Statistica Sinica **27** (2017), 1037-1057
doi:https://doi.org/10.5705/ss.202015.0354

AN IMPROVED CORRECTED SCORE ESTIMATOR FOR THE PROPORTIONAL HAZARDS MODEL WITH TIME-DEPENDENT COVARIATES MEASURED WITH ERROR AT INFORMATIVE OBSERVATION TIMES

Xiao Song

University of Georgia

Abstract: We consider the proportional hazards model with time-dependent covariates measured with error at informative observation times under shared random effects models. Although various approaches have been proposed to deal with measurement error for time-dependent covariates, very limited research has been done when the observation times are informative. We propose a new corrected score estimator that allows the observation times to depend on the survival time, the random effects, or other covariates. Compared to existing conditional score and corrected score approaches, it relaxes the requirement on non-informative observation times, may substantially improve the efficiency, and is much more robust to deviations from normality of the error. The performance of the estimator is evaluated via simulation studies and by application to data from an HIV clinical trial.

Key words and phrases: Conditional score, corrected score, informative observation times, joint modeling, measurement error.

1. Introduction

In biomedical studies, information is often collected on a time-to-event (e.g., survival time or failure time) and some covariates. It is of interest to characterize the relationship between the time-to-event and the covariates using a survival model such as the proportional hazards model. Standard inference requires knowing the values of the covariates at each event time. However, these may not be available for time-dependent covariates as they are usually measured intermittently. In addition, the observations may be subject to error. A further complication is that the observation times may be *informative* in that they are related to the previously observed covariates or even the survival time. For example, patients with more severe disease status may appear more often for hospital check-ups in observational studies or have more missing observations during follow-ups in clinical trials.

An example is AIDS Clinical Trial Group (ACTG) 175, a randomized clinical trial to compare four antiretroviral therapies in HIV-infected subjects (Hammer et al. (1996)). During the study, 2,467 subjects were recruited between December 1991 and October 1992 and followed until November 1994. CD4 count, as a reflection of immune status, was scheduled to be measured for each participant about every 12 weeks after randomization. Observations of CD4 count are subject to substantial biological variation and measurement error. An objective was to assess the effect of treatment and CD4 count on the time to AIDS or death. Some patients had missing CD4 measurements during the follow-up before the event or censoring time, with an average missing rate of 18% after week 12. We focused on the data after week 12 for reasons explained in Section 6. The missing rate seems to be significantly associated with the last observed log-10 transformed CD4 count based on the logistic GEE model (p-value $< 2e^{-16}$) — if the last observed CD4 count increased by 10, the odds of missing decreased by a factor of 0.385. We also assessed whether missingness depends on the estimated intercept and slope of the log-10 transformed CD4 count trajectory via the logistic GEE model. Both were significant with p-values less than e^{-10}. This indicates that the observation times are informative.

To circumvent the complication of the intermittent error-prone measurements on time-dependent covariates, early attempts imputed their values at each event time through the "last value carried forward" and the "naive regression". These approaches could lead to biased estimation and erroneous inference (Prentice (1982); Tsiatis and Davidian (2001)). A popular approach uses a shared random effects joint model that assumes that the longitudinal observations follow a mixed effects model and the survival time depends on the random effects of the mixed effects model through a proportional hazards model.

Various approaches have been proposed under the joint model framework assuming *non-informative* observation times, including regression calibration (e.g., Pawitan and Self (1993); Tsiatis, DeGruttola and Wulfsohn (1995); Dafni and Tsiatis (1998); Liao et al. (2011)) and likelihood-based approaches (Faucett and Thomas (1996); Wulfsohn and Tsiatis (1997); Song, Davidian and Tsiatis (2002a); Ding and Wang (2008); Xu, Baines and Wang (2014)). Regression calibration can reduce bias compared to the naive methods, but is still inconsistent (Tsiatis and Davidian (2001)). Hsieh, Tseng and Wang (2006) showed that the likelihood methods are robust to misspecification of the random effect distribution when there is rich enough longitudinal information. However, the likelihood based approaches are usually computationally intensive, and they generally assume that the censoring time is independent of survival time and the underlying random effects, which can be questionable in certain situations. Two appealing alternative approaches are the conditional score (Tsiatis and Davidian (2001); Song, Davidian and Tsiatis (2002b)) and the corrected score (Wang (2006)). Both methods require no distributional assumptions on the random effects. But they require the observation times to be independent of the survival time, the random effects and other covariates, they make inefficient use of the longitudinal data – only the longitudinal information by each event time is used to estimate the least square substitutes of the true covariates, and they can be sensitive to the normality error assumption.

Informative observation times have attracted considerable attention when the focus is on estimation of the longitudinal trajectories (Lin, Scharfstein and Rosenheck (2004); Liang, Lu and Ying (2009); Sun, Sun and Liu (2007); Sun et al. (2012)). But very limited research has been done for shared random effects models when the main interest is in assessing covariate effects on survival time — to the best of our knowledge, only a likelihood approach was proposed by adding a third model for the observation times (Liu, Huang and O'Quigley (2008)), which poses additional challenges in implementation. There is a lack of flexible and computationally efficient approaches.

To fill this gap, we develop an improved corrected score approach that allows the observation times to depend on the survival time, the random effects, or other covariates. Instead of globally correcting the overall bias of the naive estimating function, we use local correction to correct each biased term separately. We utilize the least square estimates based on all available longitudinal observations, which may substantially improve the efficiency over the existing corrected score and conditional score estimators, and can be more robust to deviations from normality of the error. As there is no need to model the observation times, the approach is much simpler to implement than the likelihood-based approaches.

The paper is organized as follows. In Section 2, we give the definition of the model. In Section 3, the existing conditional score and corrected score estimators are described and the inconsistency of the estimators is shown under informative observation times. In Section 4, we propose the improved corrected score estimator and derive its asymptotic properties. The finite sample performance of the estimator is assessed by simulation studies in Section 5, and is illustrated by an application to the ACTG 175 data in Section 6. We conclude with discussion in Section 7.

2. Model Definition

Let T denote the survival time and C the censoring time. The observed survival data are $V = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Let Z denote p time-independent covariates. For simplicity, we consider a single time-dependent covariate $X(t)$; it is straightforward to extend to multiple time-dependent covariates. The covariate process $X(t)$ is not observed directly; rather, longitudinal measurements of $X(t)$ are taken at ordered times $u = (u_1, \ldots, u_m)^T$ with the observed values $W = (W_1, \ldots, W_m)$. Usually, $u_m \leq V$.

Suppose that the longitudinal covariate process follows the linear mixed effects model

$$X(t) = f^T(t)\alpha, \qquad W_j = X(u_j) + e_j,$$

where $f(t)$ is a known q-dimensional function of t, α is a q-dimensional random effect, and $j = 1, \ldots, m$. This allows flexibility in modeling the longitudinal trajectory via polynomial or spline models. No distributional assumption is placed on α, nor is one needed. The error e_j is assumed to be normal with mean zero and variance σ^2, and independent of (T, C, Z, u) given α. For simplicity, we assume that the errors $e = (e_1, \ldots, e_m)^T$ are independent across time; this can be relaxed as discussed in Section 7.

