrix} \mathbf{f}_1^T(u) & \mathbf{0}_{q_2} & \ldots & \mathbf{0}_{q_K} \\ \mathbf{0}_{q_1} & \mathbf{f}_2^T(u) & \ldots & \mathbf{0}_{q_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{q_1} & \mathbf{0}_{q_2} & \ldots & \mathbf{f}_K^T(u) \end{bmatrix}, \quad (K \times q).$$

Since $\mathbf{X}_i(u) = \mathbf{F}(u)\boldsymbol{\alpha}_i$, following Song *et al.* (2002a), conditional on $\{\boldsymbol{\alpha}_i, \mathbf{Z}_i, \mathbf{t}_i\}$, the covariance of $\hat{\mathbf{X}}_i(u)$ is equal to $\mathbf{F}(u)\mathbf{G}_i(\boldsymbol{\xi}_0)\mathbf{F}^T(u)$, where

$$\mathbf{G}_i(\boldsymbol{\xi}_0) = \begin{bmatrix} \mathbf{G}_{i11}(\boldsymbol{\xi}_0) & \mathbf{G}_{i12}(\boldsymbol{\xi}_0) & \cdots & \mathbf{G}_{i1K}(\boldsymbol{\xi}_0) \\ \mathbf{G}_{i21}(\boldsymbol{\xi}_0) & \mathbf{G}_{i22}(\boldsymbol{\xi}_0) & \cdots & \mathbf{G}_{i1K}(\boldsymbol{\xi}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{iK1}(\boldsymbol{\xi}_0) & \mathbf{G}_{iK2}(\boldsymbol{\xi}_0) & \cdots & \mathbf{G}_{iKK}(\boldsymbol{\xi}_0) \end{bmatrix}, \quad (q \times q),$$

and $\mathbf{G}_{ikk'}(\boldsymbol{\xi}_0) = \sigma_{kk'}\{\mathbf{F}_{ik}^T\mathbf{F}_{ik}\}^{-1}\mathbf{F}_{ik}^T\mathbf{I}_{ikk'}\mathbf{F}_{ik'}\{\mathbf{F}_{ik'}^T\mathbf{F}_{ik'}\}^{-1}, (q_k \times q_{k'})$.

With similar arguments to those for $\tilde{\mathbf{U}}(\boldsymbol{\beta}; \mathbf{X})$, $\tilde{\mathbf{U}}(\boldsymbol{\beta}; \hat{\mathbf{X}})$ converges to $\mathbf{U}(\boldsymbol{\beta}; \hat{\mathbf{X}})$, which, by some simple algebra, is equal to

$$\mathbf{U}(\boldsymbol{\beta}; \mathbf{X}) - \int_0^\tau \mathrm{E}\left\{Y(u)\mathbf{H}(u, \boldsymbol{\xi}_0)\boldsymbol{\beta}\right\} du,$$

where

$$\mathbf{H}_i(u, \boldsymbol{\xi}_0) = \begin{bmatrix} \{\mathbf{F}(u)\mathbf{G}_i(\boldsymbol{\xi}_0)\mathbf{F}^T(u) & \mathbf{0}_{K\times r} \\ \mathbf{0}_{r\times K} & \mathbf{0}_{r\times r} \end{bmatrix},$$

and $\mathbf{0}_{a\times b}$ is a $a \times b$ zero matrix. Therefore, an unbiased "pseudo-score" can be obtained by adding a consistent estimator of $\int_0^\tau \mathrm{E}\left\{Y(u)\mathbf{H}(u, \boldsymbol{\xi}_0)\boldsymbol{\beta}\right\} du$ to $\mathbf{U}(\boldsymbol{\beta}; \hat{\mathbf{X}})$. Hence we propose the corrected pseudo-score estimating equations

$$\hat{\mathbf{U}}(\boldsymbol{\beta}, \boldsymbol{\xi}_0) = n^{-1} \sum_{i=1}^n \mathbf{Q}_i(\boldsymbol{\beta}, \boldsymbol{\xi}_0) = 0, \tag{6}$$

A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

where

$$\mathbf{Q}_i(\boldsymbol{\beta},\boldsymbol{\xi}) = \int_0^\tau \left(\left\{\hat{\mathbf{S}}_i(u)-\overline{\hat{\mathbf{S}}}(u)\right\}\left[dN_i(u)-Y_i(u)\mathbf{S}_i^T(u)\boldsymbol{\beta}du\right]+Y_i(u)\mathbf{H}_i(u,\boldsymbol{\xi})\boldsymbol{\beta}du\right),$$

$\hat{\mathbf{S}}_i(u) = \left\{\hat{\mathbf{X}}_i^T(u),\mathbf{Z}_i^T\right\}^T$, and $\overline{\hat{\mathbf{S}}}(u) = \frac{\sum_{i=1}^n Y_i(u)\hat{\mathbf{S}}_i(u)}{\sum_{i=1}^n Y_i(u)}$. The resulting estimator takes the explicit form given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n\int_0^\tau Y_i(u)\left[\left\{\hat{\mathbf{S}}_i(u)-\overline{\hat{\mathbf{S}}}(u)\right\}^{\otimes 2}-\mathbf{H}_i(u,\boldsymbol{\xi}_0)\right]du\right)^{-1} \times\left[\sum_{i=1}^n\int_0^\tau\left\{\hat{\mathbf{S}}_i(u)-\overline{\hat{\mathbf{S}}}(u)\right\}dN_i(u)\right], \tag{7}$$

where $\mathbf{a}^{\otimes 2}=\mathbf{a}\mathbf{a}^T$.

