Statistica Sinica **16**(2006), 235-253

CORRECTED SCORE ESTIMATOR FOR JOINT MODELING OF LONGITUDINAL AND FAILURE TIME DATA

C. Y. Wang

Fred Hutchinson Cancer Research Center

Abstract: We consider Cox proportional hazards regression when longitudinal measurements are available. In some applications, one major goal is to estimate the effect of the underlying change of the longitudinal measurements on survival. One general approach considers regression analysis when some covariate variables are the underlying regression coefficients of another random effects model. For each subject, the covariate variables to the primary regression model are not observed, but can be estimated from the observed longitudinal measurements. This set-up is often called joint modeling in the literature, but it can be treated as two-stage modeling. In this paper, a corrected score estimator is investigated. Comparisons are made with a naive estimator, a regression calibration estimator, a risk set regression calibration estimator, and a conditional score estimator. Similar to the conditional score estimator, the corrected score estimator does not need the assumption of an underlying distribution of the random effects for each subject. Under some regularity conditions, the proposed corrected score estimator is shown to be consistent and asymptotically normally distributed. Simulation results under various random effects distributions are presented.

Key words and phrases: Empirical process, estimating equation, measurement error, proportional hazards, random effects.

1. Introduction

In this paper, we are concerned with the regression relationship between survival of a disease outcome and some longitudinal measurements. There are many important applications in biomedical studies. A particular application is to investigate the relationship between longitudinal serum hormone concentrations and cardiovascular diseases. One simple statistical approach to this problem is to consider the long-term average of the concentrations as the covariate for each subject. This approach has the advantage of being simple, but it does not take into consideration measurement times and longitudinal trajectories. If the age interval of interest of a study cohort is between 50 and 70, and if the trajectory of the longitudinal serum hormone concentrations is approximately linear in this interval, then it would be interesting to investigate whether the underlying baseline serum hormone concentration and the rate of change of longitudinal measurements are major risk factors.

A second example is to consider the relationship between breast cancer and longitudinal percents of fat intakes. In the Women's Health Initiative Dietary Modification trial, there were about 48,000 subjects recruited and 40% of them were randomized into the dietary intervention group. Although the original goal of the study was to reduce the intervention group's percent fat intake by 16%, compliance to dietary change is challenging to most study subjects over an average follow-up of 8.5 years. Therefore, subjects in the intervention group often had lower fat intakes in the beginning but had increasing fat intakes longitudinally. On the other hand, many subjects in the control group mildly reduced their fat intakes longitudinally, because during the trial they learned how to improve their health via food intake. In this application, it is interesting to study whether the underlying baseline percent fat intake and the rate of change of longitudinal fat intakes are major risk factors for the onset of breast cancer.

Cox (1972) regression is a common tool for survival data. If measurement error is not of concern, then either time-independent or time-dependent modeling of longitudinal covariate data may be applied. If longitudinal covariates are considered as replicates, then time-independent modeling may be applied. But, if the longitudinal measurements for each subject have a nonconstant trajectory, then time-dependent modeling is often involved. For time-independent modeling, the issue of measurement error in covariates has been well studied in the last two decades (Prentice (1982)). Under this setting, the underlying long-term average is considered as the true covariate and hence measurement error may be either from the measuring process per se, or from the noise process associated with the long-term average. For time-dependent modeling of the longitudinal data, each subject's trajectory is usually unrestricted. However, some methodology issues may arise if measurement error is a concern. Recently, Tsiatis and Davidian (2001) proposed a conditional score estimator when the longitudinal covariate data are linear in a time-dependent setting.

In addition to the models described above, Wang, Wang and Wang (2000) studied some methods under the modeling that the covariates of interest are the regression parameters of a random effects model. For example, if each subject's longitudinal data follow a linear model then the covariate variables may be the intercept and slope parameters of the linear model. A naive estimator for this problem replaces the unobserved covariates, such as the intercept and slope parameters, by their least square estimates. However, this procedure will generally be biased. Wang et al. investigated the regression calibration (RC) estimator which replaces the unobserved parameters of the random effects model by their conditional expectation given the observed longitudinal data. The RC estimator reduces significant biases from the naive estimator under general situations, but it may have unacceptable biases under some situations, such as large relative risk parameters.

In Section 2, we describe how the statistical problem is modeled. A naive estimator and the RC estimator are reviewed, and a risk set regression calibration estimator is also described. In Section 3, we review the conditional score estimator. The corrected score estimator for this problem is proposed in Section 4. Note that in Cox regression with additive measurement error, Nakamura (1992) proposed a corrected score estimator assuming known measurement error. Nakamura's estimator can be easily applied when the numbers of replicates are equal for all study subjects. Buzas (1998), Kong and Gu (1999), Huang and Wang (2000) and Hu and Lin (2002) have provided some further developments on this type of corrected score approach. In this paper, we are concerned with a more complicated longitudinal covariates data structure having subject-specific trajectories. Rather than being treated as replicates, longitudinal data follow a second regression model. Our problem is often called joint modeling in the literature, but it can be treated as estimation in a two-stage regression model. Developing a corrected score estimator is non-trivial because the measuring times and the numbers of observations vary among study subjects. An intensive simulation study is presented in Section 6.

2. Naive and Regression Calibration Estimation

We first introduce notation and modeling for our problem. The first stage is the longitudinal covariates model. Assume that there are n study subjects and for subject i, $i = 1, \ldots, n$, longitudinal measurements W_{ij} are taken at k_i different times t_{ij}. Suppose the $W_{ij} \equiv W_i(t_{ij}), j = 1, \ldots, k_i$, follow a random effects model (Laird and Ware (1982))

$$W_{ij} = \boldsymbol{D}'_{ij}\boldsymbol{X}_i + U_{ij},\ i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{D}_{ij}$ is a $q \times 1$ observed covariate vector, often with the first element being 1, and $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iq})'$ contains the ith set of random coefficients. One special case is the intercept-slope model that $\boldsymbol{D}_{ij} = (1, t_{ij})'$ and $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$, where X_{i1} is the initial exposure level and X_{i2} is the rate of change. Thus (1) is the covariate model which specifies the covariates of interest.

Now we consider the second stage model. Let T_i^0 be the survival time of the ith subject, and C_i be the censoring time. The response consists of observed variables $T_i \equiv \min(T_i^0, C_i)$ and $\delta_i \equiv I(T_i^0 \le C_i)$, where $I(\cdot)$ is the indicator function. Of interest is the relationship between survival time T^0 and covariate vector $\boldsymbol{X}_i$, but T_i^0 is subject to censoring and thus is not fully observed. Assuming that T_i^0 is independent of C_i given $\boldsymbol{X}_i$, the model semiparametrically formulates the cumulative hazard function $\Lambda(\cdot \mid \boldsymbol{X}_i)$ of T_i^0 given $(\boldsymbol{X}_i, \boldsymbol{W}_i)$ as

$$\Lambda(dt \mid \boldsymbol{X}_i, \boldsymbol{W}_i) = \Lambda_0(dt)\exp(\boldsymbol{\beta}'_0\boldsymbol{X}_i), \tag{2}$$

where $\boldsymbol{\beta}_0$ is a vector of parameters of interest and $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function. In (2), we assume that the effect of longitudinal $\boldsymbol{W}_i$ on survival is from the random effects $\boldsymbol{X}_i$. This conditional independence of $\boldsymbol{W}_i$ and T_i given $\boldsymbol{X}_i$ is often called the surrogacy condition. If $\boldsymbol{X}_i$, $i = 1, \ldots, n$ were available, the (normalized) partial-score estimating equation evaluated at time limit τ solves

$$\widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta}, \boldsymbol{X}, \tau) \equiv \widehat{\mathcal{E}} \int_0^\tau \left[\boldsymbol{X} - \frac{\widehat{\mathcal{E}}\{\boldsymbol{X}\exp(\boldsymbol{\beta}'\boldsymbol{X})I(T \geq t)\}}{\widehat{\mathcal{E}}\{\exp(\boldsymbol{\beta}'\boldsymbol{X})I(T \geq t)\}}\right] d\{\delta I(T \leq t)\} = \mathbf{0},$$

where $\widehat{\mathcal{E}}(\boldsymbol{X})$ denotes the empirical average of any random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

To this problem, one intuitive approach is to estimate the underlying $\boldsymbol{X}_i$ using say the least square estimates, and to apply these estimates as the covariates for subject i. However, this is often associated with serious bias. The bias can be easily seen by considering the case $q = 1$, which leads to a classical measurement error problem with $W_{ij} = X_i + U_{ij}$.