Suppose that the hazard of failure depends on the covariates $X(t)$ and Z through the proportional hazards model

$$\lambda(t|X) = \lambda_0(t) \exp\left\{\beta_0 X(t) + \gamma_0^T Z\right\}, \tag{2.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and β_0 and γ_0 are regression coefficients. We assume that the survival time T is independent of the censoring time C given α and Z. Let $\{(T_i, C_i, V_i, \Delta_i, a_i, Z_i, W_i, e_i, u_i, m_i) : i = 1, \ldots, n\}$ be independent and identically distributed samples of $(T, C, V, \Delta, a, Z, W, e, u, m)$. Our interest focuses on estimation of the parameters $\theta_0 = (\beta_0, \gamma_0^T)^T$ using the observed data $\{(V_i, \Delta_i, Z_i, W_i, u_i, m_i) : i = 1, \ldots, n\}$. Let β and γ denote elements in the parameter space of β_0 and γ_0, respectively, and $\theta = (\beta, \eta^T)^T$.

3. Existing Conditional Score and Corrected Score Estimators

For now, we assume the error variance σ^2 is known. Both the conditional score and corrected score approaches use the least square estimates $\widehat{X}_i(t) = f^T(t)\widehat{\alpha}_i(t)$ of $X_i(t)$, where $\widehat{\alpha}_i(t)$ is the least square estimator of α_i based on the longitudinal observations on the ith subject by time t, $u_i(t) = \{u_{ij} : u_{ij} \leq t\}$.

This requires at least q observations by time t. Let $m_i(t)$ be the number of observations in $u_i(t)$, $N_i(t) = I(V_i \le t, \Delta_i = 1, m_i(t) \ge q)$ be the counting process for the events, $Y_i(t) = I(V_i \ge t,\ m_i(t) \ge q)$ be the at-risk process, and $\sigma_X^2(\widehat{X}_i(t))$ be the variance of $\widehat{X}_i(t)$ conditional on $(\alpha_i, u_i(t))$.

3.1. Conditional score estimator

The conditional score approach (Tsiatis and Davidian (2001)) treats the true covariate $X_i(t)$ as a nuisance parameter and constructs a "sufficient statistic" $S_i(t;\beta) = \widehat{X}_i(t) + \sigma_X^2(\widehat{X}_i, t)\beta dN_i(t)$ for $X_i(t)$ based on the distribution of $(dN_i(t), \widehat{X}_i(t))$ conditional on $(\alpha_i, Z_i, u_i(t),\ Y_i(t) = 1)$. The conditional hazard given $S_i(t;\beta)$ is then used to derive the estimating equation

$$\widehat{U}^{cd}(\theta; Y, N, \widehat{X}, Z) = n^{-1}\sum_{i=1}^{n}\int_0^L \left\{ \left(S_i(t,\beta), Z_i^T\right)^T - \frac{\widehat{G}_1^{cd}(t,\theta; Y, S, Z)}{\widehat{G}_0^{cd}(t,\theta; Y, S, Z)} \right\} dN_i(t) = 0, \tag{3.1}$$

at a given time L. Here for $r = 0, 1, 2$, $\widehat{G}_r^{cd}(t,\theta; Y, S, Z) = n^{-1}\sum_{i=1}^n H_{r,i}^{cd}(t,\theta; Y, S, Z)$, and

$$H_{r,i}^{cd}(t,\theta; Y, S, Z) = Y_i(t)\left(S_i(t,\beta), Z_i^T\right)^{T\otimes r} \exp\left\{ \beta S_i(t,\beta) + \gamma^T Z_i - \sigma_X^2(\widehat{X}_i(t))\frac{\beta^2}{2} \right\}, \tag{3.2}$$

where for a vector a, $a^{\otimes r} = 1, a, aa^T$ for $r = 0, 1, 2$, respectively. The conditional independence of $dN_i(t)$ and $\widehat{X}_i(t)$ is essential to construct the "sufficient statistic" based on their joint distribution (Tsiatis and Davidian (2001)); this would not work if the least square estimates were calculated using all the longitudinal observations.

3.2. Corrected score estimator

Wang (2006) proposed a corrected score estimating function by directly subtracting the bias from the naive estimating function that replaces the unobserved $X(t)$ by $\widehat{X}(t)$. A variation of the corrected score approach was considered in Song and Wang (2008) with a slight modification on the estimating function to have a form closer to the conditional score estimating function. The modified corrected score estimating function can be written as

$$\begin{aligned}
&\widehat{U}^{cr}(\theta; Y, N, \widehat{X}, Z)\\
&= n^{-1}\sum_{i=1}^{n}\int_0^L \left\{ \left(\widehat{X}_i(t), Z_i^T\right)^T - \frac{\widehat{G}_1^{cr}(t,\theta; Y, \widehat{X}, Z)}{\widehat{G}_0^{cr}(t,\theta; Y, \widehat{X}, Z)} + (\sigma_X^2(\widehat{X}_i(t))\beta, 0_p^T)^T \right\} dN_i(t)\\
&= 0,
\end{aligned} \tag{3.3}$$

where for $r = 0, 1, 2$, $\widehat{G}_r^{cr}(t, \theta; Y, \widehat{X}, Z) = n^{-1} \sum_{i=1}^n H_{r,i}^{cr}(t, \theta; Y, \widehat{X}, Z)$, and

$$H_{r,i}^{cr}(t, \theta; Y, \widehat{X}, Z) = Y_i(t)\Big(\widehat{X}_i(t), Z_i^T\Big)^{T\otimes r} \exp\Big\{\beta \widehat{X}_i(t) + \gamma^T Z_i - \sigma_X^2(\widehat{X}_i(t))\frac{\beta^2}{2}\Big\}.$$

When the longitudinal covariates are observed at the same times for all subjects, the modified corrected score estimator is the same as the original corrected score estimator by Wang (2006). We refer to these estimators as simple corrected score estimators, and they are asymptotically equivalent to the conditional score estimator (Song and Wang (2008)).

3.3. Inconsistency under informative observation times

The simple corrected score approaches directly subtract the overall bias from the naive estimating function (Wang (2006)) or with some adjustment (Song and Wang (2008)). We refer to this type of correction as *global correction*. As the overall bias depends on the unknown joint distribution of V_i, Δ_i, X_i and $\widehat{X_i}(t)$, which is not easy to evaluate in general, it is calculated under the simple assumption of non-informative observation times. Consequently, the methods do not work when the observation times are informative. We investigate the inconsistency of the modified corrected score estimator based on (3.3), which is similar to the original corrected score estimator (Wang (2006)).