In practice, the vector of error variance $\boldsymbol{\xi}_0$ is usually unknown. Following Song *et al.* (2002a), an unbiased estimator of $\boldsymbol{\xi}_0$ is $\hat{\boldsymbol{\xi}}$, with element $\sigma_{kk'}$ estimated by

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

where $m_{ikk'}$ is the number of common time points that covariates k and k' are observed, $\mathbf{R}_{ik}=\mathbf{P}_{ik}\mathbf{W}_{ik}$, $\mathbf{W}_{ik}=\{W_{ik}(t_{ik1}),\ldots,W_{ik}(t_{ikm_{ik}})\}^T$, $\mathbf{P}_{ik}=\mathbf{I}_{m_{ik}}-\mathbf{F}_{ik}(\mathbf{F}_{ik}^T\mathbf{F}_{ik})^{-1}\mathbf{F}_{ik}^T$, $\mathbf{I}_{m_{ik}}$ is a m_{ik}-dimensional identity matrix.

4. Large Sample Properties

From (6) and (8), the combined estimating equations for $(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)$ are

$$\boldsymbol{\Phi}(\boldsymbol{\beta},\boldsymbol{\xi}) = n^{-1}\sum_{i=1}^n\boldsymbol{\varphi}_i(\boldsymbol{\beta},\boldsymbol{\xi}) = 0, \tag{9}$$

where $\boldsymbol{\varphi}_i(\boldsymbol{\beta},\boldsymbol{\xi}) = \left\{\mathbf{Q}_i^T(\boldsymbol{\beta},\boldsymbol{\xi}),\mathbf{h}_i^T(\boldsymbol{\xi})\right\}^T$, $\mathbf{h}_i(\boldsymbol{\xi})$ is a vector with the elements $I(m_{ik}>q_k,m_{ik'}>q_{k'},m_{ikk'}>0)[\mathbf{R}_{ik}^T\mathbf{I}_{ikk'}\mathbf{R}_{ik'}-\sigma_{kk'}\mathrm{tr}\{\mathbf{P}_{ik}\mathbf{I}_{ikk'}\mathbf{P}_{ik'}\mathbf{I}_{ikk'}^T\}]$, $k\geq k'$.

Let $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\xi}})$ be the solution to (9). Let $\|\mathbf{a}\|$ denote the norm of vector $\mathbf{a}$. To derive the asymptotic properties of $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\xi}})$, we make the following assumptions:

A. $\Pr\{Y_i(\tau)=1\}>0$;

B. For $k,k'=1,\ldots,K$,

$$\sup_{u\in[0,\tau]}\|\mathbf{f}_k(u)\|<\infty,\qquad \sup_{u\in[0,\tau]}\|d\mathbf{f}_k(u)/du\|<\infty,$$

$$\mathrm{E}\left\|\boldsymbol{\alpha}_{ik}\right\|^4<\infty,\qquad \mathrm{E}\left\|\{\mathbf{F}_{ik}^T\mathbf{F}_{ik}\}^{-1}\mathbf{F}_{ik}^T\mathbf{e}_{ik}\right\|^4<\infty,\qquad \mathrm{E}\left\|\mathbf{Z}_i\right\|^4<\infty,$$

$$\mathrm{E}\left[\sup_{u\in[0,\tau]}\left|\mathbf{f}_k^T(u)\{\mathbf{F}_{ik}^T\mathbf{F}_{ik}\}^{-1}\mathbf{F}_{ik}^T\mathbf{I}_{ikk'}\mathbf{F}_{ik'}\{\mathbf{F}_{ik'}^T\mathbf{F}_{ik'}\}^{-1}\mathbf{f}_{k'}(u)\right|^2\right]<\infty,$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

C. $\mathbf{\Gamma}_0 = \mathrm{E}\left\{-\frac{\partial \mathbf{U}(\boldsymbol{\beta};\mathbf{X})}{\partial \boldsymbol{\beta}^T}\right\} = \int_0^\tau \mathrm{E}\left(\left[Y(u)\mathbf{S}^{\otimes 2}(u)\right] - \frac{\left[\mathrm{E}\{Y(u)\mathbf{S}(u)\}\right]^{\otimes 2}}{\mathrm{E}\{Y(u)\}}\right) du$ is positive definite;

D. $\mathrm{E}\left|\mathrm{tr}\{\mathbf{P}_{ik}\mathbf{I}_{ikk'}\mathbf{P}_{ik'}\mathbf{I}_{ikk'}^T\}\right| < \infty$, $\mathrm{E}\left[I(m_{ik} > q_k, m_{ik'} > q_{k'}, m_{ikk'} > 0)\mathrm{tr}\{\mathbf{P}_{ik}\mathbf{I}_{ikk'}\mathbf{P}_{ik'}\mathbf{I}_{ikk'}^T\}\right] \neq 0$, $\mathrm{E}\left\{\left|\mathbf{e}_{ik}^T\mathbf{P}_{ik}\mathbf{I}_{ikk'}\mathbf{P}_{ik'}\mathbf{e}_{ik'}\right|^2\right\} < \infty$, $k, k' = 1, \ldots, K$.