To reduce bias, Wang et al. (2000) studied the RC estimator. The RC estimator may be implemented by replacing the unobserved $\boldsymbol{X}_i$ with its conditional expectation given the observed longitudinal measurements, $\boldsymbol{W}_i$. Assume that $\boldsymbol{X}_i$ is multivariate-normal with mean $\boldsymbol{\mu}_x$, and variance-covariance matrix $\boldsymbol{\Sigma}_x$. Also suppose $\boldsymbol{U}_i \equiv (U_{i1}, \ldots, U_{ik_i})'$ follows a multivariate normal distribution with $\mathcal{E}(\boldsymbol{U}_i|\boldsymbol{X}_i) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{U}_i|\boldsymbol{X}_i) = \boldsymbol{\Sigma}_{ui} = \mathrm{diag}(\sigma_u^2, \ldots, \sigma_u^2)$. Let $\boldsymbol{D}_i$ be the ith design matrix where the jth row contains $\boldsymbol{D}'_{ij}$. Then

$$\begin{pmatrix}\boldsymbol{X}_i \\ \boldsymbol{W}_i\end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{D}_i\boldsymbol{\mu}_x\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x\boldsymbol{D}'_i \\ \boldsymbol{D}_i\boldsymbol{\Sigma}_x & \boldsymbol{D}_i\boldsymbol{\Sigma}_x\boldsymbol{D}'_i + \boldsymbol{\Sigma}_{ui}\end{pmatrix}\right\}.$$

As a result, the calibration function can be calculated by noting that

$$\mathcal{E}(\boldsymbol{X}_i|\boldsymbol{W}_i) = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_x\boldsymbol{D}'_i(\boldsymbol{D}_i\boldsymbol{\Sigma}_x\boldsymbol{D}'_i + \boldsymbol{\Sigma}_{ui})^{-1}(\boldsymbol{W}_i - \boldsymbol{D}_i\boldsymbol{\mu}_x). \tag{3}$$

The RC estimator solves $\widehat{\boldsymbol{\Psi}}\{\boldsymbol{\beta}, \mathcal{E}(\boldsymbol{X}|\boldsymbol{W})\} = \mathbf{0}$. This involves estimating the nuisance parameters $\boldsymbol{\mu}_x$, $\boldsymbol{\Sigma}_x$ and σ_u^2. Let $\boldsymbol{X}_i^{(ls)}$ be the least square estimator from (1) for each subject, *i.e.* $\boldsymbol{X}_i^{(ls)} = (\boldsymbol{D}'_i\boldsymbol{D}_i)^{-1}\boldsymbol{D}'_i\boldsymbol{W}_i$. Let $\boldsymbol{\eta}$ be the vector of nuisance parameters, and $\boldsymbol{\Sigma}_{Ri} = (\boldsymbol{D}'_i\boldsymbol{D}_i)^{-1}\boldsymbol{D}'_i\boldsymbol{\Sigma}_{ui}\boldsymbol{D}_i(\boldsymbol{D}'_i\boldsymbol{D}_i)^{-1}$. If $k_i > q$ for $i = 1, \ldots, n$, then one convenient choice of estimating equations $\boldsymbol{\Phi}_N(\boldsymbol{W}_i, \boldsymbol{\eta})$ for nuisance parameters involved in the distribution of $\boldsymbol{X}$, and U is

$$\begin{cases} \displaystyle\sum_{i=1}^n k_i\left\{\boldsymbol{X}_i^{(ls)} - \boldsymbol{\mu}_x\right\} = \mathbf{0}; \\ \displaystyle\sum_{i=1}^n k_i\left[\left\{\boldsymbol{X}_i^{(ls)} - \boldsymbol{\mu}_x\right\}\left\{\boldsymbol{X}_i^{(ls)} - \boldsymbol{\mu}_x\right\}' - \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{Ri}\right] = \mathbf{0}; \\ \displaystyle\sum_{i=1}^n k_i\left[\left\{\boldsymbol{W}_i - \boldsymbol{D}_i\boldsymbol{X}_i^{(ls)}\right\}'\left\{\boldsymbol{W}_i - \boldsymbol{D}_i\boldsymbol{X}_i^{(ls)}\right\} - \sigma_u^2(k_i - q)\right] = \mathbf{0}. \end{cases} \tag{4}$$

The first equation of (4) is for the estimation of $\boldsymbol{\mu}_x$; the second equation of (4) is for the estimation of $\boldsymbol{\Sigma}_x$, and unbiasedness follows because $\mathcal{E}[\{\boldsymbol{X}_i^{(ls)} - \boldsymbol{\mu}_x\}\{\boldsymbol{X}_i^{(ls)} - \boldsymbol{\mu}_x\}'] = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{Ri}$; the last equation of (4) is based on the assumption that U_{ij} is independent of $U_{ij'}$ for $j \neq j'$.

The RC estimator often performs well when the relative risk parameters are moderate, although it is an inconsistent estimator. Wang et al. (2000) presented situations when the RC estimator may have sizable biases. One way to refine this approach is to estimate the calibration function $\mathcal{E}(\boldsymbol{X}_i|\boldsymbol{W}_i)$ within each risk set. In the classical additive measurement error model, Tsiatis, DeGrutolla and Wulfsohn (1995) proposed risk set regression calibration, and Dafni and Tsiatis (1998) further investigated the method in joint modeling. They applied empirical Bayes estimates to estimate the covariate values at each time point at which an event occurs, based on the observed history of the observed longitudinal data among individuals who did not have an event up to that time point. In the additive measurement error model, a risk set regression calibration was investigated by Xie, Wang and Prentice (2001). In the simulation study, a risk set regression calibration estimator is implemented similar to the idea given above, and at each event time the calibration function $\mathcal{E}(\boldsymbol{X}_i|\boldsymbol{W}_i)$ is approximated by (3) and (4), using just the observed covariate history among individuals who are at risk.

3. Conditional Score Estimation

Motivated by the conditional score estimation of Stefanski and Carroll (1987) for the generalized linear model, Tsiatis and Davidian (2001) proposed a conditional score estimation in Cox regression when the underlying time dependent covariates follow a linear model. The conditional score estimator does not need the distributional assumption on $\boldsymbol{X}$, but the model assumption on $\boldsymbol{U}$ is needed. This modeling of covariates is slightly different from our modeling that has the covariates as random effects coefficients. Let $\boldsymbol{X}_i^{(ls)}(u)$ be the least square estimator from (1) using observations by time u. Let $N_i(u) = I[\delta_i = 1, T_i \leq u, t_{iq} \leq u]$ be the counting process and $Y_i(u) = I[T_i \geq u, t_{iq} \leq u]$ be the at risk process. The conditional score estimation for more general modeling of covariates was further investigated in Song, Davidian and Tsiatis (2002). In our problem, they showed that conditional on $Y_i(u) = 1$, $\mathbf{Q}_i(u, \boldsymbol{\beta}) = \boldsymbol{X}_i^{(ls)}(u) + \boldsymbol{\Sigma}_{Ri}(u)\boldsymbol{\beta} dN_i(u)$ is a complete sufficient statistic for $\boldsymbol{X}_i$, where $\Sigma_{Ri}(u)$ is the variance of $\boldsymbol{X}_i^{(ls)}(u)$, see (6). Hence, at each time u, conditioning on $\mathbf{Q}_i(u, \boldsymbol{\beta})$ would remove the dependence of the conditional distribution on the random effect $\boldsymbol{X}_i$. They showed that the conditional intensity process $\lim_{du \to 0} \mathrm{pr}\{dN_i(u) = 1|\mathbf{Q}_i(u, \boldsymbol{\beta}), Y_i(u)\}$ is equal to $\lambda_0(u)\exp\{\boldsymbol{\beta}'\mathbf{Q}_i(u, \boldsymbol{\beta}) - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{Ri}(u)\boldsymbol{\beta}/2\}Y_i(u)$. As a result, the conditional score estimator solves

$$\sum_{i=1}^{n}\int_{0}^{\tau}\left\{\mathbf{Q}_i(u,\boldsymbol{\beta})-\frac{\boldsymbol{V}_i(u,\boldsymbol{\beta})}{V_0(u,\boldsymbol{\beta})}\right\}dN_i(u)=0,$$

where $\boldsymbol{V}_m(u,\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} Y_i(u)\mathbf{Q}_i^m(u,\boldsymbol{\beta})\exp\{\boldsymbol{\beta}'\mathbf{Q}_i(u,\boldsymbol{\beta}) - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{Ri}(u)\boldsymbol{\beta}/2\}$, $m = 0,1$. The asymptotic variance of the conditional score estimator can be obtained by a sandwich estimator. Details are given in Tsiatis and Davidian (2001) and Song, Davidian and Tsiatis (2002).

4. Corrected Score Estimation

Let $\tilde{\boldsymbol{X}} = (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$ for any vector $\boldsymbol{X}$. Generally speaking, if $\widehat{\boldsymbol{\Psi}}(\tilde{\delta},\tilde{T},\tilde{\boldsymbol{X}})$ is a full data score such that $\mathcal{E}\{\widehat{\boldsymbol{\Psi}}(\tilde{\delta},\tilde{T},\tilde{\boldsymbol{X}})\} = 0$, then $\widehat{\boldsymbol{\Psi}}_c(\tilde{\delta},\tilde{T},\tilde{\boldsymbol{W}})$ is a *corrected score* if $\mathcal{E}\{\widehat{\boldsymbol{\Psi}}_c(\tilde{\delta},\tilde{T},\tilde{\boldsymbol{W}})|\tilde{T},\tilde{\delta},\tilde{\boldsymbol{X}}\} = \widehat{\boldsymbol{\Psi}}(\tilde{T},\tilde{\delta},\tilde{\boldsymbol{X}})$. A corrected score estimator does not need the distributional assumption on $\boldsymbol{X}$, but the model assumption on $\boldsymbol{U}$ is needed. In our problem, $\widehat{\boldsymbol{\Psi}}(\tilde{\delta},\tilde{T},\tilde{\boldsymbol{X}})$ is the partial likelihood score, and we use $\widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta},\boldsymbol{X})$ for notational convenience. Let

$$\widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta},\boldsymbol{X},t) = n^{-1}\sum_{i=1}^{n}\int_{0}^{t}\{\boldsymbol{X}_i - \mathbf{E}(\boldsymbol{\beta},\boldsymbol{X},s)\}dN_i(s) = \mathbf{0}, \tag{5}$$

where $\mathbf{E}(\boldsymbol{\beta},\boldsymbol{X},t) = \boldsymbol{S}^{(1)}(\boldsymbol{\beta},\boldsymbol{X},t)/S^{(0)}(\boldsymbol{\beta},\boldsymbol{X},t)$, $\boldsymbol{S}^{(m)}(\boldsymbol{\beta},\boldsymbol{X},t) = n^{-1}\sum_{j=1}^{n} Y_j(t)$ $\boldsymbol{X}_j^m\exp(\boldsymbol{\beta}'\boldsymbol{X}_j), m = 0,1$. Then the partial likelihood score without measurement error can be written as $\widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta},\boldsymbol{X},\tau)$. It is easily seen that $\boldsymbol{\Psi}(\boldsymbol{\beta},\boldsymbol{X},\tau)$ converges in probability to

$$\boldsymbol{\Psi}(\boldsymbol{\beta},\tau) = \mathcal{E}\Big[\int_{0}^{\tau}\{\boldsymbol{X} - e(\boldsymbol{\beta},t)\}dN(t)\Big],$$

where $e(\boldsymbol{\beta},t) = \boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)/s^{(0)}(\boldsymbol{\beta},t)$, $\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t) = \mathcal{E}\{\boldsymbol{X}\exp(\boldsymbol{\beta}'\boldsymbol{X})Y(t)\}$, $s^{(0)}(\boldsymbol{\beta},t) = \mathcal{E}\{\exp(\boldsymbol{\beta}'\boldsymbol{X})Y(t)\}$. The limit can also be written as $\mathcal{E}[\int_0^\tau\{\boldsymbol{X} - e(\boldsymbol{\beta},t)\}dM(t)]$, where $M(t) = N(t) - \int_0^t Y(s)\exp(\boldsymbol{\beta}_0'\boldsymbol{X})\lambda_0(s)ds$. There is a unique root of $\boldsymbol{\Psi}(\boldsymbol{\beta},\tau) = 0$. Roughly speaking, a corrected score $\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta},\boldsymbol{W},\tau)$ provides a valid estimating function because $\mathcal{E}\{\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta},\boldsymbol{W},\tau)|\tilde{T},\tilde{\delta},\tilde{\boldsymbol{X}}\}$ converges to $\boldsymbol{\Psi}(\boldsymbol{\beta},\tau)$ in probability.

4.1. First-order estimator

Recall that $\boldsymbol{X}_i^{(ls)}(t) = (\boldsymbol{D}_i'(t)\boldsymbol{D}_i(t))^{-1}\boldsymbol{D}_i'(t)\boldsymbol{W}_i(t)$, where $\boldsymbol{D}_i(t)$ and $\boldsymbol{W}_i(t)$ are the ith design matrix and observed covariates described in Section 2, except using data points up to time t. As described, the main step of our estimator is to seek an estimating function based on the observed data such that its conditional expectation given the full data is the same as the original partial likelihood score. Based on $\boldsymbol{X}_i^{(ls)}(u)$, at a time t, we now formulate an induced measurement error model as

$$\begin{cases} \boldsymbol{X}_i^{(ls)}(t) = \boldsymbol{X}_i(t) + \boldsymbol{R}_i(t), \\ \mathcal{E}\{\boldsymbol{R}_i(t)|\boldsymbol{X}_i(t)\} = 0; \\ \text{cov}\{\boldsymbol{R}_i(t)|\boldsymbol{X}_i(t)\} = \boldsymbol{\Sigma}_{Ri}(t) \\ \qquad\qquad = \{\boldsymbol{D}_i'(t)\boldsymbol{D}_i(t)\}^{-1}\boldsymbol{D}_i'(t)\boldsymbol{\Sigma}_{ui}(t)\boldsymbol{D}_i(t)\{\boldsymbol{D}_i'(t)\boldsymbol{D}_i(t)\}^{-1}. \end{cases} \tag{6}$$

Note that in (5), by (6), $\mathcal{E}(\boldsymbol{X}_i^{(ls)}|\boldsymbol{X}_i) = \boldsymbol{X}_i$. Hence, what is left to carry out a corrected score is to calculate $\mathcal{E}\{\mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{T}, \tilde{\delta}, \tilde{\boldsymbol{X}}\}$ for a time point t. To resolve some technical issues, we treat the number of longitudinal data for subject i, k_i, as a random variable, writing K_i for k_i. For given $K_i = k_i$, we also treat the measuring times, $t_{ij}, j = 1, \ldots, k_i$ as random, and hence $\boldsymbol{\Sigma}_{Ri}(t)$ is random as well. This assumption is not restrictive. The method can be applied even if k_i's are fixed but not the same for all subjects. Under this situation, we may consider these fixed k_i's as being sampled from an unknown underlying distribution.

By direct calculation, under Condition (A2) of the Appendix, as $n \to \infty$,

$$\mathcal{E}\{\mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{T}, \tilde{\delta}, \tilde{\boldsymbol{X}}\} = \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}, t) + \frac{\widehat{\mathcal{E}}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}}{\widehat{\mathcal{E}}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}} + O_p(n^{-1}).$$

Therefore, a first order corrected score (CS) estimator is to solve

$$\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}, \boldsymbol{W}, \tau) \equiv n^{-1}\sum_{i=1}^{n}\int_0^{\tau}[\boldsymbol{X}_i^{(ls)}(t) - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t) + \mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\}]dN_i(t) = \mathbf{0}, \tag{7}$$

where $\mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\} = [\sum_{j=1}^n \boldsymbol{\beta}'\boldsymbol{\Sigma}_{Rj}\exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{Rj}(t)\boldsymbol{\beta}/2\}]/[\sum_{j=1}^n \exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{Rj}(t)\boldsymbol{\beta}/2\}]$.

Hence, the first-order conditional score estimating function $\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}, \boldsymbol{W})$ has the same limit, $\boldsymbol{\Psi}(\boldsymbol{\beta})$, as that from the estimating score with $\boldsymbol{X}$ available, *i.e.*, $\tilde{\Psi}(\boldsymbol{\beta}, \boldsymbol{X})$.

Theorem 1. *Under Conditions* (A1)$-$(A6), *let* $\tilde{\boldsymbol{\beta}}$ *be a solution to* (7). *Then* $\tilde{\boldsymbol{\beta}}$ *exists and is unique in a neighborhood of* $\boldsymbol{\beta}_0$ *with probability converging to one as* $n \to \infty$, *and* $\tilde{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ *in probability. In addition,* $n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *is asymptotically normally distributed with mean* $\mathbf{0}$ *and variance given in* (10).

4.2. Second-order estimator

In order to have better finite sample performance, we consider a second order approximation. It can be shown by some calculations given in the Appendix that as $n \to \infty$, and $\beta \to 0$,

$$\begin{aligned}&\mathcal{E}\{\mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{T}, \tilde{\delta}, \tilde{\boldsymbol{X}}\}\\ &= \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}, t) + \mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\}\Big[1 - \mathcal{D}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\}\frac{n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2}\Big] + o_p(n^{-1}\boldsymbol{\beta}),\end{aligned}$$

where $\mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\} = \mathcal{E}[\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}/2\}]/\mathcal{E}[\exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}/2\}]$ and $\mathcal{D}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\} = \mathcal{E}[\exp\{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}\}]/\big(\mathcal{E}[\exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}/2\}]\big)^2$. Note that

$$\mathcal{E}\Big[\frac{n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)\}^2}|\tilde{T}, \tilde{\delta}, \tilde{\boldsymbol{X}}\Big] = \mathcal{D}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}(t)\}\frac{n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2} + o_p(n^{-1}). \quad (8)$$

Therefore, the second-order CS estimator solves

$$\sum_{i=1}^{n}\int_0^t \Big[\boldsymbol{X}_i^{(ls)}(u) - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, u) + \mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(u)\}\Big[1 - \frac{n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)\}^2}\Big]\Big] dN_i(u) = \mathbf{0}. \quad (9)$$

Because the difference between the first-order estimating score and the second-order estimating score is of order $O_p(n^{-1})$, the improvement of the second-order estimator is primarily on finite sample performance. Any further correction using a higher order does not seem appealing in terms of calculation. Hence, we refer the CS estimator to the second-order estimator hereafter.

4.3. Covariance estimation

The implementation of the CS estimator can be done by Newton-Raphson iteration. A simple sandwich estimator can be applied to estimate the standard error of the CS estimator. Let $\boldsymbol{\Theta}_0 = (\boldsymbol{\beta}_0', \sigma_{u0}^2)'$, where σ_{u0}^2 is the true nuisance parameter involved in $\boldsymbol{\Sigma}_R$ as described in Section 2. Define $\boldsymbol{\Phi}(\boldsymbol{\Theta}, \boldsymbol{W}_i) = [\{\boldsymbol{\Phi}_P(\boldsymbol{\Theta}, \boldsymbol{W}_i)\}', \{\boldsymbol{\Phi}_N(\sigma_u^2, \boldsymbol{W}_i)\}']'$, where $\boldsymbol{\Phi}_P(\boldsymbol{\Theta}, \boldsymbol{W}_i)$ is given in (12) in the Appendix, and $\boldsymbol{\Phi}_N(\sigma_u^2, \boldsymbol{W}_i)$ is given in the last equation of (4). Denote the solution of $\sum_{i=1}^n \boldsymbol{\Phi}(\boldsymbol{\Theta}, \boldsymbol{W}_i) = \mathbf{0}$ by $\widehat{\boldsymbol{\Theta}}$ and let $\boldsymbol{G}_n(\boldsymbol{\Theta}) = n^{-1}\sum_{i=1}^n (\partial/\partial\boldsymbol{\Theta})\boldsymbol{\Phi}(\boldsymbol{W}_i, \boldsymbol{\Theta})$. Then the variance of $n^{1/2}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$ can be estimated by the robust sandwich estimator

$$\boldsymbol{G}_n^{-1}(\widehat{\boldsymbol{\Theta}})\Big[n^{-1}\sum_{i=1}^n\{\boldsymbol{\Phi}(\widehat{\boldsymbol{\Theta}}, \boldsymbol{W}_i)\}^{\otimes 2}\Big]\{\boldsymbol{G}_n^{-1}(\widehat{\boldsymbol{\Theta}})\}', \quad (10)$$

where $\mathbf{a}^{\otimes 2}$ denotes $\mathbf{a}\mathbf{a}'$ for any vector $\mathbf{a}$. Numerical derivatives can be easily applied to calculate $\boldsymbol{G}(\Theta)$. Under the conditions for the consistency result described above, the robust sandwich estimator can be shown to be consistent.

5. Estimation of Baseline Cumulative Hazard Function

Estimation of the baseline cumulative hazard function $\Lambda_0(\cdot)$ is often of interest. Generally, after a method for regression coefficients is applied, the estimation of baseline cumulative hazard function can be developed by modifying the Breslow estimator of Λ_0, which can be written as

$$\tilde{\Lambda}_0(t;\boldsymbol{\beta},\boldsymbol{X}) = \int_0^t \frac{d\widehat{\mathcal{E}}\{\delta I[T \le s]\}}{\widehat{\mathcal{E}}\{\exp(\boldsymbol{\beta}'\boldsymbol{X})I[T \ge s]\}}.$$

Of course, the above estimator can not be applied directly since $\boldsymbol{X}$ is not available, and replacing $\boldsymbol{X}$ by $\boldsymbol{X}^{(ls)}$ may lead to bias. It is easily seen that the above estimator is consistent since it is a functional of the empirical processes with limit

$$\Lambda_0(t) = \int_0^t \frac{d\mathcal{E}\{\delta I[T \le s]\}}{\mathcal{E}\{\exp(\boldsymbol{\beta}'\boldsymbol{X})I[T \ge s]\}}.$$

By simple calculation, it can be seen that $\{S^{(0)}(\boldsymbol{\beta},\boldsymbol{X},t)\}^{-1} = \mathcal{E}[\exp\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}/2\}]$ $\mathcal{E}[\{S^{(0)}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)\}^{-1}|\tilde{\boldsymbol{X}},\tilde{T},\tilde{\delta}\}] + o_p(1)$. Therefore, a consistent estimator for $\Lambda_0(\cdot)$ can be obtained as

$$\widehat{\Lambda}_0(t;\boldsymbol{\beta},\boldsymbol{X}^{(ls)}) = \int_0^t \frac{\widehat{\mathcal{E}}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(s)}\frac{\boldsymbol{\beta}}{2}}\}d\widehat{\mathcal{E}}\{\delta I[T \le s]\}}{\widehat{\mathcal{E}}\{\exp(\boldsymbol{\beta}'\boldsymbol{X}^{(ls)})I[T \ge s]\}} = \int_0^t \frac{\widehat{\mathcal{E}}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(s)}\frac{\boldsymbol{\beta}}{2}}\}d\overline{N}(s)}{S^{(0)}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},s)},$$

where $\overline{N} = \sum_{i=1}^n N_i/n$. The asymptotic distribution of the estimator can be established by some theory on empirical processes. For the classical additive measurement error model, Huang and Wang (2000) proposed similar estimators for Λ_0 and showed that they converge weakly to a zero mean Gaussian process. The above estimator's distribution theory can be obtained likewise, except with more complicated calculations.

6. Simulation Study

This section provides some results from a Monte-Carlo study that compares several methods. Various distributional assumptions for $\boldsymbol{X}_i$ are considered. In Table 1, $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ were generated from a bivariate normal distribution with mean $\boldsymbol{\mu}_x = (0,0)'$, $\mathrm{var}(X_{i1}) = 1$, $\mathrm{var}(X_{i2}) = 0.25$, $\mathrm{corr}(X_{i1}, X_{i2}) = -0.1$, $\sigma_u = 0.4$. Six repeated measurements W_{ij} were simulated from the model $W_{ij} = X_{i1} + X_{i2}t_{ij} + U_{ij}$, where the t_{ij} were uniformly distributed in $[0.5j-2, 0.5j-1.9]$, for $j = 1,\ldots,6$. Failure times were generated by the hazard function $\lambda(t;\boldsymbol{X}_i) = 0.2\exp(\boldsymbol{\beta}'\boldsymbol{X}_i)$. A common censoring time was used for all subjects such that the censoring percentage was 50%, and the random error process satisfied $\boldsymbol{\Sigma}_{ui} = \sigma_u^2\mathbf{I}_6$. The observed longitudinal data were available only before the event time. The true parameters and sample sizes n used in each simulation study are shown in the corresponding tables. In the tables, the "biases" were calculated by taking the average of $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}$ from 200 replicates, "SD" denotes the sample standard deviation of the estimators, "mean(SE)" denotes the average of the estimated standard errors of the estimators. The 95% confidence interval coverage probabilities and mean square errors (MSEs) are also included.

Table 1. Simulation results when (X_1, X_2) is bivariate normal.

	$n = 400$					$n = 800$				
	Naive	RC	RRC	COR-S	CON-S	Naive	RC	RRC	COR-S	CON-S
	$\boldsymbol{\beta} = \ln(2), -\ln(2)$									
β_1										
bias	-0.015	-0.003	0.001	0.003	0.003	-0.017	-0.007	-0.003	0.000	0.000
SD	0.076	0.078	0.079	0.080	0.080	0.056	0.057	0.060	0.062	0.061
mean(SE)	0.087	0.077	0.079	0.080	0.079	0.053	0.055	0.057	0.059	0.058
95% cov.	0.930	0.935	0.945	0.945	0.945	0.945	0.940	0.930	0.935	0.935
MSE	0.006	0.006	0.006	0.006	0.006	0.003	0.003	0.004	0.004	0.004
β_2										
bias	0.078	0.008	0.004	0.002	0.003	0.069	-0.007	-0.010	-0.003	0.004
SD	0.130	0.150	0.158	0.165	0.162	0.094	0.116	0.119	0.123	0.121
mean(SE)	0.136	0.155	0.161	0.167	0.163	0.093	0.119	0.121	0.131	0.126
95% cov.	0.920	0.965	0.965	0.965	0.955	0.880	0.965	0.950	0.950	0.955
MSE	0.023	0.023	0.025	0.027	0.026	0.014	0.014	0.014	0.015	0.015
	$\boldsymbol{\beta} = (\ln(5), -\ln(5))$									
β_1										
bias	-0.098	-0.060	-0.049	0.009	0.007	-0.108	-0.080	-0.065	0.004	0.002
SD	0.098	0.102	0.104	0.116	0.115	0.066	0.071	0.076	0.082	0.081
mean(SE)	0.099	0.102	0.103	0.117	0.123	0.069	0.074	0.077	0.091	0.087
95% cov.	0.830	0.880	0.915	0.950	0.980	0.635	0.790	0.895	0.960	0.960
MSE	0.019	0.014	0.013	0.013	0.013	0.016	0.011	0.010	0.007	0.007
β_2										
bias	0.229	0.072	0.052	-0.013	-0.010	0.212	0.068	0.050	-0.000	0.003
SD	0.128	0.144	0.163	0.198	0.192	0.101	0.132	0.140	0.160	0.154
mean(SE)	0.148	0.171	0.175	0.195	0.192	0.097	0.132	0.145	0.150	0.158
95% cov.	0.670	0.950	0.955	0.975	0.980	0.405	0.915	0.920	0.955	0.950
MSE	0.070	0.026	0.029	0.039	0.037	0.055	0.022	0.022	0.026	0.024

NOTE: Nuisance parameters are $\boldsymbol{\mu}_x = (0,0)'$, $\mathrm{var}(X_1) = 1$, $\mathrm{var}(X_2) = 0.25$, $\mathrm{corr}(X_1, X_2) = -0.1$, $\sigma_u = 0.4$. RC, RRC, COR-S and CON-S denote the regression calibration estimator, risk set regression calibration estimator, corrected score estimator and conditional score estimator, respectively. For $j = 1, \ldots, 6$, the t_{ij} are uniformly distributed within $[0.5j - 2, 0.5j - 1.9]$. The censoring rate is 0.5.

The results from the upper portion of Table 1 show that the RC, risk set regression calibration (RRC), corrected score and conditional score estimators perform reasonably well when the relative risk is moderate. Note that to show the inconsistency problem of the RC estimator and RRC estimators under this moderate risk setting, we may increase the total sample size n to say 5,000, in which case the coverage probabilities for the 95% confidence intervals of the RC estimates may be less than 90%. The lower portion of Table 1 shows that the RC estimator may have bias when the relative risk is large. It is seen that the RRC estimator has very good finite sample performance in most cases, except for a minor bias problem when $\boldsymbol{\beta} = (\ln(5), -\ln(5))$. The corrected score estimates presented here are the second order ones. The first order corrected score estimator was examined, and it was generally not as good as the second order corrected score estimator in terms of both bias and efficiency. The conditional score estimator is generally slightly better than the corrected score estimator, but the difference is small, especially with small relative risk parameters, small measurement errors and large sample sizes.

Table 2 considers a non-normal model for the distribution of $\boldsymbol{X}_i$. The actual X_{i1} was generated from a mixture of two normals with means $(1/5^{1/2}, -2/5^{1/2})$, variances $(4/5, 1/5)$, and the mixture percentages were (2/3, 1/3). Under this mechanism, X_{i1} has mean 0 and variance 1, X_{i2} was generated similarly to that of X_{i1} but with variance 0.25, and $\mathrm{corr}(X_{i2}, X_{i1}) = -0.1$. The performance of the various methods are quite similar to those reported in Table 1 for moderate relative risk. When the relative risk parameters are large, such as $(\ln(5), -\ln(5))$, generally the biases and standard errors are slightly larger than those in Table 1. Although not presented in the tables, some other random effects models for X_{i1} and X_{i2} have been examined. For uniform distributions, the results are rather similar to Table 1. For chi-square distributions, which are more skewed, the results are more similar to Table 2, but the biases and standard errors for most estimates are bigger. Another fact not seen in the tables is that when the measurement error σ_u is increased, the biases and standard errors of various estimators will generally increase. The effect of increasing σ_u is rather similar to that of increasing $|\boldsymbol{\beta}|$, except that increasing σ_u may lead to more divergences in many estimation procedures.

Overall, the proposed corrected score estimator is rather similar to, but slightly less efficient than the conditional score estimator. The RC estimator and RRC estimator are both generally very efficient. The biases are generally small except when $\boldsymbol{\beta} = (\ln(5), -\ln(5))$, under which they may have coverage probabilities of less than 90%. Increasing the sample size will reduce the coverage probabilities of the RC and RRC estimators. This phenomenon was confirmed in Xie et al. (2001) in the case of an additive covariate measurement error model,

Table 2. Simulation results when X_1 and X_2 are independently sampled from two mixtures of normals.

	$n = 400$					$n = 800$				
	Naive	RC	RRC	COR-S	CON-S	Naive	RC	RRC	COR-S	CON-S
	$\boldsymbol{\beta} = \ln(2), -\ln(2)$									
β_1										
bias	-0.016	-0.006	0.001	-0.001	-0.001	-0.017	-0.007	-0.001	-0.003	-0.002
SD	0.074	0.075	0.078	0.080	0.078	0.051	0.052	0.054	0.055	0.054
mean(SE)	0.073	0.075	0.076	0.079	0.078	0.051	0.052	0.054	0.055	0.054
95% cov.	0.940	0.930	0.935	0.945	0.950	0.940	0.950	0.945	0.945	0.960
MSE	0.006	0.006	0.006	0.006	0.006	0.003	0.003	0.003	0.003	0.003
β_2										
bias	0.082	-0.009	0.012	-0.002	0.001	0.072	-0.002	-0.001	-0.014	-0.012
SD	0.155	0.169	0.183	0.195	0.194	0.101	0.112	0.119	0.130	0.126
mean(SE)	0.138	0.156	0.166	0.189	0.179	0.097	0.110	0.117	0.133	0.126
95% cov.	0.885	0.915	0.920	0.925	0.930	0.875	0.930	0.930	0.940	0.950
MSE	0.031	0.028	0.033	0.038	0.037	0.015	0.012	0.014	0.017	0.016
	$\boldsymbol{\beta} = (\ln(5), -\ln(5))$									
β_1										
bias	-0.085	-0.058	-0.025	0.020	0.018	-0.095	-0.069	-0.038	0.005	0.004
SD	0.106	0.109	0.114	0.128	0.127	0.075	0.078	0.081	0.091	0.090
mean(SE)	0.095	0.099	0.101	0.120	0.120	0.067	0.070	0.071	0.081	0.085
95% cov.	0.800	0.865	0.910	0.930	0.950	0.680	0.790	0.840	0.940	0.965
MSE	0.018	0.015	0.014	0.017	0.016	0.015	0.011	0.008	0.008	0.008
β_2										
bias	0.238	0.067	0.078	-0.018	-0.012	0.225	0.055	0.060	-0.026	-0.024
SD	0.154	0.178	0.189	0.250	0.235	0.104	0.122	0.130	0.168	0.160
mean(SE)	0.147	0.175	0.183	0.250	0.243	0.103	0.124	0.129	0.171	0.172
95% cov.	0.660	0.890	0.895	0.960	0.960	0.435	0.915	0.910	0.950	0.960
MSE	0.080	0.036	0.042	0.063	0.055	0.061	0.018	0.021	0.029	0.026

NOTE: Nuisance parameters are $\boldsymbol{\mu}_x = (0,0)'$, $\text{var}(X_1) = 1$, $\text{var}(X_2) = 0.25$, $\text{corr}(X_1, X_2) = -0.1$, $\sigma_u = 0.4$. RC, RRC, COR-S and CON-S denote the regression calibration estimator, risk set regression calibration estimator, corrected score estimator and conditional score estimator, respectively. For $j = 1, \ldots, 6$, the t_{ij} are uniformly distributed within $[0.5j - 2, 0.5j - 1.9]$. The censoring rate is 0.5.

which can be treated as a special case of the joint modeling in this paper. Note that MSE is used as a reference here and it serves as a good criteria to compare differences between the conditional score estimator and the corrected score estimator. MSE may not be a good criteria to compare consistent estimators with inconsistent estimators. For example, in situations with small measurement errors the naive estimator may have the smallest MSE. In covariate measurement error, there is a trade-off between bias and efficiency and this is also seen in joint modeling. That is, the corrected score and conditional score estimators both reduce the bias, but at a cost of increasing the variance.

Acknowledgements

This research was supported by US National Institutes of Health grants: AG 15026, CA 53996. I am thankful to Dr. Xiao Song, the Editor and two anonymous referees for helpful comments.

Appendix

Appendix A. Regularity Conditions

(A1) $\{N_i, Y_i, \boldsymbol{X}_i\}$, $i = 1, \ldots, n$, are independent and identically distributed.
(A2) $\Lambda_0(t)$ is continuous and $\Lambda_0(\tau) < \infty$.
(A3) Almost surely $\boldsymbol{X}$ is bounded.
(A4) $\mathrm{pr}\{Y(t) = 1, \forall t \in [0, \tau]\} > 0$.
(A5) The number of measurements K_i for $i = 1, \ldots, n$, and t_{ij}, $j = 1, \ldots, K_i$, are random variables and $\mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\}$ has bounded variation.
(A6) $\int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}, t) s^{(0)}(\boldsymbol{\beta}, t)\lambda_0(t)dt$ is positive definite, where $\boldsymbol{v}(\boldsymbol{\beta}, t) = \boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)/ s^{(0)}(\boldsymbol{\beta}, t) - \{e(\boldsymbol{\beta}, t)\}^2$, $\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t) = \mathcal{E}\{Y(t)\boldsymbol{X}^{\otimes 2}\exp(\boldsymbol{\beta}'\boldsymbol{X})\}$.

Appendix B. Proof of Theorem 1

The corrected score (7) can be rewritten as

$$\begin{aligned}
&\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}, \boldsymbol{W}, \tau) \\
&= n^{-1}\sum_{i=1}^{n}\int_0^\tau \{\boldsymbol{X}_i^{(ls)} - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, s) + \mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(s)\}\}dM_i(s) \\
&\quad + n^{-1}\sum_{i=1}^{n}\int_0^\tau \{\boldsymbol{X}_i^{(ls)} - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, s) + \mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(s)\}\}Y_i(s)e^{\boldsymbol{\beta}'\boldsymbol{X}_i}\lambda_0(s)ds. \quad (11)
\end{aligned}$$

Let $\boldsymbol{S}_*^{(m)}(\boldsymbol{\beta}, \boldsymbol{X}, t) = \mathcal{E}\{\boldsymbol{S}^{(m)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\}$. By noting (6), it is straightforward that

$$\begin{aligned}
S_*^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t) &= S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}, \\
\boldsymbol{S}_*^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t) &= \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\} \\
&\quad + S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}.
\end{aligned}$$

Under Conditions (A1)-(A4), as in Andersen and Gill (1982, Theorem III.1), $\sup_{t,\boldsymbol{\beta}}\|S^{(m)}(\boldsymbol{\beta},\boldsymbol{X},t)-s^{(m)}(\boldsymbol{\beta},t)\| \xrightarrow{p} 0$. Hence, it follows that $\sup_{t,\boldsymbol{\beta}}\|E(\boldsymbol{\beta},\boldsymbol{X},t)-e(\boldsymbol{\beta},t)\| \xrightarrow{p} 0$, where $e(\boldsymbol{\beta},t)$ was defined in Section 4. Under Conditions (A1)-(A5), since the class of functions of bounded variation is Glivenko-Cantelli (van der Vaar and Wellner (1996, Chap. 2.10)), $\sup_{t,\boldsymbol{\beta}}\|\mathbf{E}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)-\mathbf{E}(\boldsymbol{\beta},\boldsymbol{X},t)-\mathcal{C}_n\{\boldsymbol{\beta},\Sigma_R(t)\}\| \longrightarrow 0$. Hence, the first term of (11) can be written as $n^{-1}\sum_{i=1}^n\int_0^\tau\{\boldsymbol{X}^{(ls)}-e(\boldsymbol{\beta},s)\}dM_i(s)$, which is $o_p(1)$ since it has mean 0 and variance of order $O(n^{-1})$. Therefore, it can be seen that

$$
\begin{aligned}
\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta},\boldsymbol{W},\tau) &= \int_0^\tau \boldsymbol{S}^{(1)}(\boldsymbol{\beta}_0,\boldsymbol{X},s)ds + n^{-1}\sum_{i=1}^n\int_0^\tau Y_i(s)\boldsymbol{R}_i(s)e^{\boldsymbol{\beta}_0'\boldsymbol{X}_i}\lambda_0(s)ds \\
&\quad -\int_0^\tau \mathbf{E}(\boldsymbol{\beta},\boldsymbol{X},s)S^{(0)}(\boldsymbol{\beta}_0,\boldsymbol{X},s)\lambda_0(s)ds + o_p(1) \\
&= \int_0^\tau \{s^{(1)}(\boldsymbol{\beta}_0,s)-e(\boldsymbol{\beta},s)s^{(0)}(\boldsymbol{\beta}_0,s)\}\lambda_0(s)ds + o_p(1).
\end{aligned}
$$

In the above calculation, we use (6) and that $\boldsymbol{R}_i$ and $\boldsymbol{X}_i$ are independent. Hence, the limiting estimating function is zero at $\boldsymbol{\beta}_0$. Also, as in Andersen and Gill (1982), the derivative is a negative function of $\boldsymbol{\beta}$. We have thus shown the consistency of $\tilde{\boldsymbol{\beta}}$.

To derive the limiting distribution of $\boldsymbol{\beta}$, we note that $n^{1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = \{n^{-1}(\partial/\partial\boldsymbol{\beta})\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}_*,\boldsymbol{W},\tau)\}^{-1}n^{-\frac{1}{2}}\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}_0,\boldsymbol{W},\tau)$, where $\boldsymbol{\beta}_*$ is between $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$. Further,

$$
\begin{aligned}
&n^{\frac{1}{2}}\widehat{\boldsymbol{\Psi}}_c(\boldsymbol{\beta}_0,\boldsymbol{W},t) \\
&= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\{\boldsymbol{X}_i^{(ls)}(s)-\mathbf{E}(\boldsymbol{\beta}_0,\boldsymbol{X}^{(ls)},s)+\mathcal{C}_n\{\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_R(s)\}\}dM_i(s) \\
&\quad + n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\Big[\boldsymbol{X}_i^{(ls)}(s)-\mathbf{E}(\boldsymbol{\beta}_0,\boldsymbol{X}^{(ls)},s)+\mathcal{C}_n\{\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_R(s)\}\Big]Y_i(s)e^{\boldsymbol{\beta}_0'\boldsymbol{X}_{i(s)}}\lambda_0(s)ds \\
&\equiv A_1 + A_2.
\end{aligned}
$$

It is easily seen that

$$
\begin{aligned}
A_1 &= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\{\boldsymbol{X}_i^{(ls)}-e(\boldsymbol{\beta}_0,s)\}dM_i(s) \\
&\quad - n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t\{\mathbf{E}(\boldsymbol{\beta}_0,\boldsymbol{X}^{(ls)},s)-e(\boldsymbol{\beta}_0,s)-\mathcal{C}_n(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_R(s)\}dM_i(s).
\end{aligned}
$$

By Example 2.11.16 of van der Vaart and Wellner (1996, p.215), $n^{1/2}\overline{M}$ converges in $l^\infty[0,1]$ to a zero-mean Gaussian process $\mathcal{W}_M$ with continuous paths, where $\overline{M} = \sum_{i=1}^n M_i/n$. By the Strong Embedding Theorem (Shorack and Wellner (1986, pp.147-148)) there exists a new probability space in which $(n^{1/2}\overline{M}, \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \cdot, \boldsymbol{X}^{(ls)}), S^{(0)}(\boldsymbol{\beta}, \cdot, \boldsymbol{X}^{(ls)}))$ converges *a.s.* to $(\mathcal{W}_M, \boldsymbol{s}_*^{(1)}(\boldsymbol{\beta}, \cdot), s_*^{(0)}(\boldsymbol{\beta}, \cdot))$, where $\boldsymbol{s}_*^{(1)}(\boldsymbol{\beta}, t) = \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)\mathcal{E}(e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\boldsymbol{\beta}/2}) + s^{(0)}(\boldsymbol{\beta}, t)\mathcal{E}(\boldsymbol{\beta}'\boldsymbol{\Sigma}_R e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\boldsymbol{\beta}/2})$, $s_*^{(0)}(\boldsymbol{\beta}, t) = s^{(0)}(\boldsymbol{\beta}, t)\mathcal{E}(e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\boldsymbol{\beta}/2})$. By Lemma A.3 of Bilias et al. (1997), $n^{1/2}\int_0^t \mathbf{E}(\boldsymbol{\beta}_0, \boldsymbol{X}^{(ls)}, s)d\overline{M}(s) \to \int_0^t\{e(\boldsymbol{\beta}_0, s) + \mathcal{C}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_R)\}d\mathcal{W}_M(s)$ *a.s.* uniformly in t. Likewise, $n^{1/2}\int_0^t[e(\boldsymbol{\beta}_0, s) + \mathcal{C}_n\{\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_R(s)\}]d\overline{M}(s) \to \int_0^t[e(\boldsymbol{\beta}_0, s) + \mathcal{C}\{\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_R(s)\}]d\mathcal{W}_M(s)$ *a.s.* uniformly in t. Hence, $A_1 = n^{-1/2}\sum_{i=1}^n\int_0^t\{\boldsymbol{X}^{(ls)} - e(\boldsymbol{\beta}_0, s)\}dM_i(s) + o_p(1)$.

To approximate A_2 by a sum of independent variables, we note that

$$
\begin{aligned}
&S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)\\
&= \mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t) + n^{-1}\sum_{i=1}^n Y_i(t)e^{\boldsymbol{\beta}'\boldsymbol{X}_i}[e^{\boldsymbol{\beta}'\boldsymbol{R}_i(t)} - \mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]\\
&= \mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}\{S^{(0)}(\boldsymbol{\beta}, t, \boldsymbol{X}) + n^{-\frac{1}{2}}Q_0(\boldsymbol{\beta}, t)\},
\end{aligned}
$$

where $Q_o(\boldsymbol{\beta}, t) = n^{-1/2}\sum_{i=1}^n Y_i(t)e^{\boldsymbol{\beta}'\boldsymbol{X}_i}\Big([e^{\boldsymbol{\beta}'\boldsymbol{R}_i(t)}/\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}/2}\}] - 1\Big)$. Similarly,

$$
\begin{aligned}
&\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)\\
&= \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}/2}\} + S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}/2}\}\\
&\quad + n^{-1}\sum_{i=1}^n Y_i(t)e^{\boldsymbol{\beta}'\boldsymbol{X}_i}\{\boldsymbol{X}^{(ls)}e^{\boldsymbol{\beta}'\boldsymbol{R}_i(t)} - \boldsymbol{X}_i\mathcal{E}(e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}})\\
&\quad - \mathcal{E}(\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}})\\
&= \mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}\left[\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t) + S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\} + n^{-\frac{1}{2}}\mathbf{Q}_1(\boldsymbol{\beta}, t)\right],
\end{aligned}
$$

where $\mathbf{Q}_1(\boldsymbol{\beta}, t) = n^{-1/2}\sum_{i=1}^n Y_i(t)e^{\boldsymbol{\beta}'\boldsymbol{X}_i}([\boldsymbol{X}^{(ls)}e^{\boldsymbol{\beta}'\boldsymbol{R}_i(t)}/\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}/2}\}] - \boldsymbol{X}_i - \mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\})$. Hence

$$
\begin{aligned}
&\mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)\\
&= \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}, t) + \mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\} + n^{-\frac{1}{2}}\{\frac{\mathbf{Q}_1(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)}\}\\
&\quad - [\mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}, t) + \mathcal{C}\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(t)\}]\{\frac{n^{-\frac{1}{2}}Q_0(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)}\} + o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

By direct calculation, we have

$$A_2 = -n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \Big(\Big[\boldsymbol{X}^{(ls)} - e(\boldsymbol{\beta}_0, s) - \mathcal{C}\{\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_R(s)\} \Big] \frac{e^{\boldsymbol{\beta}_0' \boldsymbol{R}_i(s)}}{\mathcal{E}\{e^{\boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{R(s)} \boldsymbol{\beta}_0/2}\}}$$
$$- \{\boldsymbol{X}_i^{(ls)} - e(\boldsymbol{\beta}_0, s)\} \Big) Y_i(s) e^{\boldsymbol{\beta}_0' \boldsymbol{X}_i} \lambda_0(s) ds$$
$$+ n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \{\mathbf{E}(\boldsymbol{\beta}_0, \boldsymbol{X}, s) - e(\boldsymbol{\beta}_0, s)\} \Big[\frac{e^{\boldsymbol{\beta}_0' \boldsymbol{R}_i(s)}}{\mathcal{E}\{e^{\boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{R(s)} \boldsymbol{\beta}_0/2}\}} - 1 \Big]$$
$$\times Y_i(s) \boldsymbol{\beta}_0' \boldsymbol{X}_i \lambda_0(s) ds + o_p(1).$$

By using the argument of the Strong Embedding Theorem in proving the first term of A_1, it can be shown that the third term of A_2 converges to zero in probability. As a result,

$$n^{\frac{1}{2}} \widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta}_0, \tau, \boldsymbol{W}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \{\boldsymbol{X}^{(ls)} - e(\boldsymbol{\beta}_0, s)\} dN_i(s)$$
$$- n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \{\mathcal{E}\{e^{\boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{R(s)} \frac{\boldsymbol{\beta}_0}{2}}\}^{-1} [\boldsymbol{X}^{(ls)} - e(\boldsymbol{\beta}_0, s) - \mathcal{C}\{\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_R(s)\}]$$
$$\times Y_i(s) e^{\boldsymbol{\beta}_0' \boldsymbol{X}_i} \lambda_0(s) ds + o_p(1).$$

The above equation is a sum of n iid variables. Therefore, the asymptotic normality of $\tilde{\boldsymbol{\beta}}$ has been shown. If all the nuisance parameters involved in $\boldsymbol{\eta}$ were known, then the variance of $n^{1/2} \widehat{\boldsymbol{\Psi}}(\boldsymbol{\beta}, t, \boldsymbol{W})$ can be approximated by $n^{-1} \sum_{i=1}^{n} \boldsymbol{\Phi}_P(\boldsymbol{\beta}, \boldsymbol{W}_i)^{\otimes 2}$, where

$$\boldsymbol{\Phi}_P(\boldsymbol{\beta}, \boldsymbol{W}_i) = \int_0^\tau \{\boldsymbol{X}_i^{(ls)} - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, s) + \mathcal{C}_n\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_R(s)\}\} dN_i(s)$$
$$- \int_0^\tau \{\boldsymbol{X}_i^{(ls)} - \mathbf{E}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, s)\} Y_i e^{\boldsymbol{\beta}' \boldsymbol{X}_i} \{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, s)\}^{-1} \frac{dN_i(s)}{n}. \quad (12)$$

Appendix C. Calculation for Obtaining Second-order Corrected Score Estimator

Recall that $\boldsymbol{S}_*^{(m)}(\boldsymbol{\beta}, \boldsymbol{X}, t) = \mathcal{E}\{\boldsymbol{S}^{(m)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t) | \tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\}$ for $m = 0, 1$. By a Taylor series for $f(x, y) = y/x$, we have that

$$\mathcal{E}\left\{ \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)}{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)} \Big| \tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta} \right\}$$
$$= \frac{\boldsymbol{S}_*^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)}{S_*^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)} + \{S_*^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^{-2} \Big[\frac{\boldsymbol{S}_*^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)}{S_*^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)} \mathrm{var}\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t) | \tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\}$$
$$- \mathrm{cov}\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t), \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t) | \tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\} \Big] + o_p(n^{-1}). \quad (13)$$

By direct calculation,

$$\begin{aligned}
&\mathrm{var}\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\}\\
&= n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)\left[\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\} - \{\mathcal{E}(e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}})\}^2\right],
\end{aligned}$$

$$\begin{aligned}
&\mathrm{cov}\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t), \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)|\tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\}\\
&= n^{-1}\boldsymbol{S}^{(1)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\} + n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_R e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}\\
&\quad -n^{-1}\boldsymbol{S}^{(1)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2\\
&\quad -n^{-1}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}.
\end{aligned}$$

By using the above equations, (13) can be further reduced to

$$\begin{aligned}
&\mathcal{E}\Big\{\frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)}{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}^{(ls)}, t)}\Big|\tilde{\boldsymbol{X}}, \tilde{T}, \tilde{\delta}\Big\}\\
&= \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)}{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)} + \frac{\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\frac{\boldsymbol{\beta}}{2}}\}}{\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}}\Big[1 - \frac{n^{-1}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2}\Big]\\
&\quad +2\frac{n^{-1}\mathcal{E}\{\boldsymbol{\beta}\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}}{\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2}\\
&\quad -2\frac{n^{-1}\mathcal{E}\{\boldsymbol{\beta}\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2}\\
&\quad +\frac{n^{-1}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}S^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^3[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2}\\
&\quad -\frac{n^{-1}S^{(1)}(\boldsymbol{\beta}, \boldsymbol{X}, t)S^{(0)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2} - \frac{n^{-1}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}S^{(1)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]^2}\\
&\quad +\frac{n^{-1}S^{(1)}(2\boldsymbol{\beta}, \boldsymbol{X}, t)}{\{S^{(0)}(\boldsymbol{\beta}, \boldsymbol{X}, t)\}^2}. \qquad (14)
\end{aligned}$$

By direct calculation, it can be shown that if $\boldsymbol{\beta} \to 0$,

$$\frac{\mathcal{E}\{\boldsymbol{\beta}\boldsymbol{\Sigma}_R e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}}{\mathcal{E}\{\boldsymbol{\beta}\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\} - \mathcal{E}\{\boldsymbol{\beta}\boldsymbol{\Sigma}_R(t)e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\boldsymbol{\beta}}\}[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{R(t)}\frac{\boldsymbol{\beta}}{2}}\}]} \to \infty.$$

That is, the difference of the 3th and 4th terms of the right-hand side of (14) is of a smaller order than the third term of (14). Likewise, the last four terms of (14) can be shown to be

$$o_p\left(\frac{n^{-1}\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}}\}}{[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}]^3}\right)$$

if $n \to \infty$ and $\boldsymbol{\beta} \to 0$. Therefore, (14) can be written as

$$\begin{aligned}
&\mathcal{E}\Big\{\frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)}{S^{(0)}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)}\Big|\tilde{\boldsymbol{X}},\tilde{T},\tilde{\delta}\Big\}\\
&=\frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},\boldsymbol{X},t)}{S^{(0)}(\boldsymbol{\beta},\boldsymbol{X},t)}+\frac{\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\frac{\boldsymbol{\beta}}{2}}\}}{\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\frac{\boldsymbol{\beta}}{2}}\}}\Big[1-\frac{n^{-1}\mathcal{E}\{e^{2\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\boldsymbol{\beta}}\}S^{(0)}(2\boldsymbol{\beta},\boldsymbol{X},t)}{[\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}]^2\{S^{(0)}(2\boldsymbol{\beta},\boldsymbol{X},t)\}^2}\Big]\\
&\quad+o_p(n^{-1}\boldsymbol{\beta}).
\end{aligned}$$

As a result, using (8), it is easily seen that the above can be written as

$$\begin{aligned}
&\mathcal{E}\Big\{\mathbf{E}(\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)\Big|\tilde{\boldsymbol{X}},\tilde{T},\tilde{\delta}\Big\}\\
&=\mathbf{E}(\boldsymbol{\beta},\boldsymbol{X},t)+\frac{\mathcal{E}\{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R(t)\frac{\boldsymbol{\beta}}{2}}\}}{\mathcal{E}\{e^{\boldsymbol{\beta}'\boldsymbol{\Sigma}_R\frac{\boldsymbol{\beta}}{2}}\}}\Big(1-\mathcal{E}\Big[\frac{n^{-1}S^{(0)}(2\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)}{\{S^{(0)}(2\boldsymbol{\beta},\boldsymbol{X}^{(ls)},t)\}^2}\Big|\tilde{\boldsymbol{X}},\tilde{T},\tilde{\delta}\Big]\Big)\\
&\quad+o_p(n^{-1}\boldsymbol{\beta}).
\end{aligned}$$

Therefore, we have shown (9) as the second-order corrected score estimating equation.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662-682.

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error.*J. Statist. Plann. Inference* **67**, 247-257.

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

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

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

Hu, C and Lin, D. Y. (2002). Cox regression with covariate measurement error. *Scand. J. Statist.* **29**, 637-655.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. *J. Amer. Statist. Assoc.* **95**, 1209-1219.

Kong, F. H. and Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statist. Sinica* **9**, 953-969.

Laird, N. M. and Ware, J. H. (1982). Random-effect models for longitudinal data. *Biometrics* **38**, 963-974.

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

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-342.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

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

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716.

Tsiatis, A. A. and Davidian, D. (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 count in patients with AIDS. *J. Amer. Statist. Assoc.* **90**, 27-37.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Wang, C. Y., Wang, N. and Wang, S (2000). Regression analysis when covariates are regression parameters of a random effect model for observed longitudinal measurements. *Biometrics***56**, 487-495.

Xie, S. X., Wang, C. Y. and Prentice, R. L. (2001). A risk set calibration method for failure time regression using a covariate reliability sample. *J. Roy. Statist. Soc. Ser. B* **63**, 855-870.

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.

E-mail: cywang@fhcrc.org

(Received January 2004; accepted November 2004)