With some algebra, the limit of $\widehat{U}^{cr}(\theta; Y, N, \widehat{X}, Z)$ in (3.3) can be written as

$$U^{cr}(\theta; Y, N, \widehat{X}, Z) = U^I(\theta; Y, N, X, Z) + b(t, \theta; Y, N, \widehat{X}, X, Z), \tag{3.4}$$

where

$$U^I(\theta; Y, N, X, Z) = E\left[\int_0^L \left\{(X_i(t), Z_i)^T - \frac{G_1(t, \theta; Y, X, Z)}{G_0(t, \theta; Y, X, Z)}\right\} dN_i(t)\right],$$

$$b(t, \theta; Y, N, \widehat{X}, X, Z) = E\Bigg(\int_0^L \Bigg\{ \frac{E\Big[Y_i(t)(\sigma_X^2(\widehat{X}_i(t))\beta, 0_p^T)^T \exp\big\{\beta X_i(t) + \gamma^T Z_i\big\}\Big]}{G_0(t, \theta; Y, X, Z)} - (\sigma_X^2(\widehat{X}_i(t))\beta, 0_p^T)^T \Bigg\} dN_i(t)\Bigg),$$

and $G_r(t, \theta; Y, X, Z) = E(Y_i(t)(X_i(t), Z_i)^{T\otimes r} \exp\{\beta X_i(t) + \gamma^T Z\})$ for $r = 0, 1, 2$. Here $U^I(\theta; Y, N, X, Z)$ is the limit of the weighted partial likelihood function with weight $I(m_i(t) \geqslant q)$, which equals zero when $\theta = \theta_0$. If the observation times do not depend on α, V, and Z, it is clear that $\sigma_X^2(\widehat{X}_i(t))\beta$ is independent of $Y_i(t)\exp\{\beta X_i(t) + \gamma^T Z_i\}$ and $N_i(t)$. It follows that $b(t, \theta; Y, N, \widehat{X}, X, Z) = 0$ and then the consistency of the estimator. However, if the observation times depend on α, V, or Z, in general $b(t, \theta_0; Y, N, \widehat{X}, X, Z) \neq 0$ and the estimator is not consistent.

4. Improved Corrected Score Estimator

Instead of globally correcting the overall bias of the naive estimating function, we propose correcting the biased terms separately, which we refer to as *local correction.*

The naive estimating function that substitutes $\widehat{X}(t)$ for $X(t)$ can be written as

$$n^{-1}\sum_{i=1}^{n}\int_0^L \left(\widehat{X}_i(t), Z_i^T\right)^T dN_i(t) - \int_0^L \frac{\widehat{G}_1^N(t,\theta;Y,\widehat{X},Z)}{\widehat{G}_0^N(t,\theta;Y,\widehat{X},Z)} dn^{-1}\sum_{i=1}^{n} N_i(t), \quad (4.1)$$

where for $r=0,1$, $\widehat{G}_r^N(t,\theta;Y,\widehat{X},Z) = n^{-1}\sum_{i=1}^{n} H_{r,i}^N(t,\theta;Y,\widehat{X},Z)$,

$$H_{r,i}^N(t,\theta;Y,\widehat{X},Z) = Y_i(t)\left(\widehat{X}_i(t), Z_i\right)^{T\otimes r}\exp\left\{\beta\widehat{X}_i(t)+\gamma^T Z_i\right\}.$$

It has the same form as the partial likelihood estimating function with weight $I(m_i(t)\geqslant q)$. Bias arises from substituting $\widehat{X}$ for X. The function (4.1) contains four empirical processes, three of which contain $\widehat{X}$. Among the three, with respect to the corresponding processes using the true X, $n^{-1}\sum_{i=1}^{n}\int_0^L(\widehat{X}_i(t), Z_i^T)^T dN_i(t)$ is not biased, but $\widehat{G}_1^N(t,\theta;Y,\widehat{X},Z)$ and $\widehat{G}_0^N(t,\theta;Y,\widehat{X},Z)$ are. Instead of subtracting the induced overall bias, we propose correcting each empirical process separately. Specifically, for $r=0,1$, we replace $\widehat{G}_r^N(t,\theta;Y,\widehat{X},Z)$ by $\widehat{G}_r^{cr*}(t,\theta;Y,\widehat{X},Z)$ $= n^{-1}\sum_{i=1}^{n} H_{r,i}^{cr*}(t,\theta;Y,\widehat{X},Z)$ with

$$\begin{aligned} H_{r,i}^{cr*}(t,\theta;Y,\widehat{X},Z) &= Y_i(t)\left\{\widehat{X}_i(t)-\sigma_X^2(\widehat{X}_i(t))\beta, Z_i^T\right\}^{T\otimes r} \\ &\quad\times\exp\left\{\beta^T\widehat{X}_i(t)+\gamma^T Z_i-\sigma_X^2(\widehat{X}_i(t))\frac{\beta^2}{2}\right\}. \end{aligned}$$

These can be shown to be unbiased regardless of whether the observation times are informative or not. Note that $\widehat{G}_0^{cr*}(t,\theta;Y,\widehat{X},Z) = \widehat{G}_0^{cr}(t,\theta;Y,\widehat{X},Z)$. Using the local correction technique, we obtain the estimating equation

$$\begin{aligned} \widehat{U}^{cr*}(\theta;Y,N,\widehat{X},Z) &= n^{-1}\sum_{i=1}^{n}\int_0^L\left\{\left(\widehat{X}_i(t), Z_i^T\right)^T - \frac{\widehat{G}_1^{cr*}(t,\theta;Y,\widehat{X},Z)}{\widehat{G}_0^{cr}(t,\theta;Y,\widehat{X},Z)}\right\}dN_i(t) \\ &= 0. \end{aligned} \quad (4.2)$$

Comparing to (3.3), the estimating function (4.2) does not have a separate correction term for bias, but $\widehat{G}_1^{cr}(t,\theta;Y,X,Z)$ is replaced by $\widehat{G}_1^{cr*}(t,\theta;Y,\widehat{X},Z)$ with $H_{1,i}^{cr}(t,\theta;Y,X,Z)$ replaced by $H_{1,i}^{cr*}(t,\theta;Y,\widehat{X},Z)$, where $\widehat{X}_i(t)$ is replaced by $\widehat{X}_i(t)-\sigma_X^2(\widehat{X}_i(t))\beta$ in the linear term. This works even when the observation times are informative.

We wish to modify the corrected score approach so that it will be more efficient. Note that the least square estimates $\widehat{X}_i(t)$ in (3.3) are calculated using the longitudinal observations by time t only, the observations after t are not used. In addition, at any event time, subjects with less than q observations by then are excluded in the calculation even though there are more observations available later. One reason of using $\widehat{X}(t)$ might be due to the intuition that hazard may only depend on information by time t. However, although the hazard model (2.1) depends on the time-dependent covariate $X_i(t)$, it essentially depends on the underlying random effects α_i, which does not change over time and can be estimated using all the longitudinal observations available for subject i, usually depending on V_i. The resulting least square estimator $\widehat{\alpha}_i^*$ of α_i is more efficient than $\widehat{\alpha}_i$. Thus, to improve the efficiency, it is tempting to replace $\widehat{X}_i(t)$ by $\widehat{X}_i^*(t) = f^T(t)\widehat{\alpha}_i^*$, and correspondingly replace the variance $\sigma_X^2(\widehat{X}_i(t))$ by the variance $\sigma_X^2(\widehat{X}_i^*(t))$ of $\widehat{X}_i^*(t)$ conditional on α_i and u_i, the counting process $N_i(t)$ by $N_i^*(t) = I(V_i \leq t, \Delta_i = 1, m_i \geq q)$ and the at-risk process $Y_i(t)$ by $Y_i^*(t) = I(V_i \geq t,\ m_i \geq q)$. Thus all subjects with at least q observations are included in the calculation regardless whether the observations are before or after time t. However, this does not work with the simple corrected score approaches as the resulting estimating functions are no longer consistent, even under non-informative observation times. This can be easily seen from (3.4) as the limit of $\widehat{U}^{cr}(\theta; Y^*, N^*, \widehat{X}^*, Z)$ is in the same form and the corresponding bias $b(t, \theta; Y^*, N^*, \widehat{X}^*, X, Z) \neq 0$ in general as $\sigma_X^2(\widehat{X}_i^*(t))$ depends on V_i. On the contrary, we may apply the replacement to the estimating function (4.2). Specifically, we propose the improved corrected score estimating equation using $\widehat{X}_i^*(t)$:

$$\widehat{U}^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z) = 0. \tag{4.3}$$

The asymptotic properties of the improved estimator are summarized in the following theorem, with the proof given in the Appendix.