Under these assumptions, we can show that $\left(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\xi}}^T\right)^T$ is consistent and asymptotically normal with variance consistently estimated by the sandwich estimator $n^{-1}\hat{\mathbf{V}} = n^{-1}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\left\{\hat{\mathbf{A}}^{-1}\right\}^T$, where $\hat{\mathbf{A}} = n^{-1}\partial\mathbf{\Phi}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}})/\partial(\boldsymbol{\beta}^T, \boldsymbol{\xi}^T)$, $\hat{\mathbf{B}} = n^{-1}\sum_{i=1}^n \boldsymbol{\varphi}_i^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}})\boldsymbol{\varphi}_i^{*T}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}})$, $\boldsymbol{\varphi}_i^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) = (\hat{\boldsymbol{\omega}}_i^T(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}), \mathbf{h}_i^T(\hat{\boldsymbol{\xi}}))^T$,

$$\begin{aligned}
\hat{\boldsymbol{\omega}}_i(\boldsymbol{\beta}, \boldsymbol{\xi}) \quad = \quad & \int_0^\tau \Bigg(\left[\hat{\mathbf{S}}_i(u) - \frac{\hat{\mathcal{E}}\left\{Y(u)\hat{\mathbf{S}}(u)\right\}}{\hat{\mathcal{E}}\left\{Y(u)\right\}}\right]\left\{dN_i(u) - Y_i(u)\hat{\mathbf{S}}_i^T(u)\boldsymbol{\beta}\right\} \\
& - \left[\frac{Y_i(u)\hat{\mathbf{S}}_i(u)}{\hat{\mathcal{E}}\left\{Y(u)\right\}} - \frac{Y_i(u)\hat{\mathcal{E}}\left\{Y(u)\hat{\mathbf{S}}(u)\right\}}{\hat{\mathcal{E}}^2\left\{Y(u)\right\}}\right]\hat{\mathcal{E}}\left[\{N(u)\} - Y(u)\hat{\mathbf{S}}^T(u)\boldsymbol{\beta}\right] \\
& + Y_i(u)\mathbf{H}_i(u, \boldsymbol{\xi})\boldsymbol{\beta}\Bigg) du.
\end{aligned}$$

The proof is sketched in the Appendix.

5. Simulation Studies

To evaluate the performance of this approach, we conducted simulations under the following scenarios, which mimicked typical AIDS clinical trials. In all cases, for each of $n = 500$ or 1000 subjects, the longitudinal observations W_{ij} of a time-dependent covariate $X_i(u) = \alpha_{i0} + \alpha_{i1}u$ were generated at times $t_{ij} = 0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80$ weeks according to the linear mixed effects model (2), with a 10% missingness rate at any time after week 16. The random effects followed a truncated normal distribution taking only negative samples for the slope; the underlying normal has mean $(2.603, -0.0023)$ and distinct covariance matrix elements $\{\mathrm{var}(\alpha_{i0}), \mathrm{cov}(\alpha_{i0}, \alpha_{i1}), \mathrm{var}(\alpha_{i1})\} = (0.146, 0.00265, 0.00324)$. The errors were generated from either a normal or a skewed bimodal mixture of normal as described in Davidian and Gallant (1993) with mixing proportion $p = 0.3$ and sep=3. For both, the mean is zero and variance $\sigma^2 = 0.20$ or 0.60. A treatment indicator Z_i with $\mathbf{P}(Z_i = 1) = 0.5$ was generated. The hazard relationship was taken to be $\lambda_i(u) = \lambda_0(u) + \gamma X_i(u) + \eta Z_i$ with $\gamma = -0.01$, $\eta = 0$ and $\lambda_0 = 0.05$. Censoring was generated from an exponential distribution with mean 110 weeks and truncated at the maximum follow-up time 80 weeks, leading to a censoring rate of 24%.

For each scenario, 1000 Monte Carlo data sets were simulated. For each data set, we fitted the model above four ways: (i) using Lin and Ying's estimator, called the 'ideal' estimator, in which the true $X_i(u)$ is used; (ii) using naive regression where the true values of $X_i(u)$ in Lin and Ying's estimator were substituted by the ordinary least square estimates; (iii) using the corrected pseudo-score estimator with $\hat{X}_i(u)$ computed based the observations prior to u. (iv) using the corrected pseudo-score estimator given in section 3. To

Collection of Biostatistics
Research Archive

discriminate (iii) from (iv), we called (iii) type I corrected pseudo-score estimator and (iv) type II corrected pseudo-score estimator, respectively. In all cases, 95% Wald confidence intervals for γ and η were constructed. Results are given in Table 1. In all cases and for both sample sizes, the corrected pesudo-score estimators show negligible bias close to that of the unachievable 'ideal' estimator, and their coverage probabilities are close to the nominal level. In contrast, the naive regression approach can yield biased estimates and coverage probabilities well below the nominal level. In addition, the type II corrected pesudo-score estimator is more efficient than the type I estimator, as expected.

The simulation evidence suggests that the corrected pseudo-score estimators yield reliable inferences regardless of the distribution of error under moderate sample sizes.

6. Application to ACTG 175

We apply the approach to the ACTG 175 data. We are interested in assessing the effect of CD4 counts on time to AIDS or death in antiretroviral-naive patients. In the study, 1,067 patients had no antiretroviral theropy at baseline. Figure 1 presents $\log_{10}$-transformed CD4 profiles for 10 randomly selected subjects and shows an apparent initial increase, with a peak at week 12, followed by a linear decline. Only 1 event occurred before week 12. For simplicity, we consider the post-twelve-week data and assume $X_i(u) = \alpha_{i0} + \alpha_{i1}u$ represents "inherent" $\log_{10}$ CD4 count for subject i at time u and model (2) with error variance σ^2. The primary analysis found zidovudine alone to be inferior to the other three therapies; thus, further investigations focused on two treatment groups, zidovudine alone and the combination of the other three. We took the hazard for AIDS or death to be as in (3) with $Z_i = I($ treatment $\neq$ zidovudine$)$. We estimated the regression parameters using the naive and corrected score approaches.

Results are shown in table 2. The naive regression estimate of γ is smaller in magnitude. The corrected pseudo-score estimates are close with smaller standard errors for the type II estimates. The results suggested that there is a strong relationship between hazard and log(CD4); moreover, once this is taken into account, there appears to be no association between prognosis and treatment, that is, the treatment effect is mediated through log(CD4). The estimated error variance is 0.010.

7. Discussion

We have proposed a semiparametric approach for a joint model for survival and longitudinal data with assumptions on only the variance structure of the error process and no distributional assumption on the random effects. The estimator has a closed form formula and is simple to compute. It uses efficiently the longitudinal information by utilizing the least square estimates based on all longitudinal observations. The same idea could be applied to the corrected score approach for the proportional hazards model.

The model as presented here may be extended to more complicated situations. For example, it is possible to consider models with trajectories nonlinear in individual random effects, and models with more complicated error structure. Moreover, the model may be extended to involve nonlinear functions of the random effects in the hazard using linear approximation by analogy to Song *et al.* (2002a). In such cases, the estimator might no longer have an explicit form.

Collection of Biostatistics
Research Archive

Acknowledgements

This research was supported by NIH grants CA90747 and AI29168 (Huang) and U01-AI46702 (Song), and career development funds from the University of Washington Department of Biostatistics (Song).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper217

Appendix A. Large sample properties of $\hat{\beta}$

For now, we assume the vector of error variances $\boldsymbol{\xi}$ is known. Then $\hat{\boldsymbol{\beta}}$ satisfies $\hat{\mathbf{U}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\xi}_0) = 0$. First, we show the consistency of $\hat{\boldsymbol{\beta}}$. Let $\mathbf{B}$ be a compact set including $\boldsymbol{\beta}_0$ as a internal point. Since $\hat{\mathbf{U}}(\boldsymbol{\beta}, \boldsymbol{\xi}) = \tilde{\mathbf{U}}(\boldsymbol{\beta}; \hat{\mathbf{X}}) + n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \mathbf{H}_i(u, \boldsymbol{\xi}) \boldsymbol{\beta} du$, using the empirical mean operator $\hat{\mathcal{E}}$, at $\boldsymbol{\xi} = \boldsymbol{\xi}_0$, (6) can be rewritten as

$$\begin{aligned}
\hat{\mathbf{U}}(\boldsymbol{\beta}, \boldsymbol{\xi}_0) \quad = \quad & \int_0^{\tau} \Bigg(d\hat{\mathcal{E}} \left\{ \hat{\mathbf{S}}(u) N(u) \right\} - \hat{\mathcal{E}} \left\{ \frac{d\hat{\mathbf{S}}(u)}{du} N(u) \right\} du - \hat{\mathcal{E}} \left\{ Y(u) \hat{\mathbf{S}}(u) \hat{\mathbf{S}}^T(u) \boldsymbol{\beta} \right\} du \\
& - \frac{\hat{\mathcal{E}} \left\{ Y(u) \hat{\mathbf{S}}(u) \right\}}{\hat{\mathcal{E}} \left\{ Y(u) \right\}} \left[d\hat{\mathcal{E}} \left\{ N(u) \right\} - \hat{\mathcal{E}} \left\{ Y(u) \hat{\mathbf{S}}^T(u) \boldsymbol{\beta} \right\} du \right] \\
& + \hat{\mathcal{E}} \{ Y(u) \mathbf{H}(u, \boldsymbol{\xi}_0) \boldsymbol{\beta} \} du \Bigg), \qquad (10)
\end{aligned}$$