Theorem 1. *Under conditions C1–C5 in the Appendix, a solution* $\widehat{\theta}^* = (\widehat{\beta}^*, \widehat{\gamma}^{*T})^T$ *to* (4.3) *exists and converges to* θ_0 *almost surely. In addition,* $n^{1/2}(\widehat{\theta}^* - \theta_0)$ *is asymptotically normal with mean zero and variance*

$$\begin{aligned}\Omega^{cr*} = &\{\Gamma^{*-1}(\theta_0; Y^*, N^*, \widehat{X}^*, Z)\}^{-1} var\{\varphi_i^*(\theta_0; Y^*, N^*, \widehat{X}^*, Z)\} \\ &\Gamma^{*-1}(\theta_0; Y^*, N^*, \widehat{X}^*, Z),\end{aligned} \tag{4.4}$$

where

$$\Gamma^*(\theta; Y^*, N^*, \widehat{X}^*, Z)$$
$$= \int_0^L \left(\frac{G_2^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)G_0^{cr}(t,\theta;Y^*,\widehat{X}^*,Z) - G_1^{cr*\otimes 2}(t,\theta;Y^*,\widehat{X}^*,,Z)}{G_0^{cr2}(t,\theta;Y^*,\widehat{X}^*,Z)} \right.$$
$$\left. - \frac{E\left[Y_i^*(t) diag(\sigma_X^2(\widehat{X}_i^*(t)), 0_p^T)^T \exp\left\{\beta \widehat{X}_i^*(t) + \gamma^T Z_i - \sigma_X^2(\widehat{X}_i^*(t))\beta^2/2\right\}\right]}{G_0^{cr}(t,\theta;Y^*,N^*,\widehat{X}^*,Z)} \right)$$
$$dE\{N_i^*(t)\},$$

$$\varphi_i^*(\theta; Y^*, N^*, \widehat{X}^*, Z)$$
$$= \int_0^L \left\{ (\widehat{X}_i^*(t), Z_i^T)^T - \frac{G_1^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)}{G_0^{cr}(t,\theta;Y^*,\widehat{X}^*,Z)} \right\} dN_i^*(t)$$
$$- \int_0^L \left\{ \frac{H_{1,i}^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)}{G_0^{cr}(t,\theta;Y^*,\widehat{X}^*,Z)} - \frac{H_{0,i}^{cr}(t,\theta;Y^*,\widehat{X}^*,Z)G_1^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)}{G_0^{cr2}(t,\theta;Y^*,\widehat{X}^*,Z)} \right\}$$
$$d\{EN_i^*(t)\},$$

$$G_r^{cr*}(t,\theta;Y^*,N^*,\widehat{X}^*,Z) = E\{H_r^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)\}, \qquad r = 0,1,2.$$

To compare the efficiency of the improved corrected score estimator with the simple estimators, we consider the estimator $\widehat{\theta}$ obtained from (4.2), which is equivalent to the simple corrected score estimators when the time-dependent covariates are observed at the same times for all subjects. The asymptotic variance of $n^{1/2}(\widehat{\theta} - \theta_0)$ is

$$\Omega^{cr} = \{\Gamma^{*-1}(\theta_0; Y, N, \widehat{X}, Z)\}^{-1} var\left\{\varphi_i^*(\theta_0; Y, N, \widehat{X}, Z)\right\} \Gamma^{*-1}(\theta_0; Y, N, \widehat{X}, Z),$$

which just replaces $\widehat{X}^*$, Y^*, N^* by $\widehat{X}$, Y, N in (4.4), respectively. We show in the Appendix that $\Omega^{cr} - \Omega^{cr*}$ is nonnegative definite ($\Omega^{cr} \geq_{pd} \Omega^{cr*}$). Hence the improved corrected score estimator is more efficient than the simple estimators.

Another significant gain from this improvement is that the improved corrected score estimator does not require the observation times of the longitudinal data to be independent of V, α, and Z, which is required by the conditional score and simple corrected score estimators. In addition, the least square estimate $\widehat{\alpha}_i^*$ needs to be calculated only once for each subject, while $\widehat{a}_i(t)$ needs to be calculated separately at each failure time before V_i.

In practice, the error variance σ^2 is usually unknown. It can be estimated by the method of moment estimate $\widehat{\sigma}^2$ (Tsiatis and Davidian (2001)). The improved estimators can be obtained by replacing σ^2 by $\widehat{\sigma}^2$ in (4.3). The estimators are still consistent and normal. The asymptotic distribution can be derived by stacking the estimating functions for σ^2 and the estimating functions in (4.3). The variances can be estimated by replacing the unknown parameters $(\beta, \gamma, \sigma^2)$ by $(\widehat{\beta}, \widehat{\gamma}, \widehat{\sigma}^2)$ and the expectations by empirical averages.

In some situations, $X(t)$ may follow the same longitudinal trajectory after the event time and may be measured intermittently until being censored. This can happen when the event is not terminal (death) and $X(t)$ is endogenous. We can calculate $\widehat{X}_i^*$ using all the longitudinal observations and the improved corrected score estimator still works.

5. Simulation Studies

We conducted simulation studies to evaluate the performance of the improved corrected score estimator under various scenarios. We first considered the case when the observation times were independent of the survival time and the covariates. Here, the survival time depended on a single covariate $X(t) = \alpha_0 + \alpha_1 t$ through the proportional hazards model, with $\lambda_0(t) = I(t \geq 2)$ and $\beta_0 = -1$. The random effects (α_0, α_1) were generated from a normal distribution with mean $(4.173, -0.0103)$, variance $(1.24, 0.003)$ and covariance -0.0114. The censoring time was generated from an exponential distribution with mean 300 and truncated at 80, leading to a censoring rate of 37%. The covariate $X(t)$ was observed at times $u^* = 0, 2, 8, 16, 24, \ldots, 80$ before the survival time and the censoring time. On average there were 6.00 observations per subject. The measurement error was generated from a normal distribution with mean 0 and variance σ^2.

We generated 1,000 simulated data sets with sample size $n = 500$ or $1,000$ and $\sigma^2 = 0.15$ or 0.30. For each data set, the regression coefficient β_0 was estimated in five ways: (i) using the "ideal" approach where the true values of $X(t)$ were used; (ii) using the conditional score approach; (iii) using the simple corrected score approach based on (3.3); (iv) using the improved corrected score approach; (v) using the conditional score approach with $\widehat{X}(t)$ replaced by $\widehat{X}^*(t)$; (vi) using the simple corrected score approach with $\widehat{X}(t)$ replaced by $\widehat{X}^*(t)$. For all methods, we calculated the estimate and standard error and constructed the 95% Wald confidence interval.

The results are shown in Table 1. The simple corrected score approach does not converge or has outlier estimates for some simulated data sets when the error variance is relatively large ($\sigma^2 = 0.30$). Its performance is close to that of the conditional score estimator when the error variance is relatively small ($\sigma^2 = 0.15$), but the standard deviation is still larger. The conditional score and the improved

Table 1. Simulation results in the case of a single covariate.

		$n = 500$					$n = 1,000$				
		Bias	BiasM	SD	SE	CP	Bias	BiasM	SD	SE	CP
	ideal	−0.002	−0.001	0.052	0.052	0.953	−0.001	−0.000	0.038	0.037	0.948
$\sigma^2 = 0.15$	CDS	−0.007	−0.006	0.080	0.076	0.938	−0.004	−0.003	0.061	0.056	0.928
	CRS	−0.029	−0.024	0.093	0.087	0.956	−0.016	−0.014	0.067	0.061	0.923
	ICRS	−0.008	−0.006	0.065	0.065	0.956	−0.005	−0.003	0.048	0.046	0.943
	CDSA	−0.426	−0.413	0.152	0.128	0.056	−0.428	−0.425	0.114	0.099	0.012
	CRSA	1.542	−0.489	4.802	0.461	0.269	0.005	−0.538	2.745	0.244	0.048
$\sigma^2 = 0.30$	CDS	−0.013	−0.015	0.112	0.094	0.900	−0.006	−0.009	0.090	0.073	0.885
	CRS	0.671	−0.059	2.346	0.255	0.874	0.051	−0.066	1.026	0.157	0.951
	ICRS	−0.019	−0.016	0.083	0.080	0.949	−0.011	−0.010	0.060	0.056	0.940
	CDSA	−0.742	−0.726	0.230	0.167	0.017	−0.751	−0.746	0.184	0.132	0.006
	CRSA	6.989	6.962	1.041	0.618	0.007	7.366	7.349	0.909	0.464	0.005

CDS, conditional score; CRS, corrected score; ICRS, improved corrected score; CDSA, conditional score with $\widehat{X}(t)$ replaced by $\widehat{X}^*(t)$; CRSA, corrected score with $\widehat{X}(t)$ replaced by $\widehat{X}^*(t)$; BiasM, empirical median bias; SD, empirical standard deviation; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

corrected score estimators show negligible bias close to the "ideal" estimator. The improved corrected score estimator has better coverage probabilities and smaller standard deviations than the conditional score estimator, with relative efficiencies ranging between 1.53 to 2.24. The conditional score and simple corrected score estimators with $\widehat{X}(t)$ replaced by $\widehat{X}^*(t)$ are biased.

We also conducted simulations with an additional time-independent covariate Z in the proportional hazards model. The covariate Z was generated from a Bernoulli distribution with probability 0.5. The corresponding regression coefficient was $\gamma_0 = -1$. The censoring rate was about 57%. Table 2 shows the results from methods (i)-(iv) for $\sigma^2 = 0.15$, which indicates that the improved corrected score approach also improves the efficiency of estimation of γ_0.

We modified the scenario so that the observation times depended on the survival time or the underlying random effects. We considered two cases with a single covariate $X(t)$ where the observations may be missing after baseline with the missing probability (1) $\{1+\exp(1+0.5\alpha_0+30\alpha_1)\}^{-1}$ or (2)$\{1+\exp(1+0.05V)\}^{-1}$. The regression parameter β_0 was estimated using methods (i)-(iv). The average numbers of observations per subject were 5.62 and 5.77, respectively. The results for $\sigma^2 = 0.15$ are presented in Table 3. In both cases the conditional score and the simple corrected score estimators show obvious bias and poor coverage

Table 2. Simulation results in the case of two covariates.

		$n = 500$				$n = 1,000$			
		Bias	SD	SE	CP	Bias	SD	SE	CP
β	ideal	−0.009	0.062	0.059	0.946	−0.005	0.044	0.042	0.940
	CDS	−0.018	0.091	0.081	0.924	−0.011	0.064	0.059	0.934
	CRS	−0.044	0.109	0.093	0.948	−0.024	0.069	0.064	0.931
	ICRS	−0.017	0.076	0.072	0.942	−0.010	0.053	0.051	0.945
γ	ideal	−0.001	0.153	0.155	0.958	−0.001	0.117	0.109	0.930
	CDS	−0.004	0.167	0.166	0.957	−0.002	0.123	0.117	0.933
	CRS	−0.010	0.171	0.169	0.954	−0.005	0.124	0.118	0.934
	ICRS	−0.004	0.157	0.159	0.965	−0.004	0.119	0.112	0.931

CDS, conditional score; CRS, corrected score; ICRS, improved corrected score; SD, empirical standard deviation; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

Table 3. Simulation results with observation times depending on V or α.

	$n = 500$					$n = 1,000$				
	Bias	BiasM	SD	SE	CP	Bias	BiasM	SD	SE	CP
					case 1					
ideal	−0.001	0.002	0.053	0.055	0.951	−0.001	−0.001	0.037	0.039	0.957
CDS	−0.140	−0.131	0.106	0.093	0.657	−0.161	−0.156	0.076	0.071	0.394
CRS	6.506	6.783	2.184	1.193	0.061	7.218	7.252	1.499	0.981	0.018
ICRS	−0.042	−0.039	0.091	0.088	0.950	−0.038	−0.033	0.061	0.061	0.930
					case 2					
ideal	−0.001	0.003	0.054	0.054	0.950	0.000	0.001	0.037	0.038	0.954
CDS	−0.180	−0.179	0.097	0.089	0.482	−0.201	−0.197	0.071	0.068	0.143
CRS	6.389	7.900	3.888	1.129	0.229	6.617	8.384	4.096	0.878	0.224
ICRS	−0.033	−0.030	0.079	0.077	0.952	−0.031	−0.029	0.053	0.054	0.933

CDS, conditional score; CRS, corrected score; ICRS, improved corrected score; BiasM, empirical median bias; SD, empirical standard deviation; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

probabilities, while the improved corrected score estimator still works well.

We also conducted simulations to evaluate the sensitivity of the approaches on deviations from normality of the error. The scenario was the same as that for Table 1 except that the error was generated from non-normal distributions with mean zero and variance $\sigma^2 = 0.15$ by mixing two normal distributions $N(\mu_1, \sigma_N^2)$ and $N(\mu_2, \sigma_N^2)$ with $\mu_1 = (1-p)s\sigma_N$, $\mu_2 = -ps\sigma_N$ and $\sigma_N^2 = \sigma^2/(1+p(1-p)s^2)$, where p is the mixing proportion and the distance between the means of the two normal distributions is $s\sigma_N$. We considered two cases, (a) a skewed bimodal

Table 4. Simulation results with mixture of normal error.

	$n = 500$				$n = 1,000$			
	Bias	SD	SE	CP	Bias	SD	SE	CP
ideal	−0.006	0.053	0.052	0.948	0.000	0.037	0.037	0.949
				case (a)				
CDS	−0.067	0.089	0.080	0.860	−0.062	0.061	0.060	0.837
CRS	−0.105	0.115	0.101	0.881	−0.081	0.069	0.067	0.819
ICRS	−0.026	0.070	0.067	0.942	−0.014	0.048	0.047	0.950
				case (b)				
CDS	−0.079	0.083	0.080	0.841	−0.079	0.062	0.059	0.752
CRS	−0.125	0.113	0.105	0.866	−0.105	0.073	0.069	0.726
ICRS	−0.023	0.069	0.066	0.936	−0.013	0.048	0.046	0.948

CDS, conditional score; CRS, corrected score; ICRS, improved corrected score; SD, empirical standard deviation; SE, average of estimated standard errors; CP, coverage probability of the 95% Wald confidence interval.

mixture of two normals with $p = 0.3$ and $s = 3$, and b) a symmetric bimodal mixture of two normals with $p = 0.5$ and $s = 10$. The densities of the two distributions are shown in Figure 3. Table 4 shows the results from methods (i)-(iv). The conditional score and the simple corrected score estimators both show some bias and have coverage probabilities obviously below the nominal level, while the improved corrected score estimator seems more robust to the deviations from normality. One possible explanation is that $\widehat{X}^*(t)$ is closer to normal than $\widehat{X}(t)$.

Simulation studies were also conducted under other scenarios: When the random effects followed a skewed mixture of normal distribution, the results were similar; When the longitudinal covariate was observed following the same trend after the event time, the improved corrected score estimator gained more efficiency by using all the longitudinal observations in calculating $\widehat{X}^*(t)$. Overall, the simulation evidence suggests that the improved corrected score estimator outperforms the conditional score and the simple corrected score estimators. The computation times for the three approaches were comparable in our simulation studies. Although the improved corrected score approach only requires calculating the least square estimate once for each subject, it involves additional computation induced by the extra term in $\widehat{G}_1^{cr*}$ versus $\widehat{G}_1^{cr}$, especially in computing the variance.

6. Application

We applied the proposed approach to the ACTG 175 data. Interest is in

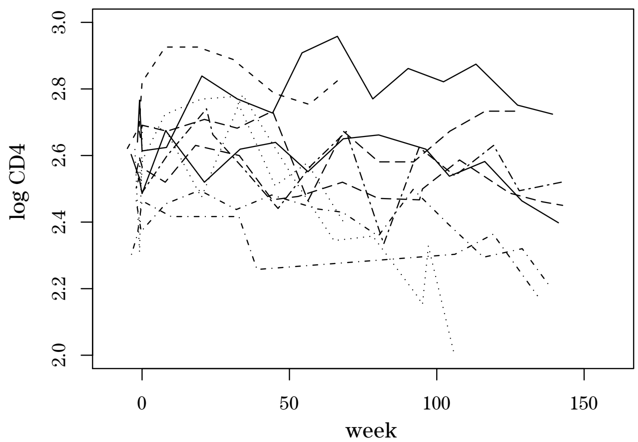

Figure 1. Trajectories of log CD4 for 10 randomly selected subjects.

Table 5. Results for ACTG 175 data.

	β		γ	
	Est	SE	Est	SE
CDS	−2.208	0.202	0.141	0.258
CRS	−2.342	0.262	0.230	0.306
ICRS	−2.378	0.176	−0.091	0.178

CDS, conditional score; CRS, corrected score; ICRS, improved corrected score; Est, estimate; SE, standard error.

assessing the effect of CD4 count and treatment on time to AIDS or death. There were 308 events observed during the study with an average of 8.2 CD4 measurements per subject. 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 an approximate linear decline. The logarithmic transformation is usually used for CD4 count to achieve approximate within-subject normality and constant variance. Because only nine events occurred by week 12, for simplicity, we considered the data after week 12 and assumed $X(u) = \alpha_0 + \alpha_1 u$ represents the inherent log10 CD4 count at time u. Figure 2 shows the residual plots from the least square estimates and the corresponding Q-Q plot. It seems reasonable to assume constant error variance, and the error distribution may be short-tailed relative to the normal but symmetric.

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. Let $Z(u) = I(\text{treatment} \neq \text{zidovudine alone})$ be the treatment indicator. The hazard model includes the

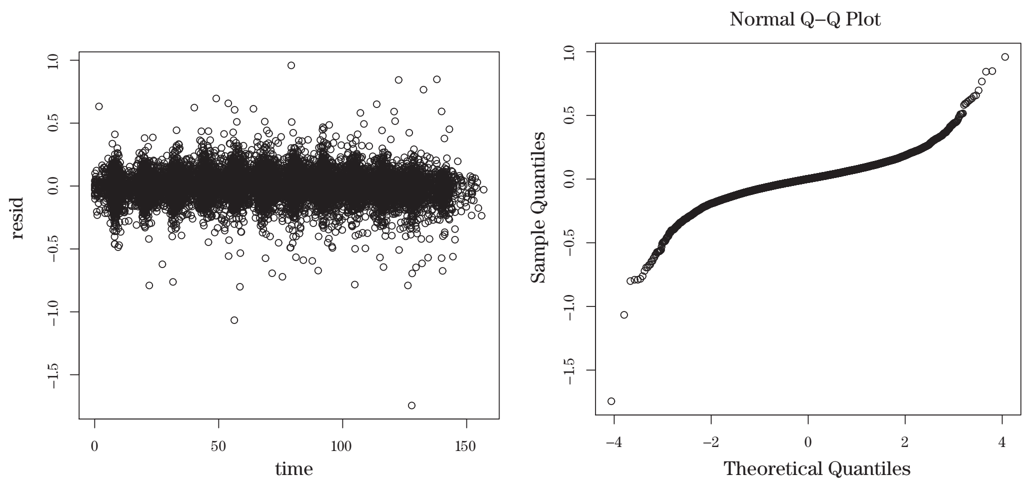

Figure 2. Left: residual plot; Right, Q-Q plot of the residuals.

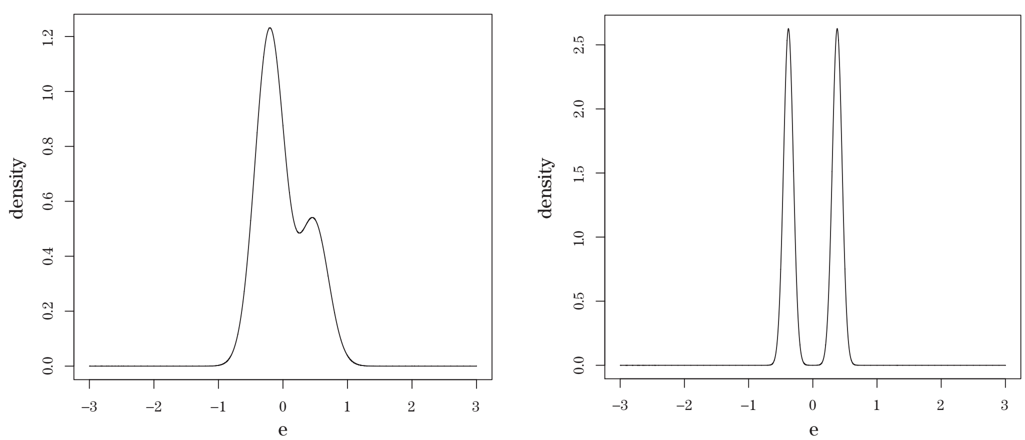

Figure 3. Densities of non-normal errors. Left: (a) $p = 0.3$, $s = 3$; Right: (b) $p = 0.5$, $s = 10$.

two covariate $X(u)$ and $Z(u)$.

We estimated the regression coefficients using the conditional score, simple corrected score (3.3), and improved corrected score approaches. The results are shown in Table 5. The estimates of β are similar for all three methods. The improved corrected score estimate of γ is negative while the conditional score and simple corrected score estimates are positive, although none is significant. This may be related to the informative observation times. The improved corrected score estimates of β and γ have smaller standard errors, which reflects efficiency gain over the conditional score and the simple corrected score estimates.

7. Discussion

We have proposed an improved corrected score estimator for the proportional hazards model with time-dependent covariates intermittently measured with error. The estimator may substantially improve the efficiency over the conditional score and the simple corrected score estimators, and is more robust to deviations from the normality of the error. In addition, it does not require the observation times to be non-informative.

We have focused on improving the estimation of the regression coefficients. The same technique may be applied to improve the Breslow-type estimator of the cumulative baseline hazard (Song and Wang (2008)) by substituting $\widehat{X}_i^*(u)$ for $\widehat{X}_i(u)$.

The improved corrected score approach can be easily extended to multiple error-contaminated time-independent covariates with informative number of replicates or time-dependent covariates with informative observation times (Song, Davidian and Tsiatis (2002b); Song and Wang (2008)), or more flexible models such as the varying coefficient proportional hazards model (Song and Wang (2008)) and longitudinal models with subject-specific changepoints (Tapsoba, Lee and Wang (2011)). For simplicity, we have focused on the case when the errors are independent across time. This can be generalized to other error correlation structures, such as the exponential correlation ((Diggle et al., 2002, p.56)). In addition, the survival model can be generalized to include functions of the random effects (Song, Davidian and Tsiatis (2002b)).

Acknowledgment

This research was partially supported by NSF grant DMS-1106816 and NIH grants HL121347 and 1R21CA201207.

Appendix

A.1. Regularity conditions

To derive the asymptotic properties of the improved estimator, we need some regularity conditions.

C1. $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is continuous in $[0, L]$, and $\Lambda_0(L) < \infty$;

C2. $P(V \geq L, m \geq q) > 0$;

C3. $\sup_{t\in[0,L]} E\left\{X^2(t)\right\} < \infty$, $E(Z^T Z) < \infty$. For a compact neighborhood

$\mathcal{N}(\theta_0)$ of θ_0,

$$\sup_{t\in[0,L],\theta\in\mathcal{N}(\theta_0)} E\left[\left(X^2(t)+Z^TZ\right)\left\{\exp\left(2\beta X(t)+2\gamma^TZ\right)\right\}\right]<\infty.$$

C4. $\sup_{t\in[0,L]}E\left\{\sigma_X^2(\widehat{X}_i^*(t))\right\}<\infty.$

C5. The matrix

$$\Gamma(\theta;Y^*,N^*,X,Z)=\int_0^L\left\{\frac{G_2(t,\theta;Y^*,N^*,X,Z)}{G_0(t,\theta;Y^*,N^*,X,Z)}-\frac{G_1^{\otimes 2}(t,\theta;Y^*,N^*,X,Z)}{G_0^2(t,\theta;Y^*,N^*,X,Z)}\right\}dE\left\{N^*(t)\right\}$$

is positive definite.

A.2. Proof of Theorem 1

First consider existence and consistency. Note that $\widehat{U}^{cr*}(\theta;Y^*,N^*,\widehat{X}^*,Z)$ can be rewritten as

$$\widehat{U}^{cr*}(\theta;Y^*,N^*,\widehat{X}^*,Z)=\widehat{E}\left[\int_0^L(\widehat{X}^*(t),Z^T)^TdN^*(t)\right]-\int\frac{\widehat{G}_1^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)}{\widehat{G}_0^{cr}(t,\theta;Y^*,\widehat{X}^*,Z)}d\widehat{E}\left\{N^*(t)\right\},\qquad\text{(A.1)}$$

where $\widehat{E}$ is the operator for empirical average such that $\widehat{E}(a)=n^{-1}\sum_{i=1}^n a_i$. By the extended strong law of large number (Appendix III, Andersen and Gill (1982)), under conditions C2-C4, the empirical processes in (A.1) converge almost surely to their limits uniformly for $t\in[0,L],\theta\in\mathcal{N}(\theta_0)$. By the chain law, $\widehat{U}^{cr*}(\theta;Y^*,N^*,\widehat{X}^*,Z)$ converges uniformly almost surely for $\theta\in\mathcal{N}(\theta_0)$ to

$$U^{cr*}(\theta;Y^*,N^*,\widehat{X}^*,Z)=E\left[\int_0^L(\widehat{X}^*(t),Z^T)^TdN^*(t)\right]-\int_0^L\left[\frac{G_1^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z)}{G_0^{cr}(t,\theta;Y^*,\widehat{X}^*,Z)}dE\left\{N^*(t)\right\}\right].$$

By iterated expectations, it can be shown that

$$E\left\{\int_0^L(\widehat{X}^*(t),Z_i^T)^TdN^*(t)\right\}=E\left[\int_0^L(X(t),Z_i^T)^TdN_i^*(t)\right],$$
$$G_r^{cr*}(t,\theta;Y^*,N^*,\widehat{X}^*,Z)=G_r(t,\theta;Y^*,N^*,X,Z)\qquad\text{(A.2)}$$

for $r=0,1$. Note that $G_0^{cr*}(t,\theta;Y^*,N^*,\widehat{X}^*,Z)=G_0^{cr}(t,\theta;Y^*,N^*,\widehat{X}^*,Z)$ is bounded away from zero. By counting process theory, under condition C1,

$$E\left[\int_0^L (X(t), Z_i^T)^{Tr} dN_i^*(t)\right] = \int_0^L G_r^{cr*}(t, \theta_0; Y^*, \widehat{X}^*, Z) d\Lambda_0(t)$$

for $r = 0, 1$. It follows that $U^{cr*}(\theta_0; V, \Delta, \widehat{X}^*) = 0$. With similar arguments, it can be shown that $\partial\widehat{U}^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z)/\partial\theta^T$ converges uniformly almost surely to $-\Gamma^*(\theta; Y^*, N^*, \widehat{X}^*, Z)$. With some algebra, it can be shown that $\Gamma^*(\theta; Y^*, N^*, \widehat{X}^*, Z) = \Gamma(\theta; Y^*, N^*, X, Z)$. Under condition C5, θ_0 is the unique zero crossing of $U^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z)$ in a neighborhood of θ_0. The existence and consistency of $\widehat{\theta}^*$ then follows.

Next, we show the asymptotic normality. By a Taylor expansion of $U^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z)$ at θ_0,

$$\begin{aligned} 0 &= U^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z) \\ &= \widehat{U}^{cr*}(\theta_0; Y^*, N^*, \widehat{X}^*, Z) + \frac{\partial\widehat{U}^{cr*}(\widetilde{\theta}; Y^*, N^*, \widehat{X}^*, X)}{\partial\theta^T}\left(\widehat{\theta}^* - \theta_0\right), \end{aligned}$$

where $\widetilde{\theta}$ lies between θ_0 and $\widetilde{\theta}^*$. Thus

$$n^{1/2}\left(\widetilde{\theta}^* - \theta_0\right) = \left\{-\frac{\partial\widehat{U}^{cr*}(\widetilde{\theta}; Y^*, N^*, \widehat{X}^*, X)}{\partial\theta^T}\right\}^{-1} n^{1/2}\widehat{U}^{cr*}(\theta_0; Y^*, N^*, \widehat{X}^*, Z).$$

With a functional Taylor expansion, it can be shown that

$$n^{1/2}\widehat{U}^{cr*}(\theta_0; Y^*, N^*, \widehat{X}^*, Z) = n^{-1/2}\sum_{i=1}^n \varphi_i^*(\theta_0; Y^*, N^*, \widehat{X}^*, Z) + o_p(1).$$

By the uniform almost sure convergence of $\partial\widehat{U}^{cr*}(\theta; Y^*, N^*, \widehat{X}^*, Z)/\partial\theta^T$, we have

$$n^{1/2}\left(\widehat{\theta}^* - \theta_0\right) = \{\Gamma(\theta_0; Y^*, N^*, X)\}^{-1} n^{-1/2}\sum_{i=1}^n \varphi_i^*(\theta_0; Y^*, N^*, \widehat{X}^*, Z) + o_p(1). \tag{A.3}$$

The asymptotic normality then follows from the Central Limit Theorem and Slutsky's Theorem.

A.3. Proof of $\Omega^{cr} \geq_{pd} \Omega^{cr*}$

For simplicity, we assume that $I(m \geq q)$ and $I(m(t) \geq q)$ are independent of V, Δ, α, and Z. In parallel to (A.3), we can show

$$n^{1/2}\left(\widehat{\theta} - \beta_0\right) = \{\Gamma(\theta_0; Y, N, X)\}^{-1} n^{-1/2}\sum_{i=1}^n \varphi_i^*(\theta_0; Y, N, \widehat{X}, Z) + o_p(1).$$

From (A.2), it can be seen that, for $r = 0, 1, 2$,

$$G_r^{cr*}(t,\theta;Y^*,\widehat{X}^*,Z) = G_r(t,\theta;Y^I,X)E\{I(m_i \geq q)\},$$
$$G_r^{cr*}(t,\theta;Y,\widehat{X}^*,Z) = G_r(t,\theta;Y^I,X)E\{I(m_i(t) \geq q)\},$$

where $Y^I(t) = I(V \geq t)$. This, together with $N_i^*(t) \geq N_i(t)$, implies

$$\Gamma(\theta;Y^*,N^*,\widehat{X}^*,Z) \geq_{pd} \Gamma(\theta;Y,N,X,Z). \tag{A.4}$$

Note that, conditional on $\alpha_i, u_i, V_i, \Delta_i$, $(\widehat{X}_i(t), \widehat{X}_i^*(t))$ has a joint normal distribution with mean $(X(t), X(t))$ and $cov(\widehat{X}_i(t), \widehat{X}_i^*(t)) = \sigma_X^2(\widehat{X}_i^*(t))$. It follows that

$$\widehat{X}_i(t)|\widehat{X}_i^*(t), \alpha_i, u_i, V_i, \Delta_i \sim N\left(\widehat{X}_i^*(t), \sigma_X^2(\widehat{X}_i(t)) - \sigma_X^2(\widehat{X}_i(t))\right).$$

Thus, with some algebra, we have

$$E\left[H_{r,i}^{cr*}(t,\theta;Y,\widehat{X},Z)|\widehat{X}_i(t),\alpha_i,u_i,V_i,\Delta_i\right] = E\left[H_{r,i}^{cr*}(t,\theta;Y,\widehat{X}^*,Z)\right]$$

for $r = 0, 1$. It follows that $E\{\varphi_i^*(\theta;Y,N,\widehat{X},Z)|V_i,\Delta_i,\alpha_i,u_i,\widehat{\alpha}_i^*\} = \varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)$. This implies

$$cov\left\{\varphi_i^*(\theta;Y,N,\widehat{X},Z), \varphi_i^*(\theta;Y,N,\widehat{X},Z) - \varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\right\} = 0.$$

Hence

$$\begin{aligned} &var\left\{\varphi_i^*(\theta;Y,N,\widehat{X},Z)\right\} \\ &= var\left\{\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\right\} + var\left\{\varphi_i^*(\theta;Y,N,\widehat{X},Z) - \varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\right\} \\ &\geq_{pd} var\left\{\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\right\}. \end{aligned}$$

In addition, it can be easily seen that

$$\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\left[\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z) - \varphi_i^*(\theta;Y^*,N^*,\widehat{X}^*,Z)\right] = 0$$

as $\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z) = 0$ if $I(m_i(t) \geq q) = 0$ and $\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z) = \varphi_i^*(\theta;Y^*,N^*,\widehat{X}^*,Z)$ otherwise. Thus

$$cov\left\{\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z), \varphi_i^*(\theta;Y,N,\widehat{X}^*,Z) - \varphi_i^*(\theta;Y^*,N^*,\widehat{X}^*,Z\right\} = 0.$$

It follows that $var\left\{\varphi_i^*(\theta;Y,N,\widehat{X}^*,Z)\right\} \geq_{pd} var\left\{\varphi_i^*(\theta;Y^*,N^*,\widehat{X}^*,Z)\right\}$. Thus

$$var\left\{\varphi_i^*(\theta;Y,N,\widehat{X},Z)\right\} \geq_{pd} var\left\{\varphi_i^*(\theta;Y^*,N^*,\widehat{X}^*,Z)\right\}. \tag{A.5}$$

Combining (A.4) and (A.5), we have $\Omega^{cr} \geq_{pd} \Omega^{cr*}$.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10**, 1100–1120.

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

Ding, J. and Wang, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64**, 546–556.

Diggle, P. J., Heagerty, P. J., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford, New York.

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.

Hammer, S. M., Katezstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* **335**, 1081–1089.

Hsieh, F., Tseng, Y. and Wang, J. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* **62**, 1037–1043.

Liang, L., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377–384.

Liao, X., Zucker, D. M., Li, Y. and Speigelman, D. (2011). Survival analysis with error-prone time-varying covariates: A risk set calibration approach. *Biometrics* **67**, 50–58.

Lin, H., Scharfstein, D. O. and Rosenheck, D. O. (2004). Analysis of longitudinal data with irregular outcome-dependent follow-up. *Journal of the Royal Statistical Society B* **66**, 791–813.

Liu, L., Huang, X. and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64**, 950–958.

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.

Song, X., Davidian, M. and Tsiatis, A. A. (2002a). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.

Song, X., Davidian, M. and Tsiatis, A. A. (2002b). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* **3**, 511–528.

Song, X. and Wang, C. Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time varying coefficients. *Statistica Sinica* **27**, 3178–3190.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association* **102**, 1397–1406.

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association* **107**, 688–700.

Tapsoba, Jde. D., Lee, S. M. and Wang, C. Y. (2011). Joint modeling of survival time and longitudinal data with subject-specific changepoints in the covariates. *Statistics in Medicine*

30, 232–249.

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.

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.

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

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

Xu, C., Baines, P. D. and Wang, J. L. (2014). Standard error estimation using the em algorithm for the joint modeling of survival and longitudinal data. *Biostatistics* **15**, 731–744.

Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, B.S. Miller Hall, Room 213, Athens, GA 30602, U.S.A.

E-mail: xsong@uga.edu

(Received October 2015; accepted July 2016)