Under condition B, it is easy to show that

$$\mathrm{E}\left[\sup_{u \in [0,\tau]} \mathbf{S}^T(u) \mathbf{S}(u) \right] < \infty, \qquad \mathrm{E}\left[\sup_{u \in [0,\tau]} \frac{d\mathbf{X}^T(u)}{du} \frac{d\mathbf{X}(u)}{du} \right] < \infty,$$

$$\mathrm{E}\left[\sup_{\substack{u \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} \left\{ \mathbf{S}^T(u) \boldsymbol{\beta} \right\}^2 \right] < \infty, \qquad \mathrm{E}\left[\sup_{\substack{u \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} \left\{ \mathbf{S}^T(u) \boldsymbol{\beta} \right\}^2 \mathbf{S}^T(u) \mathbf{S}(u) \right] < \infty,$$

$$\mathrm{E}\left[\sup_{\substack{u \in [0,\tau] \\ \boldsymbol{\beta} \in \mathcal{B}}} \left\{ \boldsymbol{\beta}^T H^T(u, \boldsymbol{\xi}_0) \mathbf{H}(u, \boldsymbol{\xi}_0) \boldsymbol{\beta} \right\} \right] < \infty.$$

And condition A implies that $E\{Y(u)\}$ is bounded away from zero for $u \in [0, \tau]$. Hence, by using the extended law of large numbers in Appendix III of Anderson and Gill (1982), the empirical processes in (10) converge almost surely to their limits uniformly in $u \in [0, \tau]$ and $\boldsymbol{\beta} \in \mathcal{B}$ and hence $\hat{\mathbf{U}}(\boldsymbol{\beta}, \boldsymbol{\xi}_0)$ converges uniformly to $\mathbf{U}(\boldsymbol{\beta}; \mathbf{X})$ in $\boldsymbol{\beta} \in \mathcal{B}$. Since $\hat{\mathbf{U}}(\boldsymbol{\beta}, \boldsymbol{\xi}_0)$ and $\mathbf{U}(\boldsymbol{\beta}; \mathbf{X})$ are linear functions of $\boldsymbol{\beta}$ and $\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{X}) = 0$, the consistency of $\hat{\boldsymbol{\beta}}$ follows.

Next, we show that $n^{1/2}\{\hat{\mathbf{U}}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0) - \mathbf{U}(\boldsymbol{\beta}_0; \mathbf{X})\}$ converges to a normal distribution. Let $\boldsymbol{\Lambda}(u, \boldsymbol{\beta}, \boldsymbol{\xi}_0)$ denote the vector composed of the empirical processes in (10). Following the proof of Lemma 5.1 in Tsiatis (1981), under condition B, $n^{1/2}[\boldsymbol{\Lambda}(u, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0) - \mathrm{E}\{\boldsymbol{\Lambda}(u, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0)\}]$ converges to a Gaussian process. Coupled with $\hat{\mathbf{U}}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0)$ being Hadamard differentiable as a functional of $\boldsymbol{\Lambda}$, the asymptotic normality of $n^{1/2}\hat{\mathbf{U}}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0)$ follows by the functional delta method. Using the functional Taylor expansion, with some algebra, we can show that

$$n^{1/2}\hat{\mathbf{U}}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\omega}_i(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0) + o_p(1), \qquad (11)$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

where

$$\begin{aligned}\boldsymbol{\omega}_i(\boldsymbol{\beta},\boldsymbol{\xi}) &= \int_0^\tau \left(\left[\hat{\mathbf{S}}_i(u) - \frac{\mathrm{E}\left\{Y(u)\hat{\mathbf{S}}(u)\right\}}{\mathrm{E}\left\{Y(u)\right\}} \right] \left\{ dN_i(u) - Y_i(u)\hat{\mathbf{S}}_i^T(u)\boldsymbol{\beta} \right\} \right. \\ &\quad - \left[\frac{Y_i(u)\hat{\mathbf{S}}_i(u)}{\mathrm{E}\left\{Y(u)\right\}} - \frac{Y_i(u)\mathrm{E}\left\{Y(u)\hat{\mathbf{S}}(u)\right\}}{\mathrm{E}^2\left\{Y(u)\right\}} \right] \mathrm{E}\left[\{N(u)\} - Y(u)\hat{\mathbf{S}}^T(u)\boldsymbol{\beta} \right] \\ &\quad \left. + Y_i(u)\mathbf{H}_i(u,\boldsymbol{\xi})\boldsymbol{\beta} \right) du.\end{aligned}$$

By the central limit theorem, $n^{1/2}\left\{\hat{\mathbf{U}}(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0) - \mathbf{U}(\boldsymbol{\beta}_0;\mathbf{X})\right\}$ converges to a normal distribution with mean 0 and variance $\Omega(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0) = \mathrm{var}\{\boldsymbol{\omega}_i(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)\}$. By empirical processes theory, we can show that $\hat{\Omega}(\boldsymbol{\beta},\boldsymbol{\xi}_0) = n^{-1}\sum_{i=1}^n \left\{\hat{\boldsymbol{\omega}}_i(\boldsymbol{\beta},\boldsymbol{\xi}_0) - \overline{\hat{\boldsymbol{\omega}}}(\boldsymbol{\beta},\boldsymbol{\xi}_0)\right\}^{\otimes 2}$ converges almost surely to $\Omega(\boldsymbol{\beta},\boldsymbol{\xi}_0)$ uniformly in $\boldsymbol{\beta}\in\mathcal{B}$, where $\overline{\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}) = \hat{\mathcal{E}}\{\hat{\boldsymbol{\omega}}(\hat{\boldsymbol{\beta}},\boldsymbol{\xi})\}$. With simple algebra, we can show $\hat{\Omega}(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}_0) = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\omega}}_i(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}_0)\hat{\boldsymbol{\omega}}_i^T(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}_0)$.

Since $\hat{\mathbf{U}}(\boldsymbol{\beta},\boldsymbol{\xi})$ is a linear function of $\boldsymbol{\beta}$,

$$0 = n^{1/2}\hat{\mathbf{U}}(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}_0) = n^{1/2}\hat{\mathbf{U}}(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0) - \hat{\boldsymbol{\Gamma}}(\boldsymbol{\xi}_0)n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}(\boldsymbol{\xi}) &= -\frac{\partial\hat{\mathbf{U}}(\boldsymbol{\beta},\boldsymbol{\xi})}{\partial\boldsymbol{\beta}} = n^{-1}\sum_{i=1}^n \int_0^\tau \left(Y_i(u)\left[\left\{\hat{\mathbf{S}}_i(u) - \overline{\hat{\mathbf{S}}}(u)\right\}^{\otimes 2} - \mathbf{H}_i(u,\boldsymbol{\xi}) \right] \right) du \\ &= \int_0^\tau \left[\hat{\mathcal{E}}\left\{Y(u)\hat{\mathbf{S}}(u)\hat{\mathbf{S}}^T(u)\right\} - \frac{\hat{\mathcal{E}}\left\{Y(u)\hat{\mathbf{S}}(u)\right\}\hat{\mathcal{E}}\left\{Y(u)\hat{\mathbf{S}}^T(u)\right\}}{\hat{\mathcal{E}}\left\{Y(u)\right\}} - \hat{\mathcal{E}}\left\{Y(u)\mathbf{H}(u,\boldsymbol{\xi})\right\} \right] du.\end{aligned}$$

By the extended strong law of large numbers $\hat{\boldsymbol{\Gamma}}(\boldsymbol{\xi})$ converges almost surely to

$$\begin{aligned}\boldsymbol{\Gamma}(\boldsymbol{\xi}) &= \int_0^\tau \left[E\left\{Y(u)\hat{\mathbf{S}}(u)\hat{\mathbf{S}}^T(u)\right\} - \frac{E\left\{Y(u)\hat{\mathbf{S}}(u)\right\}E\left\{Y(u)\hat{\mathbf{S}}^T(u)\right\}}{E\left\{Y(u)\right\}} \right. \\ &\quad \left. - E\left\{Y(u)\mathbf{H}(u,\boldsymbol{\xi})\right\} \right] du\end{aligned}$$

uniformly in a neighborhood of $\boldsymbol{\xi}_0$. It is easy to show that $\boldsymbol{\Gamma}(\boldsymbol{\xi}_0) = \boldsymbol{\Gamma}_0$. Thus, under condition C, the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ follows from that of $n^{1/2}\hat{\mathbf{U}}(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)$ with asymptotic variance $\Sigma(\boldsymbol{\beta}_0) = \boldsymbol{\Gamma}_0^{-1}\Omega(\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)\left\{\boldsymbol{\Gamma}_0^{-1}\right\}^T$. An estimator for $\Sigma(\boldsymbol{\beta}_0)$ is $\hat{\Sigma}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Gamma}}(\boldsymbol{\xi}_0)^{-1}\hat{\Omega}(\hat{\boldsymbol{\beta}},\boldsymbol{\xi}_0)\left\{\hat{\boldsymbol{\Gamma}}(\boldsymbol{\xi}_0)^{-1}\right\}^T$. The consistency of $\Sigma(\boldsymbol{\beta}_0)$ then follows from the almost surely convergence of $\hat{\Omega}(\boldsymbol{\beta},\boldsymbol{\xi}_0)$ uniformly in $\boldsymbol{\beta}\in\mathcal{B}$.

When $\boldsymbol{\xi}$ is unknown, following Song et al.(2002a), $\hat{\boldsymbol{\xi}}$ is a consistent estimator of $\boldsymbol{\xi}_0$ under condition D. Now $\hat{\boldsymbol{\beta}}$ is obtained by replacing $\boldsymbol{\xi}$ with $\hat{\boldsymbol{\xi}}$ in (7), the denominator of which is equal to $n\hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\xi}})$. It has been shown that $\hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\xi}})$ converges almost surely to $\boldsymbol{\Gamma}(\boldsymbol{\xi})$ uniformly in a

Collection of Biostatistics
Research Archive

neighborhood of $\boldsymbol{\xi}_0$. This, together with the continuity of $\boldsymbol{\Gamma}(\boldsymbol{\xi})$ and consistency of $\hat{\boldsymbol{\xi}}$, implies that $\hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\xi}})$ converges to $\boldsymbol{\Gamma}_0$. The consistency of $\hat{\boldsymbol{\beta}}$ then follows. Using similar arguments as those when $\boldsymbol{\xi}$ is known, we can show that $n^{1/2}\left\{(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\xi}}^T)^T - (\boldsymbol{\beta}_0^T, \boldsymbol{\xi}_0^T)^T\right\}$ converges to a normal distribution with variance $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}\left\{\mathbf{A}^{-1}\right\}^T$, where $\mathbf{A} = \mathrm{E}\left\{\partial\boldsymbol{\Phi}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0)/\partial(\boldsymbol{\beta}^T, \boldsymbol{\xi}^T)\right\}$ and $\mathbf{B} = \mathrm{var}\{\boldsymbol{\varphi}^*(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0)\}$, $\boldsymbol{\varphi}_i^*(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0) = \{\boldsymbol{\omega}_i^T(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0), \mathbf{h}_i^T(\boldsymbol{\beta}_0, \boldsymbol{\xi}_0)\}^T$, and $\mathbf{V}$ is consistently estimated by $\hat{\mathbf{V}}$.

References

Knuth, D. E. (1986). *The TEXbook*. Reading, MA: Addison-Wesley.

Andersen, P.K. and Gill, R.D. (1982) Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann. Statist.*, **10**, 1100–1120.

Carrol, R.J., Ruppert, D. and Stefanski, L.A. (1995) *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

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. *Statist. Med.*, **15**, 1663–1685.

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., for the AIDS Clinical Trials Group Study 175 Study Team, (1996) A Trial Comparing Nucleoside Monotherapy With Combination Therapy in HIV-infected Adults With CD4 Cell Counts From 200 to 500 Per Cubic Millimeter. *New Engl. J. Med.*, **335**, 1081–1089.

Henderson, R., Diggle, P. and Dobson, A. (2000) Joint Modeling of Longitudinal Measurements and Event Time Data. *Biostatistics*, **4**, 465–480.

Kulich, M. and Lin, D.Y. (2000) Additive Hazards Regression With Covariate Measurement Error. *J. Am. Statist. Ass.*, **95**, 238–248.

Lin, D.Y. and Yin, Z. (1994) Semiparametric Analysis of the Addivive Risk model. *Biometrika*, **81**, 61–71.

Nakamura, T. (1992) Proportional hazards model with covariates subject to measurement error. *Biometrics*, **48**, 829-838.

Pawitan, Y. and Self, S. (1993) Modeling Disease Marker Processes in AIDS. *J. Am. Statist. Ass.*, **83**, 719–726.

Prentice, R. (1982) Covariate Measurement Errors and Parameter Estimates in a Failure Time Regression Model. *Biometrika*, **69**, 331–342.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Schluchter, M.D. (1992) Methods for the Analysis of Informatively Censored Longitudinal Data. *Statist. Med.*, **11**, 1861–1870.

Song, X., Davidian, M. and Tsiatis, A. A. (2002a) An Estimator for the Proportional Hazards Model With Multiple Longitudinal Covariates Measured With Error. *Biostatistics*, **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.

Tsiatis, A.A. (1981) A Large Sample Study of Cox's Regression Model. *Ann. Statist.*, **9**, 93–108.

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.Am. Statist. Ass.*, **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. (2001) Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events. *Appl. Statist.*, **50**, 375–387.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 1. Simulation results †

		$n = 500$							$n = 1000$						
			$\sigma^2 = 0.2$			$\sigma^2 = 0.4$				$\sigma^2 = 0.2$			$\sigma^2 = 0.4$		
		I	NR	CSI	CSII	NR	CSI	CSII	I	NR	CSI	CSII	NR	CSI	CSII
								Normal Error							
γ	B	−0.26	0.94	−0.26	−0.24	1.76	−0.30	−0.24	−0.11	1.01	−0.14	−0.13	1.80	−0.17	−0.14
	SD	2.27	2.18	2.66	2.54	2.13	2.99	2.78	1.59	1.53	1.85	1.79	1.49	2.07	1.94
	SE	2.06	2.14	2.50	2.46	2.18	2.85	2.77	1.46	1.51	1.74	1.72	1.54	1.98	1.94
	CP	92.5	91.8	92.8	94.0	86.6	93.8	94.8	92.5	88.5	93.4	94.4	77.2	94.4	95.3
η	B	−0.13	−0.17	−0.19	−0.19	−0.17	−0.18	−0.20	−0.06	−0.06	−0.06	−0.06	−0.06	−0.07	−0.06
	SD	3.47	3.36	3.71	3.69	3.35	3.75	3.70	2.38	2.31	2.55	2.53	2.30	2.58	2.54
	SE	3.39	3.27	3.63	3.61	3.26	3.66	3.62	2.38	2.30	2.54	2.53	2.30	2.56	2.54
	CP	94.5	93.6	93.7	93.8	93.7	93.9	93.9	95.4	94.7	94.4	94.8	94.4	94.3	94.7
σ^2	B	—	—	−0.25	−0.25	—	−0.51	−0.51	—	—	−0.17	−0.17	—	−0.33	−0.33
	SD	—	—	6.87	6.87	—	13.73	13.73	—	—	5.00	5.00	—	10.01	10.01
	SE	—	—	6.97	6.97	—	13.95	13.95	—	—	4.93	4.93	—	9.85	9.85
	CP	—	—	95.2	95.2	—	95.2	95.2	—	—	94.1	94.1	—	94.1	94.1
								Skewed Bimodal Error							
γ	B	−0.14	0.97	−0.21	−0.20	1.77	−0.27	−0.24	−0.09	1.04	−0.07	−0.09	1.84	−0.07	−0.09
	SD	2.25	2.12	2.59	2.49	2.07	2.92	2.72	1.62	1.54	1.87	1.80	1.49	2.08	1.95
	SE	2.05	2.13	2.49	2.46	2.19	2.84	2.78	1.46	1.51	1.74	1.73	1.54	1.98	1.95
	CP	92.8	91.9	94.7	95.2	88.0	95.3	96.2	92.5	86.8	93.8	94.4	75.8	93.7	95.2
η	B	−0.00	−0.04	−0.04	−0.03	−0.04	−0.03	−0.03	0.13	0.11	0.13	0.11	0.10	0.14	0.11
	SD	3.34	3.24	3.55	3.54	3.22	3.58	3.55	2.39	2.32	2.56	2.54	2.32	2.58	2.55
	SE	3.38	3.26	3.62	3.60	3.25	3.66	3.62	2.38	2.30	2.54	2.53	2.29	2.56	2.54
	CP	95.6	95.0	95.0	95.0	94.9	95.4	95.2	94.1	94.6	94.4	94.4	94.6	94.1	94.2
σ^2	B	—	—	−0.01	−0.01	—	−0.02	−0.02	—	—	−0.09	−0.09	—	−0.17	−0.17
	SD	—	—	6.26	6.26	—	12.51	12.51	—	—	4.51	4.51	—	9.02	9.02
	SE	—	—	6.28	6.28	—	12.57	12.57	—	—	4.44	4.44	—	8.88	8.88
	CP	—	—	95.5	95.5	—	95.5	95.5	—	—	94.1	94.1	—	94.1	94.1

† I, 'ideal' method; NR, naive regression; CSI, type I corrected pseudo-score; CSII, type II corrected pseudo-score; B, bias (×1000); SD, empirical standard deviation across simulated data sets (×1000); SE, average of estimated standard errors (×1000); CP, coverage probability (%) of 95% Wald confidence interval.

Hosted by The Berkeley Electronic Press

Table 2. Results for ACTG 175 data †

	γ		η		σ^2	
	Est	SE	Est	SE	Est	SE
NR	-0.00611	0.00088	-0.00007	0.00021	—	—
CSI	-0.00981	0.00237	0.00017	0.00033	0.01041	0.00077
CSII	-0.00936	0.00214	0.00010	0.00031	0.01041	0.00077

† NR, naive regression; CSI, type I corrected pseudo-score; CSII, type II corrected pseudo-score.

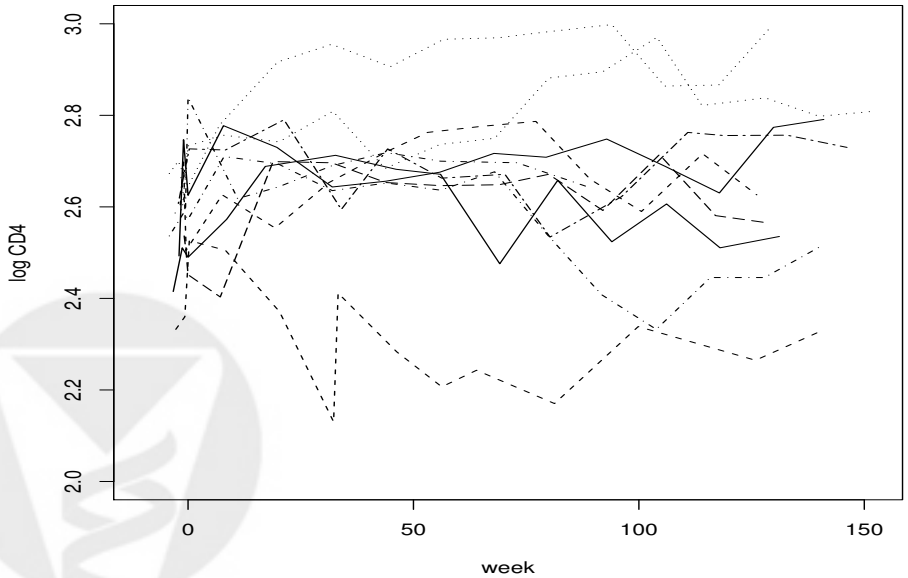

Fig. 1. Trajectories of log CD4 for 10 randomly selected subjects.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive