 Vol 33: 169–190, 2006

Inference on Survival Data with Covariate Measurement Error – An Imputation-based Approach

YI LI and LOUISE RYAN

Department of Biostatistics, Harvard University

ABSTRACT. We propose a new method for fitting proportional hazards models with error-prone covariates. Regression coefficients are estimated by solving an estimating equation that is the average of the partial likelihood scores based on imputed true covariates. For the purpose of imputation, a linear spline model is assumed on the baseline hazard. We discuss consistency and asymptotic normality of the resulting estimators, and propose a stochastic approximation scheme to obtain the estimates. The algorithm is easy to implement, and reduces to the ordinary Cox partial likelihood approach when the measurement error has a degenerate distribution. Simulations indicate high efficiency and robustness. We consider the special case where error-prone replicates are available on the unobserved true covariates. As expected, increasing the number of replicates for the unobserved covariates increases efficiency and reduces bias. We illustrate the practical utility of the proposed method with an Eastern Cooperative Oncology Group clinical trial where a genetic marker, c-*myc* expression level, is subject to measurement error.

Key words: bootstrap, covariate measurement error, Cox models, imputed partial likelihood score

1. Introduction

Analyses of survival data are often hampered by the presence of covariate measurement error. For example, in clinical trials, many biomarkers, such as blood pressure (Carroll *et al.*, 1995) and CD4 counts (Tsiatis *et al.*, 1995), are subject to measurement error, and in nutritional studies, fat intake is often measured with error (Carroll *et al.*, 1995). A large body of literature has been devoted to the measurement error problem within the proportional hazard model framework. Prentice (1982) has shown that the induced hazard function conditional on the observed covariates is also a multiplicative hazard model, but having a complicated form; consequently fitting the naive model by directly using the contaminated covariates will typically lead to biases. Several remedy methods have been proposed recently. Likelihood-based approaches, where the distribution of unobserved covariates is fully parametrically, semiparametrically and non-parametrically specified have been considered by Hu *et al.* (1998). From perspectives of estimating equation, Huang & Wang (2000), and Tsiatis & Davidian (2001) (whose method is asymptotically equivalent to that of Nakamura, 1992) have considered asymptotically consistent corrected partial likelihood approaches. Their methods are robust with respect to the assumptions on the unobserved covariates or measurement error.

In all these aforementioned methods, the cumulative baseline hazards are non-parametrically estimated by step functions with jumps at distinctive failure times. Though they are robust with respect to misspecification in the baseline hazards, the number of unknown parameters which need to be estimated increase with number of events. Hence, the computation may be complex and the efficiency may be low. A naive alternative is to consider a fully parametric model, with all the baseline hazards and the distribution of unknown covariates specified, and carry out a parametric maximum likelihood estimation, which shall yield consistent and the most efficient estimates under the correct model. However, the estimates are generally biased when the model is misspecified. In view of these difficulties and along the line of the imputational methods introduced by Satten *et al.* (1998) for interval censored data, and by Li *et al.* (2003) for clustered survival data, this article proposes a method that is intermediate between the fully parametric and non-parametric approaches for survival data with covariate measurement error. A linear spline model is assumed on the baseline hazard, but it is only used to impute the unobserved error prone covariates. Once the baseline distribution is specified, the distribution of true covariates conditional on the observed data can be calculated. A new estimating equation can be obtained by averaging the score equation for the Cox partial likelihood with respect to the conditional distribution of unobserved true covariates, given observed data. We propose to use this average partial likelihood score equation for estimating regression parameters, and give closed-form estimators of the sampling variance of our proposed estimators. The algorithms are easy to implement, and reduces to the ordinary Cox partial likelihood approach, when the measurement error has a degenerate distribution. Simulations indicate high efficiency and robustness of the estimates obtained by the proposed method. As expected, increasing the number of replicates for the unobserved covariates increases efficiency and reduces bias.

The rest of the article is structured as follows. In sections 2 and 3, we state the model and derive an average partial likelihood score equation. We show in section 4, asymptotic results and outline in section 5 a stochastic approximation scheme for constructing the estimates. In section 6, we assess via simulation the finite sample performance of the proposed methods, and in section 7, we apply the methods to the analysis of a published clinical trial (Augenlicht *et al.*, 1997) in colon cancer. We conclude with general discussion in section 8.

2. Imputed partial likelihood score equation

Let V_i and C_i be failure and censoring times, respectively, for subject i, $i = 1,\dots,m$. Suppose for each subject, an error-free covariate vector $\mathbf{Z}_i(r_1 \times 1)$, and an error-prone covariate vector $\mathbf{X}_i(r_2 \times 1)$, where $r_1 + r_2 = r$, are of interest. We suppose that the $\mathbf{X}_i$ are not directly observable, but instead, multiple error-prone replicates, $\mathbf{W}_i = (\mathbf{W}_{i1},\dots,\mathbf{W}_{in_i})$, are observed, where the number of replicates n_i is an i.i.d. random variable taking positive integer values and is independent of $\mathbf{X}_i$.

We assume that the C_i are independently and identically distributed and independent of the V_i and $\mathbf{X}_i$, conditional on $\mathbf{Z}_i$. The observed data are right censored with only $T_i = \min\{V_i, C_i\}$ and the censoring code $\delta_i = I(V_i \le C_i)$ observed, where $I(\cdot)$ denotes an indicator function. Introduce the counting process $N_i(t) = I(T_i \le t, \delta_i = 1)$, and the at-risk process $Y_i(t) = I(T_i \ge t)$. Our model specifies that, conditional on the true covariates $\mathbf{Z}_i$ and $\mathbf{X}_i$, the counting process $N_i(t)$ has an intensity function following the proportional hazards model (Cox, 1972),

$$\lim_{\Delta t \to 0+} \frac{1}{\Delta t} P\{N_i(t+\Delta t) - N_i(t) = 1 | \mathbf{X}_i, \mathbf{Z}_i, N_i(s), Y_i(s), 0 \le s \le t\} = \lambda_0(t) Y_i(t) \exp(\mathbf{X}_i'\boldsymbol{\beta}_x + \mathbf{Z}_i'\boldsymbol{\beta}_z), \quad (1)$$

where $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_z$ are the vectors of fixed effects and $\lambda_0(t)$ is an unknown baseline hazard function.

This model is completed by adding a classical non-differential measurement error structure (Carroll *et al.*, 1995),

$$\mathbf{W}_{ij} = \mathbf{X}_i + \mathbf{U}_{ij}, \quad (2)$$

where the $\mathbf{U}_{ij}$ are assumed independent of $\mathbf{X}_i$, $\mathbf{Z}_i$, and are i.i.d. with a normal distribution $N(0, \boldsymbol{\Sigma}_u)$. Here *non-differentiality* indicates the conditional law $L(T_i, \delta_i | \mathbf{X}_i, \mathbf{W}_i, \mathbf{Z}_i) =$

$L(T_i, \delta_i | \mathbf{X}_i, \mathbf{Z}_i)$, implying that, conditional on the true unobserved covariate $\mathbf{X}_i$, the observed replicates $\mathbf{W}_i$ do not contain additional information about the survival outcome (T_i, δ_i).

In what follows, we assume that the unobserved covariates $\mathbf{X}_i$ are time-invariant and are i.i.d. with a known conditional density function $f_{x|z}(\mathbf{X}|\mathbf{Z})$ or conditional distribution function $F_{x|z}(\mathbf{X}|\mathbf{Z})$. It is common in practice to postulate a normal distribution $N(\boldsymbol{\mu}_{x|z}, \boldsymbol{\Sigma}_{x|z})$ (Carroll *et al.*, 1995).

For notational convenience, we denote $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$, and likewise for $\mathbf{W}$, $\mathbf{T}$, $\boldsymbol{\Delta}$, $\mathbf{N}(t)$ and $\mathbf{Y}(t)$. Write $\bar{\mathbf{W}}_i = \sum_{j=1}^{n_i} \mathbf{W}_{ij}/n_i$, and $\bar{\mathbf{W}} = (\bar{\mathbf{W}}_1, \ldots, \bar{\mathbf{W}}_m)$. Under model (2) and non-differentiality of measure error, $\bar{\mathbf{W}}_i$ is a sufficient statistic for $\mathbf{X}_i$ conditional on T_i, δ_i and $\mathbf{Z}_i$. Hence, for convenience, we will work with $\bar{\mathbf{W}}_i$, $\bar{\mathbf{W}}$ in lieu of $\mathbf{W}_i$, $\mathbf{W}$, hereafter. Throughout, unless otherwise specified, $F(\cdot)$ represents a distribution function, $f(\cdot)$ is a density function, and expectations are taken conditionally on the observed covariates $\mathbf{Z}$.

Following Cox (1972), if $\mathbf{X}$ were observed, one would be able to estimate $\boldsymbol{\beta} = (\boldsymbol{\beta}_x, \boldsymbol{\beta}_z)$ from the 'complete' data partial likelihood score function

$$\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) = \sum_{i=1}^{m} \int_0^{\tau} \left\{ \mathbf{Z}_i^* - \frac{\mathcal{S}^{(1)}(t, \mathbf{Z}^*; \boldsymbol{\beta})}{\mathcal{S}^{(0)}(t, \mathbf{Z}^*; \boldsymbol{\beta})} \right\} \mathrm{d}N_i(t), \tag{3}$$

where $\mathbf{Z}^* = (\mathbf{Z}_1^*, \ldots, \mathbf{Z}_m^*)$, $\mathbf{Z}_i^{*\prime} = (\mathbf{X}_i', \mathbf{Z}_i')$, $\mathcal{S}^{(l)}(t, \mathbf{Z}^*; \boldsymbol{\beta}) = \sum_{i=1}^{m} \mathbf{Z}_i^{*\otimes l} Y_i(t) \exp(\mathbf{Z}_i^{*\prime}\boldsymbol{\beta})$, and $\tau < \infty$ is a constant such that $\mathrm{pr}(C_i > \tau) > 0$. In practice, τ is usually the study duration. Here, for a vector $\mathbf{u}$, $\mathbf{u}^{\otimes l} = \mathbf{u}\mathbf{u}'$ if $l = 2$, $\mathbf{u}^{\otimes l} = \mathbf{u}$ if $l = 1$, and $\mathbf{u}^{\otimes l} = 1$ if $l = 0$.

However, as $\mathbf{X}$ is not observerable, the 'complete' data partial likelihood function (3) is not calculable. Instead, we propose to estimate $\boldsymbol{\beta}$ from an average partial likelihood function, which is the conditional expectation of $\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})$ with respect to $\mathbf{X}$ over the observed quantities, i.e. the observed survival information, the observed covariates $\mathbf{Z}$ and the error-prone covariates $\bar{\mathbf{W}}$. Explicitly, we introduce

$$\mathbb{S}\{\boldsymbol{\beta}; \lambda_0(\cdot)\} = \mathrm{E}\{\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) | \mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \boldsymbol{\beta}, \lambda_0(\cdot)\}. \tag{4}$$

Denote by $\boldsymbol{\beta}_0$ and $\tilde{\lambda}_0(\cdot)$, the true values of $\boldsymbol{\beta}$ and $\lambda_0(\cdot)$, respectively.Using a double expectation theorem (e.g. Fleming & Harrington, 1991, p. 22), we can show that

$$\mathrm{E}[\mathbb{S}\{\boldsymbol{\beta}_0; \tilde{\lambda}_0(\cdot)\}] = 0 \tag{5}$$

under $\boldsymbol{\beta}_0$ and $\tilde{\lambda}_0(\cdot)$. As a result, $\mathbb{S}\{\boldsymbol{\beta}; \tilde{\lambda}_0(\cdot)\} = 0$ is indeed an unbiased estimating equation for $\boldsymbol{\beta}$, given the true value of the baseline hazard function.

The form of (4) thus motivates us to regard the unobserved true covariates as 'missing' covariates, to impute them by simulating from the conditional distribution, and to use the imputed values to construct an unbiased estimating equation; we term this simulated version of (4) an *imputed partial likelihood equation*. Indeed, data augmentation by imputing unobserved quantities has drawn large attention in recent years, and multiple imputation has become a general approach to handle missing values in regression models (see, e.g. Rubin, 1987; Rubin & Schenker, 1991).

We also remark that this proposed imputation-based methodology can easily handle the presence of interactions between the error-free and error-prone covariates, which has been rarely explored in the existing literature.

3. Construction of the estimating equation

The conditional score function (4) can be written as

$$\int \mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta};) \, \mathrm{d}F\{\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \boldsymbol{\beta}, \lambda_0(\cdot)\},$$

where the conditional distribution of $\mathbf{X}$, F has a product form

$$F\{\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \boldsymbol{\beta}, \lambda_0(\cdot)\} = \prod_{i=1}^{m} F\{\mathbf{X}_i|T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \lambda_0(\cdot)\}.$$

To fully specify this conditional distribution, instead of considering the baseline hazard function $\lambda_0(\cdot)$ in an infinite dimensional space, we impose a local parametric form on it. That is, we assume a piecewise-linear spline model for the log baseline hazard function:

$$\log \lambda_0(t, \boldsymbol{\eta}) = \eta_1 + \eta_2 t + \eta_3(t - \tau_1) + \cdots + \eta_q(t - \tau_{q-2})_+, \tag{6}$$

with knots fixed at $0 \equiv \tau_0 < \tau_1 < \cdots < \tau_{q-2}$. $x_+ = \max(x, 0)$, and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)$. In practice, the choice of knots should be data-driven, e.g. by the criteria of AIC or BIC, and Cai *et al.* (2002) recommend choosing knots densely to allow for a detailed study for the structure of the baseline hazard function. In our particular setting (for the purpose of imputation), we found that a relatively small number of knots suffice, especially when the underlying hazard is relatively smooth. Our later simulations confirm this.

Under model (2) and the normality assumption on measurement error, we can write

$$\mathrm{d}F\{\mathbf{X}_i|T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \lambda_0(\cdot, \boldsymbol{\eta})\} = \frac{L(T_i, \delta_i|\mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})\phi(\bar{\mathbf{W}}_i; \mathbf{X}_i, \boldsymbol{\Sigma}_u/n_i) \, \mathrm{d}F_{x|z}(\mathbf{X}_i|\mathbf{Z}_i)}{L(T_i, \delta_i, \bar{\mathbf{W}}_i|\mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})}, \tag{7}$$

where $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density function for a multivariate normal random variable $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $L(T_i, \delta_i|\mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})$ is the conditional likelihood for the ith subject given the covariate $\mathbf{X}_i$ under (1) and (6), i.e.

$$L(T_i, \delta_i|\mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta}) = \lambda_0^{\delta_i}(T_i, \boldsymbol{\eta}) \exp\left\{\delta_i(\mathbf{X}_i'\boldsymbol{\beta}_x + \mathbf{Z}_i'\boldsymbol{\beta}_z) - \Lambda_0(T_i, \boldsymbol{\eta}) \, \mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}_x + \mathbf{Z}_i'\boldsymbol{\beta}_z}\right\}, \tag{8}$$

and $L(T_i, \delta_i, \bar{\mathbf{W}}_i|\mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})$ is the marginal likelihood for the observed data $(T_i, \delta_i, \bar{\mathbf{W}}_i)$. That is

$$L(T_i, \delta_i, \bar{\mathbf{W}}_i|\mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta}) = \int L(T_i, \delta_i|\mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})\phi(\bar{\mathbf{W}}_i; \mathbf{X}_i, \boldsymbol{\Sigma}_u/n_i) \, \mathrm{d}F_{x|z}(\mathbf{X}_i|\mathbf{Z}_i), \tag{9}$$

which follows from the non-differentiality of the assumed measurement error. In (8) the cumulative baseline hazard $\Lambda_0(t, \boldsymbol{\eta}) = \int_0^t \lambda_0(s, \boldsymbol{\eta}) \, \mathrm{d}s$.

For notational ease, we rewrite

$$\mathbb{S}\{\boldsymbol{\beta}; \lambda_0(\cdot, \boldsymbol{\eta})\} = \mathbb{S}(\boldsymbol{\beta}, \boldsymbol{\eta}). \tag{10}$$

As $\boldsymbol{\eta}$ is unknown, we resort to a full likelihood maximization to obtain an estimate. Specifically, given an estimate of $\boldsymbol{\beta}$, we solve for $\boldsymbol{\eta}$ from the following log likelihood score equation

$$\mathbb{U}(\boldsymbol{\beta}, \boldsymbol{\eta}) = \sum_{i=1}^{m} \mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta}) = \sum_{i=1}^{m} \frac{\partial}{\partial \boldsymbol{\eta}} \log\left\{L(T_i, \delta_i, \bar{\mathbf{W}}_i|\mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})\right\}, \tag{11}$$

where $L(\cdot)$ is defined at (9).

The aforementioned scheme is equivalent to obtaining estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ by simultaneously solving

$$\begin{pmatrix} \mathbb{S}(\boldsymbol{\beta}, \boldsymbol{\eta}) \\ \mathbb{U}(\boldsymbol{\beta}, \boldsymbol{\eta}) \end{pmatrix} = 0. \tag{12}$$

Under model (1) with $\lambda_0(t; \boldsymbol{\eta})$ correctly specified and by the usual properties of maximum likelihood estimation, the estimating equations (12) are unbiased, and, consequently, would be expected to yield consistent estimates.

Even when $\lambda_0(t; \boldsymbol{\eta})$ is incorrectly specified, we may expect that (10) is less sensitive to such misspecification than is the observed data likelihood score obtained by differentiating (9) with respect to $\boldsymbol{\beta}$, because the 'complete' data partial likelihood score $\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})$ is independent of the baseline distribution. A similar observation was made by Satten *et al.* (1998), who imputed unobserved failure times in the context of independent interval-censored survival data by specifying a Weibull model on the baseline hazard. We study this more analytically in section 4 and conduct simulations in section 7 to verify the theoretical results.

4. Asymptotic theory and robustness analysis

Denote by $\tilde{\boldsymbol{\beta}}_0$, the true value of the regression coefficient $\boldsymbol{\beta}$. With the assumption that (T_i, δ_i, $\mathbf{X}_i$, $\mathbf{Z}_i$) are i.i.d., Lin & Wei (1989) have shown that the 'complete' data partial likelihood score $\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \tilde{\boldsymbol{\beta}}_0)$ can be represented, up to $o_p(1)$, as an i.i.d. sum. That is,

$$m^{-1/2}\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \tilde{\boldsymbol{\beta}}_0) = m^{-1/2}\sum_{i=1}^{m} \boldsymbol{\xi}(T_i, \delta_i, \mathbf{X}_i, \mathbf{Z}_i; \tilde{\boldsymbol{\beta}}_0) + o_p(1), \tag{13}$$

where

$$\boldsymbol{\xi}(T_i, \delta_i, \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}) = \delta_i\left\{\mathbf{Z}_i^* - \frac{s^{(1)}(T_i; \boldsymbol{\beta})}{s^{(0)}(T_i; \boldsymbol{\beta})}\right\} - e^{\mathbf{Z}_i^{*\prime}\boldsymbol{\beta}}\int_0^{T_i}\left\{\mathbf{Z}_i^* - \frac{s^{(1)}(t; \boldsymbol{\beta})}{s^{(0)}(t; \boldsymbol{\beta})}\right\}\frac{\mathrm{d}G(t)}{s^{(0)}(t; \boldsymbol{\beta})}.$$

Here, $G(t) = \mathrm{E}\{N_i(t)\}$, and $s^{(l)}(t; \boldsymbol{\beta}) = \mathrm{E}\{\mathcal{S}^{(l)}(t, \mathbf{Z}^*; \boldsymbol{\beta})\}$ for $l = 0, 1, 2$, where the expectation is taken with respect to the true (but unspecified) distribution distribution of (N, Y, $\mathbf{X}$, $\mathbf{Z}$).

Denote by $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\eta})$, the collection of unknown parameters. Following the proof of Datta *et al.* (2000) and Li *et al.* (2003), we further show that in a small neighborhood of $\boldsymbol{\gamma}$ the term-wise integration of (13) is allowable, enabling us to write

$$m^{-1/2}\mathbb{S}(\boldsymbol{\gamma}) = m^{-1/2}\sum_{i=1}^{m} \boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\gamma}) + o_p(1), \tag{14}$$

where $\mathbb{S}(\cdot)$ is defined in (10) and $\boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\gamma}) = \int \boldsymbol{\xi}(T_i, \delta_i, \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}) \times \mathrm{d}F\{\mathbf{X}_i | T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \lambda_0(\cdot, \boldsymbol{\eta})\}$. Hence, we are able to approximate the average partial likelihood score with respect to the conditional distribution of unobserved covariates using a sum of i.i.d. random variables.

Denote by $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0)$ the solution to the following equation:

$$\mathrm{E}\begin{pmatrix} \boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \boldsymbol{\gamma}) \\ \frac{\partial}{\partial \boldsymbol{\eta}}\log\{L(T_i, \delta_i, \bar{\mathbf{W}}_i | \mathbf{Z}_i; \boldsymbol{\gamma})\} \end{pmatrix} = 0, \tag{15}$$

where the expectation is taken with respect to the true distribution distribution of (N, Y, $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$). Note when the baseline hazard $\lambda_0(t; \boldsymbol{\eta})$ and the distribution of $\mathbf{X}$, $\mathbf{W}$ are correctly specified, $\boldsymbol{\beta}_0 = \tilde{\boldsymbol{\beta}}_0$, as the terms inside the braces of (15) are unbiased under the true parameters.

When implementing the proposed estimating procedure, however, we might misspecify the baseline hazard function as well as the distribution functions of **X**, **W** and measurement error **U**. Hence the proposed estimates will incorporate some asymptotic bias. Theorem 1 shows that, under regularity conditions C1–C3 (in the Appendix), the imputed partial likelihood estimates converge in probability to $\boldsymbol{\beta}_0$, the solution to (15), enabling the calculation of potential asymptotic biases.

Theorem 1

There exists a sequence of solutions $\widehat{\gamma}$ *to* (12) *such that for any given* $\epsilon > 0$, *there exists a* $K < \infty$ *and an integer* $m_0 > 0$ *such that* $\mathrm{pr}\{\widehat{\gamma} \in \mathcal{N}_{K/\sqrt{m}}(\gamma 0)\} \geq 1 - \epsilon$ *for any* $m \geq m_0$, *where* $\mathcal{N}_\rho(\gamma_0)$ *is the neighbourhood around* γ_0 *with radius* ρ.

Using (14), it can be shown readily that

$$m^{-1/2}\{\mathbb{S}(\gamma_0), \mathbb{U}(\gamma_0)\} \xrightarrow{d} N(0, \boldsymbol{\Psi}), \tag{16}$$

where $\boldsymbol{\Psi}$, the covariance matrix of $\boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)$ and $\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)$, is defined later. Asymptotic properties of the solution to (12) follow in part from (16) and are summarized in theorem 2.

Theorem 2

Let $\widehat{\gamma}$ *be a solution to* (12) *that converges to* γ_0 *in probability. Then*

$$m^{1/2}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{A}^{-1}\boldsymbol{\Psi}(\mathbf{A}^{-1})^{\mathrm{T}}$, *with* $\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}'_{12} & \boldsymbol{\Psi}_{22} \end{pmatrix}$. *Here,* $\mathbf{A}$ *is expectation of the Jacobian matrix of the score equations* (12) *and*

$$\boldsymbol{\Psi}_{11} = \mathrm{E}\{\boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)\boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)'\},$$

$$\boldsymbol{\Psi}_{12} = \mathrm{E}\{\boldsymbol{\psi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)'\},$$

$$\boldsymbol{\Psi}_{22} = \mathrm{E}\{\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma_0)'\}.$$

We now compute the asymptotic bias from (15), wherein the expectation can be evaluated, for any measurable function g, by

$$\begin{aligned}&\mathrm{E}\{g(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i)\}\\ &\quad = \sum_{\delta=0}^{1} \int g(t, \delta, \bar{\mathbf{w}}, \mathbf{z})\lambda^{\delta}(t|\mathbf{x}, \mathbf{z})S(t|\mathbf{x}, \mathbf{z})c_c^{1-\delta}(t)C_c^{\delta}(t)f_{\bar{w}|x,z}(\bar{\mathbf{w}}|\mathbf{x}, \mathbf{z})f_{x,z}(\mathbf{x}, \mathbf{z})\,\mathrm{d}t\,\mathrm{d}\bar{\mathbf{w}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{z},\end{aligned}$$

where $\lambda(t|\mathbf{x}, \mathbf{z})$ [defined in (1)], $S(t|\mathbf{x}, \mathbf{z}) = \exp(-\int_0^t \lambda(s|\mathbf{x}, \mathbf{z})\,\mathrm{d}s)$ are the true hazard function and survival function, respectively, $c_c(t)$, $C_c(t)$ are the density and survival functions for the censoring time C, and $f_{\bar{w}|x,x}$, $f_{x,z}$ are the conditional density function of $\bar{\mathbf{W}}$ and the joint density of **X**, **Z**, respectively. Numerical integration can be employed to evaluate this integral, if necessary.

To illustrate bias patterns, we calculated the asymptotic biases for the imputed partial likelihood estimates (denoted by $\hat{\beta}_I$) under the following settings: the true hazard $\lambda(t|x) = \lambda_0(t)\exp(\tilde{\beta}_0 x)$, where $\lambda_0(t) = 0$ if $t < 1$ and 2 if $t \geq 1$, $x \sim N(0, 1)$; censoring time $C \sim U(0, 4)$; the number of replicates per subject is $n_i \equiv 4$. We varied the measurement error variance σ_u^2 from 0.25 to 2 and the regression coefficient $\tilde{\beta}_0$ from 0.25 to 2. For the purpose of

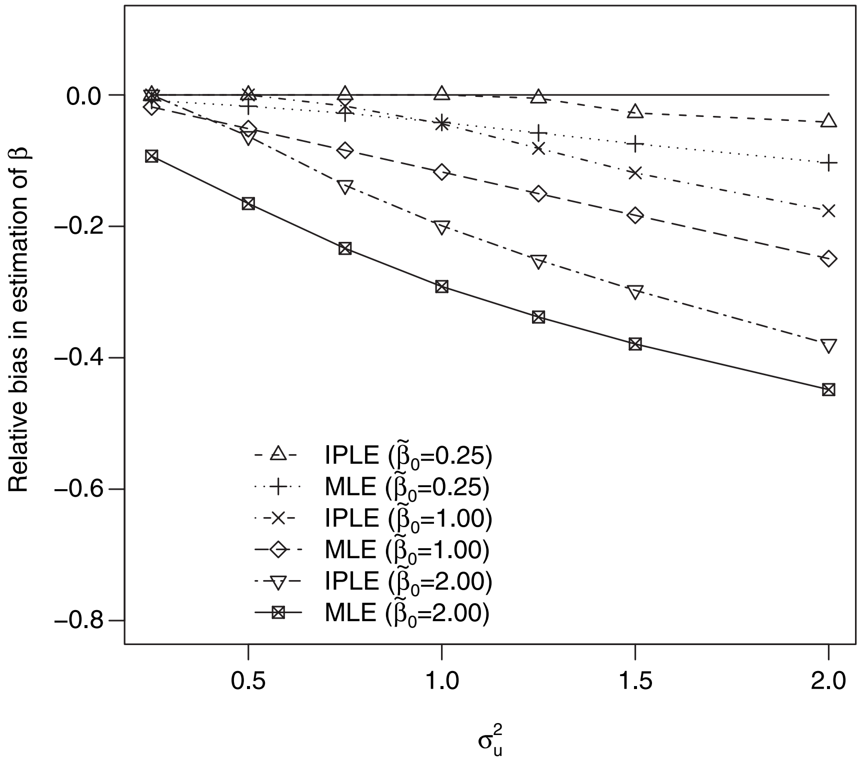

Fig. 1. Asymptotic biases of $\hat{\beta}_I$, the imputed partial likelihood estimates (IPLE), and $\hat{\beta}_L$, the parametric MLE, under various values of $\tilde{\beta}_0$ and σ_u^2. The horizontal line is the zero-line, corresponding to no bias.

comparison we also calculated the asymptotic biases for the fully parametric maximum likelihood estimates (denoted by $\hat{\beta}_L$), converging in probability to the solution to the following equation.

$$\mathrm{E}\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} \log\{L(T_i, \delta_i, \bar{\mathbf{W}}_i | \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})\} \\ \frac{\partial}{\partial \boldsymbol{\eta}} \log\{L(T_i, \delta_i, \bar{\mathbf{W}}_i | \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})\} \end{pmatrix} = 0.$$

The relative biases (|bias|/true value) were computed and depicted in Fig. 1 for various measurement error variances and true regression parameter $\tilde{\beta}_0$. When calculating $\hat{\beta}_I$ and $\hat{\beta}_L$, we assumed a linear spline model on the baseline hazard $\log \lambda_0(t, \boldsymbol{\eta}) = \eta_1 + \eta_2 t + \eta_3(t - 0.5)_+$.

The plot indicates the asymptotic biases for $\hat{\beta}_I$ were small, when $\tilde{\beta}_0$ was not too large (i.e. <1) and when the measurement error was moderate [the noise-signal ratio $(\sigma_u^2/n_i)/\sigma_x^2 < 0.25$], in which cases the absolute asymptotic biases for $\hat{\beta}_I$ were less than 0.20 and the relative biases were below 12%. We also notice that, under extreme cases, when $\tilde{\beta}_0 > 0$ and the measurement error variance was as large as 2, the absolute asymptotic bias for $\hat{\beta}_I$ was large, indicating that the proposed method may not work well. Under all the scenarios examined, the absolute asymptotic biases for the fully parametric MLE, $\hat{\beta}_L$, always exceeded those for $\hat{\beta}_I$. For example, when the measurement error variance was 2 and $\tilde{\beta}_0 = 0.25$, the relative bias for $\hat{\beta}_L$ was −0.12 compared with −0.04 for $\hat{\beta}_I$. Note that even in the absence of measurement error ($\sigma_u^2 = 0$), $\hat{\beta}_L$ was still biased (especially when $\tilde{\beta}_0$ was as large as 2.0) because the baseline hazard function was misspecified. In contrast, our proposed imputation-based method reduces to the standard Cox proportional hazards analysis in the absence of measurement error, and hence incurs no bias.

5. A stochastic approximation scheme

In practice, $\mathbf{\Sigma}_u$, $\boldsymbol{\mu}_x$ and $\mathbf{\Sigma}_x$ will typically be unknown, but can be consistently estimated by moment estimators, when replicate data are available. For example, following Carroll *et al.* (1995), we have the following $\sqrt{m}$-consistent estimating equations:

$$\widehat{\boldsymbol{\mu}}_x = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{W}_{ij}}{\sum_{i=1}^m n_i}, \qquad \widehat{\mathbf{\Sigma}}_u = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)(\mathbf{W}_{ij} - \bar{\mathbf{W}}_i)'}{\sum_{i=1}^m (n_i - 1)},$$

and

$$\widehat{\mathbf{\Sigma}}_x = \frac{\sum_{i=1}^m n_i(\bar{\mathbf{W}}_i - \widehat{\boldsymbol{\mu}}_x)(\bar{\mathbf{W}}_i - \widehat{\boldsymbol{\mu}}_x)'}{\sum_{i=1}^m n_i}.$$

We rewrite (11) as

$$\mathbb{U}(\gamma) = \int \mathbf{U}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta})\, \mathrm{d}F(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}),$$

where $\mathbf{U}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta}) = (\partial/\partial\boldsymbol{\eta}) \sum_{i=1}^m \log L(T_i, \delta_i; \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \boldsymbol{\eta})$. Both $\mathbb{U}(\gamma)$ and $\mathbb{S}(\gamma)$ in (12) are multidimensional integral with respect to the conditional distribution $F(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z})$. As neither $F(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z})$ nor $f(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z})$ has a closed form, we propose to use a stochastic approximation (see, e.g. Gu & Zhu, 2001), coupled with a sampling-importance resampling (SIR) scheme (McLachlan and Krishnan, 1997, Ch. 6), to evaluate $\mathbb{U}(\gamma)$ and $\mathbb{S}(\gamma)$ and solve (12). In contrast to the common rejection sampling approach, an advantage of the SIR algorithm is that the bound of the ratio of the candidate and the target distributions do not need to be evaluated. We proceed as follows.

Denote by γ^k, the estimate of the unknown parameter $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\eta})$ at the kth step. Assume $\mathbf{X}_{k,1},\ldots,\mathbf{X}_{k,n}$ are n i.i.d realizations following distribution $F(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \mathbf{\Theta}^k)$. Introduce $\mathcal{I}(\mathbf{X}; \boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\mathbf{S}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})$ and $\mathcal{J}(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\eta}) = (\partial/\partial\boldsymbol{\eta})\mathbf{U}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta})$. Let $I^k = (1/n)\sum_{i=1}^n \mathcal{I}(\mathbf{X}_{k,i}; \boldsymbol{\beta}^k)$ $J^k = (1/n)\sum_{i=1}^n \mathcal{J}(\mathbf{X}_{k,i}; \boldsymbol{\beta}^k, \boldsymbol{\eta}^k)$, $\mathbf{S}^k = (1/n)\sum_{i=1}^n \mathbf{S}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}^k)$, and $\mathbf{U}^k = (1/n)\sum_{i=1}^n \mathbf{U}(\mathbf{T}, \mathbf{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}^k, \boldsymbol{\eta}^k)$.

Then at the $(k+1)$-step, the updated estimate of γ is

$$\hat{\boldsymbol{\beta}}^{k+1} = \tilde{\boldsymbol{\beta}}^k - a_k\tilde{\mathbf{I}}^k\tilde{\mathbf{S}}^k,$$

$$\hat{\boldsymbol{\eta}}^{k+1} = \tilde{\boldsymbol{\eta}}^k - a_k\tilde{\mathbf{J}}^k\tilde{\mathbf{U}}^k,$$

where $\tilde{\boldsymbol{\beta}}^k = (1/k)\sum_{i=1}^k \hat{\boldsymbol{\beta}}^i$, $\tilde{\boldsymbol{\eta}}^k = (1/k)\sum_{i=1}^k \hat{\boldsymbol{\eta}}^i$, $\tilde{\mathbf{I}}^k = (1/k)\sum_{i=1}^k \mathbf{I}^i$, $\tilde{\mathbf{S}}^k = (1/k)\sum_{i=1}^k \mathbf{S}^i$, $\tilde{\mathbf{J}}^k = (1/k)\sum_{i=1}^k \mathbf{J}^i$, $\tilde{\mathbf{U}}^k = (1/k)\sum_{i=1}^k \mathbf{U}^i$. Here the gain constant a_k is defined to be $a_k = c/(k^e + g)$, where c, $g > 0$, $e \in (0, 1)$ are fixed. In practice, e is chosen to be close to 1/2 and c to be small and g is relatively large (cf. Gu & Zhu, 2001). The iteration continues until convergence. To guarantee convergence in stochastic approximation, a good starting value is often needed; our experience suggested that choosing the naive estimates as starting points is preferable.

We apply the SIR to draw samples from intractable distribution $F(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma)$ or density $f(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma)$. Specifically, we draw M values, $\mathbf{X}_1,\ldots,\mathbf{X}_M$ from a candidate density $h(\mathbf{X})$ and calculate the *importance ratios*: $r_j = f(\mathbf{X}_j|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma)/h(\mathbf{X}_j)$ for $j = 1,\ldots,M$. Then draw n values from $\mathbf{X}_1,\ldots,\mathbf{X}_M$. It can be shown that as $M/n \to \infty$, the distribution of the drawn values follows the targeted distribution (McLachlan & Krishnan, 1997, Ch. 6). In order for SIR to work well, one needs to carefully select the candidate density $h(\mathbf{X})$ such that $h(\mathbf{X})$ and the object density function $f(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma)$ have the same support, and $h(\mathbf{X})$ approximates $f(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z})$ well. We proceed as follows.

First notice that $f(\mathbf{X}|\mathbf{T}, \mathbf{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}) = \prod_{i=1}^{m} f(\mathbf{X}_i|T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i)$, and apply a Taylor series expansion to $\ell_i(\mathbf{x}) \equiv \log f(\mathbf{x}|T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i)$ centred at its mode $\tilde{\boldsymbol{\mu}}_i$, yielding

$$\ell_i(\mathbf{x}) = \ell_i(\tilde{\boldsymbol{\mu}}_i) + \frac{1}{2}(\mathbf{x} - \tilde{\boldsymbol{\mu}}_i)'\left\{\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}'}\ell_i(\mathbf{x})|_{\mathbf{x}=\tilde{\boldsymbol{\mu}}_i}\right\}(\mathbf{x} - \tilde{\boldsymbol{\mu}}_i) + r_i, \tag{17}$$

where the remainder term r_i is negligible compared to the quadratic term if $\mathbf{x}$ is close to $\tilde{\boldsymbol{\mu}}_i$ or n_i is large. From (7) and (8), $\tilde{\boldsymbol{\mu}}_i$ solves

$$0 = \frac{\partial}{\partial\mathbf{x}}\ell_i(\mathbf{x}) = \left\{\delta_i - \Lambda_0(T_i, \boldsymbol{\eta})\, e^{\boldsymbol{\beta}_x'\mathbf{x}+\boldsymbol{\beta}_z'\mathbf{Z}_i}\right\}\boldsymbol{\beta}_x - n_i\boldsymbol{\Sigma}_u^{-1}(\mathbf{x} - \bar{\mathbf{W}}_i) - \boldsymbol{\Sigma}_x^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

and

$$-\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}'}\ell_i(\mathbf{x})|_{\mathbf{x}=\tilde{\boldsymbol{\mu}}_i} = \Lambda_0(T_i, \boldsymbol{\eta})\, e^{\boldsymbol{\beta}_x'\tilde{\boldsymbol{\mu}}_i+\boldsymbol{\beta}_z'\mathbf{Z}_i}\boldsymbol{\beta}_x\boldsymbol{\beta}_x' + n_i\boldsymbol{\Sigma}_u^{-1} + \boldsymbol{\Sigma}_x^{-1}.$$

Considering (17) as a function of $\mathbf{x}$ only, the first term is a constant, whereas the second term is proportional to the logarithm of a normal density, yielding a normal approximation $N(\tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)$, where

$$\tilde{\boldsymbol{\Sigma}}_i = \left\{-\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}'}\ell_i(\mathbf{x})|_{\mathbf{x}=\tilde{\boldsymbol{\mu}}_i}\right\}^{-1}.$$

Therefore, $h(\cdot) = \prod_{i=1}^{m} \phi(\cdot; \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)$. Notice when $\min_i(n_i) \to \infty$ or norm $||\boldsymbol{\Sigma}_u|| \to 0$, $\tilde{\boldsymbol{\mu}}_i - \bar{\mathbf{W}}_i \to 0$ almost surely and $||\tilde{\boldsymbol{\Sigma}}_i|| \to 0$ almost surely for $i = 1,\ldots,m$, in which case $\bar{\mathbf{W}}_i$ can be substituted for $\mathbf{X}_i$, and the estimating equation reduces to the ordinary partial likelihood equation and attains semiparametric efficiency (Bickel *et al.*, 1993). Consequently, increasing the number of replicates for the unobserved covariates increases efficiency and reduces bias.

6. Variance estimator

The variances of the maximum likelihood estimates are conventionally calculated by inverting the Fisher information matrix. However, because (12) is not an i.i.d. sum of ordinary likelihood scores, a more in-depth analysis is required to derive an estimate of the variance matrix of $\widehat{\gamma}$, the zero of (12). The appendix establishes that the asymptotic variance of $m^{1/2}(\widehat{\gamma} - \gamma_0)$, $\mathbf{V}$, can be estimated by

$$\widehat{\mathbf{V}} = \mathbf{A}_m^{-1}(\widehat{\gamma})\widehat{\boldsymbol{\Psi}}\{\mathbf{A}_m^{-1}(\widehat{\gamma})\}', \tag{18}$$

where $\mathbf{A}_m$ and $\widehat{\boldsymbol{\Psi}}$ are consistent estimates to $\mathbf{A}$ and $\boldsymbol{\Psi}$ in theorem 2 and are given in the appendix.

Computationally, using (18) to calculate a variance estimate is complicated and time-consuming, even without accounting for the additional variability induced by estimating $\boldsymbol{\mu}_x$, $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_u$. A simple alternative is to use the bootstrap approach (Efron, 1981). Specifically, we resample m subjects, with replacement, from $(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i)|_{i=1}^{m}$ to obtain a new data set $\{T_{(i)}, \delta_{(i)}, \bar{\mathbf{W}}_{(i)}, \mathbf{Z}_{(i)}\}|_{i=1}^{m}$. Given this new dataset, we solve (12) for the estimates of $\boldsymbol{\gamma}$, in particular, $\boldsymbol{\beta}$. Such a procedure can be repeated for B times to obtain a sequence of estimates, say, $\tilde{\boldsymbol{\beta}}^{(l)}$, $l = 1,\ldots,B$. The bootstrap variance estimates can be calculated using the sample variance

$$\mathrm{var}_{\mathrm{boot}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{B-1}\sum_{l=1}^{B}\left\{\tilde{\boldsymbol{\beta}}^{(l)} - \bar{\boldsymbol{\beta}}_{\mathrm{boot}}\right\}\left\{\tilde{\boldsymbol{\beta}}^{(l)} - \bar{\boldsymbol{\beta}}_{\mathrm{boot}}\right\}',$$

where $\bar{\boldsymbol{\beta}}_{\mathrm{boot}} = (1/B)\sum_{l=1}^{B}\tilde{\boldsymbol{\beta}}^{(l)}$. In practice, it is adequate to choose a moderate number of resamplings, B, in the range of 25–100 (Lange, 1999, p. 301). We chose $B = 40$ in our simulation studies. For a review of other nonparametric techniques for obtaining the variance estimates, such as the Jackknife procedure, the smoothed bootstrap method and the half-sampling approach, see Efron (1981) and Efron & Tibshirani (1993).

7. Simulation

Simulations were performed to assess the finite sample performance of the proposed imputed partial likelihood score (PLS) estimators. For simplicity, we focus on a single covariate measurement error survival model. Of particular interest were robustness and efficiency of the proposed estimator, along with the performance of the bootstrap variance estimator.

In each simulated data set, survival times V_i were generated for each individual by the hazard $\lambda_i(t) = \lambda_0(t)\exp(\tilde{\beta}_0 X_i)$, $i = 1,\ldots,m$, where $\tilde{\beta}_0 = 1$, and the X_i were generated independently from the standard normal $N(0, 1)$. Censoring times C_i were simulated from the uniform distribution on interval $[0, \tau]$.

We considered the following combinations of experiments: the number of subjects m was set to be 50, 100 and 200, while the number of replicates n_i were distributed according to the discrete uniform distribution with mass on the integers 1–5; for $j = 1,\ldots,n_i$, the measurement error U_{ij} were independently generated from $N(0, \sigma_u^2 = 0.5)$; τ was chosen to yield two different censoring proportions 30 Per cent (light censoring) and 70 Per cent (heavy censoring). When generating the data, we chose the following three models for the baseline hazard $\lambda_0(t)$ to examine the robustness of the proposed PLS approach. A log linear model:

$$\log\lambda_0(t) = \log 2 + 2t. \tag{19}$$

Weibull model: $\lambda_0(t) = 0.5t$, which departs from model (6); and a discrete hazard model, $\lambda_0(t) = 0$ if $t < 1$ and $\lambda_0(t) = 2$ if $t \geq 1$, corresponding to a common scenario in clinical trials, in which failures are not seen immediately following treatment. For each parameter configuration, a total 300 replicated data sets were generated, and for each dataset, the regression coefficient was estimated in three methods: the naive method where the predicator was substituted by the average of its multiple measurements, the regression calibration where the unobserved covariates were replaced by their conditional expectation based on the observed surrogates, the conditional intensity estimator, derived by Tsiatis & Davidian (2001) via the conditional intensity of the counting process based on the sufficient statistic of the unobserved true covariate, the parametric maximum likelihood estimation and the imputed PLS method. When carrying out the parametric MLE and the imputed PLS estimation, we assume a linear spline model (19) on the baseline hazard with three knots (corresponding to time 0-, 33- and 67-percentile of the observed failures times) and a normal distribution on the unobserved covariate.

The averages of the estimates, the empirical standard errors were calculated. We reported the results when the baseline hazard follows (19) in Table 1. We used the mean-squared error (MSE) to summarize the performance of each estimator. The imputed PLS method is quite robust to the specification of the underlying baseline hazard, and successfully corrected biases. It gave highly efficient estimates compared with the conditional intensity estimator. For example, with a sample size of 100 (censoring proportion 30 Per cent) and a log linear baseline hazard, the relative bias and standard error for $\hat{\beta}$ calculated by the Imputed PLS method was 1 Per cent and 0.196 (MSE = 0.039), compared with those of 1.2 Per cent and 0.274 (MSE =

Table 1. *Comparisons of imputed PLS estimates, conditional intensity estimates and naive estimates*

n	Censor (%)	Imputed PLS			Conditional intensity			Naive			Regression calibration			Parametric MLE		
		$\hat{\beta}_I$	SE_e	MSE	$\hat{\beta}_C$	SE_e	MSE	$\hat{\beta}_N$	SE_e	MSE	$\hat{\beta}_R$	SE_e	MSE	$\hat{\beta}_L$	SE_e	MSE
50	30	1.007	0.302	0.091	1.193	0.594	0.390	0.782	0.218	0.095	0.951	0.263	0.072	0.999	0.217	0.057
	70	1.058	0.347	0.124	1.134	0.551	0.322	0.810	0.253	0.100	0.991	0.309	0.096	0.975	0.315	0.100
100	30	0.990	0.196	0.039	1.012	0.274	0.075	0.756	0.138	0.078	0.926	0.175	0.036	0.972	0.169	0.035
	70	1.021	0.241	0.058	1.101	0.339	0.125	0.774	0.191	0.088	0.929	0.212	0.050	0.937	0.202	0.045
200	30	0.969	0.135	0.019	1.021	0.187	0.035	0.736	0.108	0.081	0.915	0.117	0.021	0.981	0.115	0.021
	70	0.966	0.164	0.028	1.030	0.222	0.050	0.767	0.122	0.069	0.918	0.138	0.026	0.980	0.149	0.026

The true baseline hazard follows the log linear model (19).
The true regression coefficient $\tilde{\beta}_0 = 1$.
MSE: mean-squared error; SE_e: empirical standard error.

0.075) calculated by the conditional intensity approach. Similar patterns persisted across different sample sizes, censoring proportions and baseline hazards. Noticeably when the sample size is relatively small, we found the imputed PLS method outperformed the conditional intensity approach significantly in terms of efficiency and bias correction. This is because the unbiasedness of the latter estimator relies on asymptotic approximation, and, hence, is only valid when the sample size is large. We also notice that the imputed PLS performed much better than the regression calibration and parametric MLE.

We next examined the robustness of the proposed imputed PLS estimators with respect to the parametric assumptions made on the distributions underlying covariate and the measurement error, when computing the imputed PLS estimates. We generated true X_i and measurement error U_{ij} from a normal mixture

$$F = \pi N\{-(1-\pi)\mu, \sigma^2\} + (1-\pi)N\{\pi\mu, \sigma^2\}, \tag{20}$$

where π is a constant between 0 and 1. This distribution has mean 0 and variance $\theta = \pi(1-\pi)\mu^2 + \sigma^2$. Two cases for F were considered: unimodal normal mixture, that is, $\pi = 0.25$, $\mu = 0.5$, $\sigma^2 = \theta - \pi(1-\pi)\mu^2$, and bimodal normal mixture, that is $\pi = 0.50$, $\mu = 1$, $\sigma^2 = \theta - \pi(1-\pi)\mu^2$. By appropriate choice of θ, we let $\text{var}(X_i) = 1$ and $\text{var}(U_{ij}) = 0.5$. In the calculation of the imputed PLS, we however assumed normal distributions on X_i and U_{ij}. Because of space limitation, we only reported the results when the X_i and U_{ij} follow the bimodal normal mixture distributions in Table 2. Both the imputed PLS method and the conditional intensity approach gave consistent results, in contrast with the regression calibration and parametric MLE. Again, we found the imputed PLS method outperformed the conditional intensity approach in terms of efficiency. For instance, with a sample size of 100 (censoring proportion 30 Per cent), covariates and measurement errors following bimodal normal mixtures, and a Weibull baseline hazard, the relative bias and standard error for $\hat{\beta}$ calculated by the imputed PLS method was 1.7 Per cent and 0.180 (MSE = 0.033), compared with those of 8.4 Per cent and 0.322 (MSE = 0.111) calculated by the conditional intensity approach. Similar patterns presented with varied sample sizes and censoring proportions.

Finally, to explore robustness of imputed PLS estimates with respect to choices of knots, we computed the imputed PLS estimates using the linear spline model with three knots (time 0-, 33- and 67-percentiles of the observed failures times), with four knots (time 0, the first, second and third quartiles of the observed failures times) and with five knots (time 0-, 20-, 40-, 60-, 80-percentiles of the observed failures times); to examine the performance of the bootstrap variance estimator, we obtained the bootstrap standard error estimates using the procedure outlined in section 6 and compared the average of them with the empirical standard errors. We generated the data using the Weibull baseline hazard ($\lambda_0(t) = 0.5t$) and the normal covariates and measurement errors. We varied sample size from 50 to 200 and censoring proportion from 30 Per cent to 70 Per cent, and the results were listed in Table 3. Under all these scenarios, the imputed PLS method gave fairly consistent estimates and were robust to the specification of the knots. For example, with a sample size of 200 (censoring proportion 70 Per cent), the relative bias, the empirical and the bootstrap standard errors for $\hat{\beta}$ calculated using three knots were -0.9 Per cent, 0.161 and 0.152, compared with those of 1.8 Per cent, 0.152 and 0.147 (four knots) and -0.1 Per cent, 0.158 and 0.158 (five knots), respectively. Noticeably the bootstrap standard errors agreed well with the empirical standard errors.

8. Application

Clinical oncologists have become increasingly interested in assessing the role of various genetic markers in predicting patient survival and response to treatment. The Eastern Cooperative

Table 2. *Comparisons of imputed PLS estimates, conditional intensity estimates and naive estimates*

		Imputed PLS			Conditional intensity			Naive			Regression calibration			Parametric MLE		
n	Censor (%)	$\hat{\beta}_I$	SE_e	MSE	$\hat{\beta}_C$	SE_e	MSE	$\hat{\beta}_N$	SE_e	MSE	$\hat{\beta}_R$	SE_e	MSE	$\hat{\beta}_L$	SE_e	MSE
50	30	0.954	0.272	0.076	1.099	0.454	0.216	0.755	0.227	0.112	0.967	0.271	0.075	0.960	0.246	0.062
	70	1.016	0.325	0.106	1.088	0.587	0.353	0.816	0.273	0.108	1.044	0.333	0.113	0.912	0.308	0.103
100	30	0.983	0.180	0.033	1.084	0.322	0.111	0.752	0.152	0.085	0.933	0.194	0.042	0.912	0.168	0.036
	70	0.987	0.235	0.055	1.084	0.333	0.118	0.777	0.179	0.082	0.947	0.230	0.056	0.918	0.205	0.049
200	30	0.957	0.138	0.021	1.028	0.191	0.037	0.753	0.099	0.071	0.921	0.115	0.019	0.912	0.117	0.021
	70	1.002	0.160	0.026	1.049	0.218	0.050	0.757	0.122	0.074	0.942	0.145	0.024	0.940	0.147	0.025

The true covariates and measurement errors were generated from bimodal normal mixtures. The true baseline hazard follows the Weibull model.
The true regression coefficient $\tilde{\beta}_0 = 1$.
SE_e: empirical standard error.

Table 3. *Robustness of imputed PLS estimates with respect to choices of knots and the performance of the bootstrap variance estimator*

		Three knots			Four knots			Five knots		
n	Censor (%)	$\hat{\beta}$	SE_e	SE_b	$\hat{\beta}$	SE_e	SE_b	$\hat{\beta}$	SE_e	SE_b
50	30	0.986	0.279	0.268	1.029	0.276	0.273	1.032	0.279	0.278
	70	1.057	0.368	0.371	1.013	0.341	0.345	1.042	0.373	0.364
100	30	0.984	0.180	0.181	1.007	0.186	0.181	1.009	0.199	0.185
	70	1.006	0.235	0.221	1.043	0.252	0.224	1.011	0.225	0.231
200	30	0.975	0.137	0.122	0.980	0.134	0.126	0.979	0.133	0.129
	70	0.988	0.161	0.151	0.996	0.168	0.155	1.005	0.158	0.156

The true baseline hazard follows the Weibull model.
The true regression coefficient $\tilde{\beta}_0 = 1$.
SE_e: empirical standard error; SE_b: bootstrap standard error.

Oncology Group recently published a study (Augenlicht *et al.*, 1997) designed to assess associations between expression of the c-*myc* oncogene and the disease free survival and overall survival of patients treated for early stage colon cancer. We considered a subset of the cases from this clinical trial, which was also coordinated by the North Central Cancer Treatment Group. In this subset, disease progression free survival and overall survival from date of study entry was measured for a total of 92 patients randomized to receive either surgery alone or surgery plus chemotherapy, namely, Levamisole. Figs 2 and 3 give the progression free and overall survival comparisons by treatment and c-*myc* expression level. It appears that patients with higher expression of the c-*myc* gene might have an enhanced response to treatment. Statistically, then, the goal was to assess whether there was a c-*myc* effect and/or a treatment/c-*myc* interaction. Complicating the analysis was that the expression level of c-*myc* gene could not be assessed precisely, and multiple measurements had to be taken if possible. The 92 patients had a total of 124 measurements on c-*myc*, with a range of 1–6 measurements per person. The variability in the number of replicates was caused by vriation in the size of available tissues.

Let X be the true c-*myc* expression level. We assumed that X follows a normal model with mean μ_x and variance σ_x^2. The observed replicate, W, also follows a normal model with the residual variance equal to $\sigma_x^2 + \sigma_u^2$, where σ_u^2 is the measurement error variance. In our calculation, the replicates of c-*myc* expression level, W, were log transformed so as to make the normality assumption more plausible. Using the moment estimating equations given in section 5, we estimate that $\mu_x = 0.366$, $\sigma_x^2 = 0.173$, and $\sigma_u^2 = 0.047$.

We fitted two survival models for progression free survival and overall survival, separately, using the proposed imputed PLS approach. Covariates of interest included the true (log transformed) c-*myc* value, treatment (coded by TRT, with 0 = Surgery alone and 1 = Surgery+ Chemotherapy) and their interaction. For comparison, we refitted the models using the naive method, i.e. replacing X with the mean of its replicate in the model. The results are presented in Tables 4 and 5. It appears that, in both progression free survival and overall survival models, the magnitude of c-*myc* and its interaction with the treatment increased after the measurement error in c-*myc* level was taken into account, compared with the naive method. For example, in the progression free survival model, the point estimate for the main effect of c-*myc* increased from 0.747 (SE = 0.43, $p = 0.083$) by the naive method to 0.905 (SE = 0.50, $p = 0.070$) by the imputed PLS method; while that of the interaction effect of c-*myc* with treatment changed from −1.298 (SE = 0.71, $p = 0.069$) to −1.613 (SE = 0.90, $p = 0.073$); similar patterns were seen in the overall survival model.

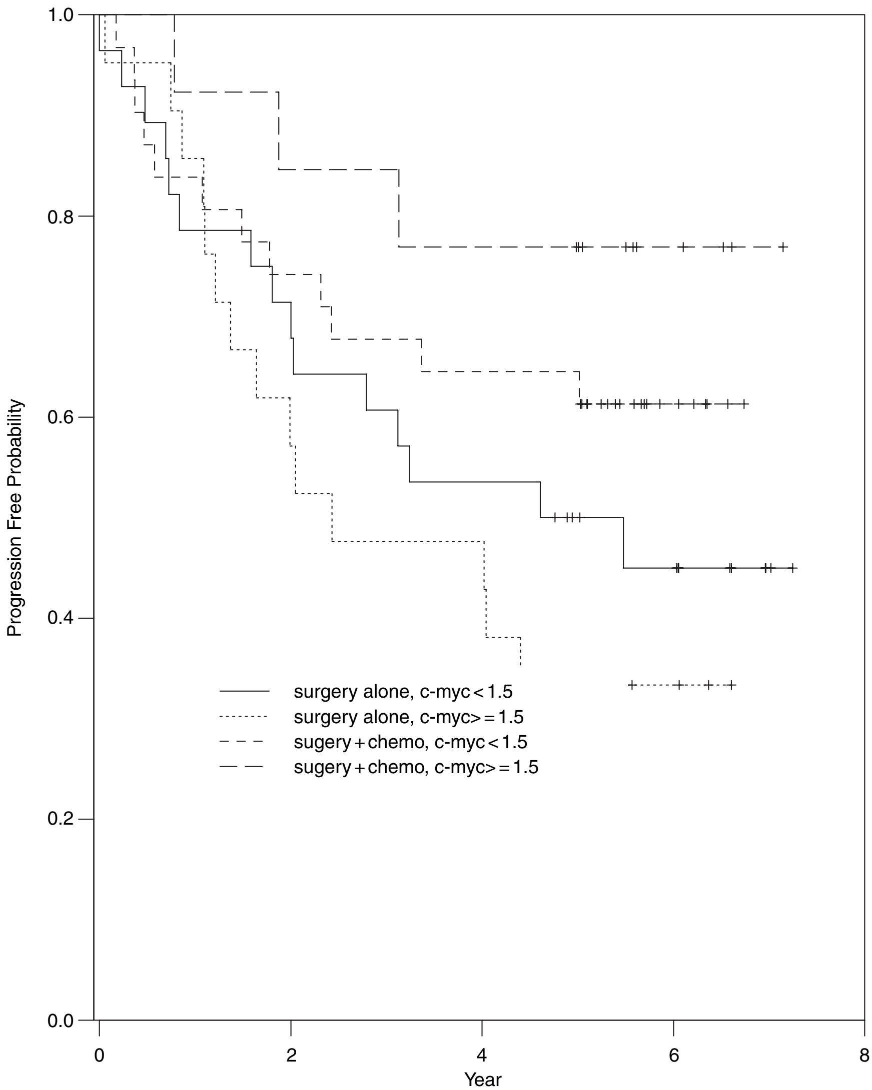

Fig. 2. Comparison of progression free survival by treatment arm and c-*myc*.

9. Discussion

In this article, we have extended the Cox partial likelihood approach to fit survival models with covariate measurement errors. Our key idea is to impute the unobserved covariates based on their conditional distributions, for which purpose, a linear spline model is assumed on the baseline hazard. We estimate the regression coefficients by solving the average PLS equations. Simulations have indicated high efficiency of the resulting estimates. Despite the dependence of our estimating equations on the parametric structure of the baseline hazard and the distribution of unobserved true covariates, analytic considerations and simulations have also revealed that the estimation of regression coefficients is quite robust to possible deviations.

An alternative strategy would be the regression calibration approach (Prentice, 1982), which replaces the unobserved true covariate in the partial likelihood score by its conditional

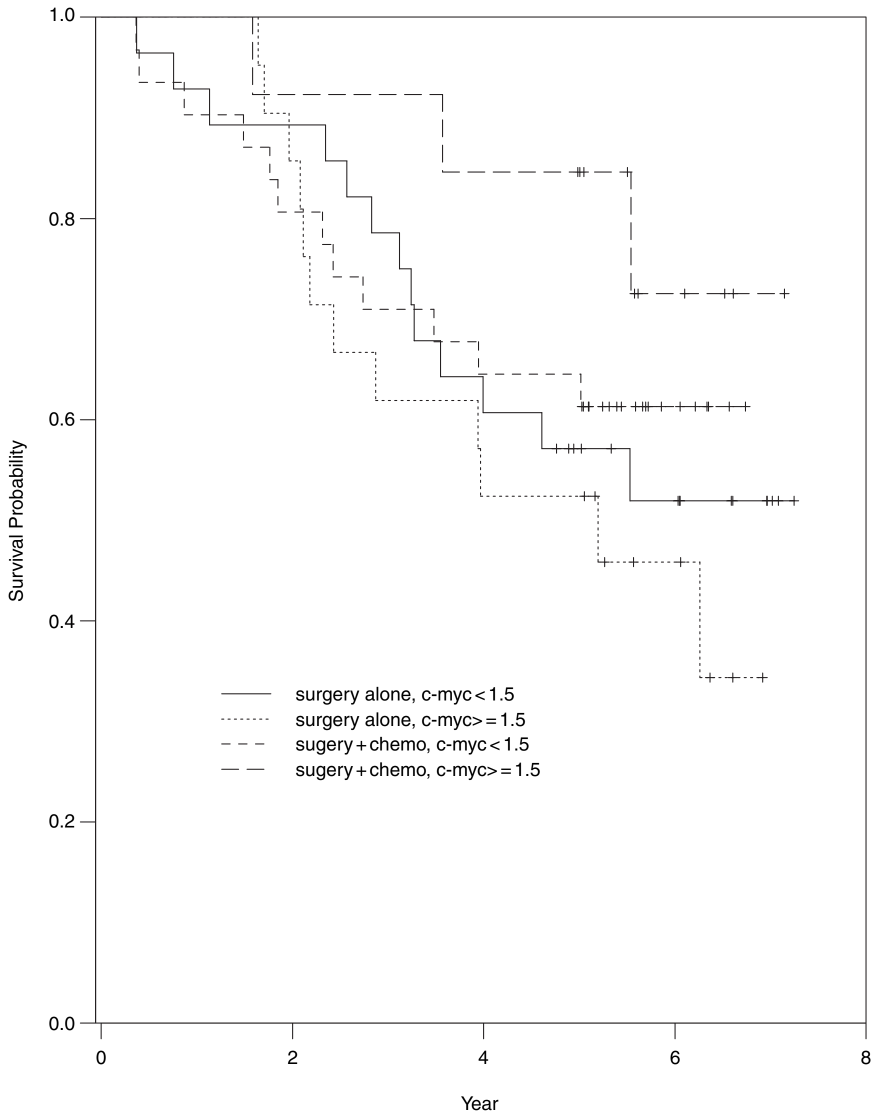

Fig. 3. Comparison of overall survival by treatment arm and c-*myc*.

Table 4. *Results of the progression free survival model for the* c-*myc study*

Covariates	Naive	Imputed PLS
c-*myc*	0.747 (0.43)	0.905 (0.50)
TRT	−0.295 (0.39)	−0.208 (0.44)
c-*myc* × TRT	−1.298 (0.71)	−1.613 (0.90)

Estimates were calculated by the Naive (ignoring measurement error) and the imputed method. Numbers inside the parentheses are estimated SEs.

expectation, given the the observed quantities at each risk set. This approach generally requires the availability of a validation data set, in which gold standard measurement are available on a subset of study subjects (see, e.g. Wang *et al.*, 2001). Recently, Xie *et al.* (2001) have developed a calibration procedure that is applicable to the reliability sample situation.

Table 5. *Analysis results of the* c-*myc data (overall survival)*

Covariates	Naive	imputed PLS
c-*myc*	0.582 (0.47)	0.711 (0.52)
TRT	−0.105 (0.40)	−0.021 (0.45)
c-*myc* × TRT	−1.126 (0.72)	−1.416 (0.90)

Estimates were calculated by the Naive (ignoring measurement error) and the imputed method. Numbers inside the parentheses are estimated SEs.

In our setting, we considered the case where replicate measurements were available on the covariates of interest. But the proposed method should extend easily to settings, where validation data samples are available.

With a slight modification, we can also extend the proposed methodology to analyse recurrent event data. In particular, (3) remains valid except that $N(t)$ will be treated as the number of events observed up to time t. The conditional likelihood of (8) should be modified to accommodate the recurrent data [cf. eq. (6.1.1) in Andersen *et al.*, 1993, Ch. 6]. All the other procedures will be intact.

Acknowledgements

The authors wish to thank the editor, an associate editor and three referees for helpful comments. This work was supported in part by NIH Grants.

References

Aitchison, J. & Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to constraints. *Ann. Math. Statist.* **29**, 813–828.

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

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Spinger-Verlag, New York.

Augenlicht, L. H., Wadler, S., Corner, G., Richards, C., Ryan, L., Multani, A. S., Pathak, S., Benson, A., Haller, D. & Heerdt, B. G. (1997). Low-level c-*myc* amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Res.* **57**, 1769–1775.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore.

Cai, T., Hyndman, R. J. & Wand, M. P. (2002). Mixed model-based hazard estimation. *J. Comput. Graph. Statist.* **11**, 784–798.

Carroll, R., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear 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.

Datta, S., Satten, G. A. & Williamson, J. M. (2000). Consistency and asymptotic normality of estimators in a regression model with interval censoring and left truncation. *Ann. Inst. Statist. Math.* **52**, 160–172.

Efron, B. (1981). Nonparametric estimates of standard error: the Jacknife, the Bootstrap and other. *Biometrika* **68**, 589–599.

Efron, B. & Tibshirani, R. (1993). *An introduction to the Bootstrap*. Chapman and Hall, New York.

Fleming, T. R. & Harrington, D. P. (1991). *Counting processes and survival analysis*. Wiley, New York.

Gu, M. G. & Zhu, H. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *J. Roy. Statist. Soc. Ser. B* **63**, 339–355.

Hu, P., Tsiatis, A. A. & Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–1419.

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

Lange, K. (1999). *Numerical analysis for statisticians*. Springer-Verlag, New York.
Li, Y., Ryan, L., Bellamy, S. & Satten, G. (2003). Inference on clustered survival data using imputed frailties. *J. Comput. Graph. Statist.* **12**, 640–662.
Lin, D. Y. & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **89**, 1074–1078.
McLachlan, G. J. & Krishnan, T. (1997). *The EM algorithm and extension*. John Wiley and Sons, New York.
Nakamura, T. (1992). Proportional hazards model 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.
Reid, N. & Crépeau, H. (1985). Influence functions for the proportional hazards regression. *Biometrika* **72**, 1–9.
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley, New York.
Rubin, D. B. & Schenker, N. (1991). Multiple imputation in health-care databases: an overview and some applications. *Statist. Med.* **10**, 585–598.
Satten, G., Datta, S. & Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval censored data. *J. Amer. Statist. Assoc.* **93**, 318–327.
Tsiatis, A. A., De Gruttola, V. & Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90**, 27–37.
Tsiatis, A. A. & Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.
Wang, C. Y., Xie, S. X. & Prentice, R. L. (2001). Recalibration based on an approximate relative risk estimator in Cox regression with missing covariates. *Statist. Sinica* **11**, 1081–1104.
Xie, S. X., Wang, C. Y. & Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *J. Roy. Statist. Soc. Ser. B* **63**, 855–870.

Received January 2003, in final form October 2004

Yi Li, Department of Biostatistics, Harvard University, Boston, MA 02494, USA.
E-mail: yili@jimmy.harvard.edu

Appendix

Technical details

Regularity Conditions. For the ith subject, we introduce the score function with respect to the conditional density of the unobserved covariates as follows:

$$\begin{aligned}\mathbf{U}_{x_i,\gamma}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma) &= \frac{\partial}{\partial \boldsymbol{\gamma}} \log\{f(\mathbf{X}_i | T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma)\} \\ &= \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log\{f(\mathbf{X}_i | T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma)\}, \frac{\partial}{\partial \boldsymbol{\eta}} \log\{f(\mathbf{X}_i | T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i; \gamma\}\right] \\ &= [\mathbf{U}_{x_i,\boldsymbol{\beta}}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma), \mathbf{U}_{x_i,\boldsymbol{\eta}}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma)]. \qquad (21)\end{aligned}$$

Denote by $\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)$, $\mathbf{U}_{x,\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)$ and $\mathbf{U}_{x,\boldsymbol{\eta}}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)$, the sums of the corresponding terms in (21) over subjects. We further denote a ρ-neighbourhood of γ by $\mathcal{N}_\rho(\gamma) = \{\gamma' \in \mathcal{B} : ||\gamma' - \gamma|| < \rho\}$, where $||\cdot||$ denotes an Euclidean norm. With the notations introduced above and those established in sections 2 and 3, we stipulate the following regularity conditions:

(C1) The sequences (T_i, δ_i, $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{W}_i$, n_i) are i.i.d.
(C2) The sequences $\{\frac{\partial \boldsymbol{\Psi}}{\partial \gamma}\}$, $\{\frac{\partial \mathbf{U}}{\partial \gamma}\}$, $\{\boldsymbol{\xi}\mathbf{U}_{x,\gamma}\}$, $\{\frac{\partial \mathbf{U}_{x,\gamma}}{\partial \gamma}\}$ and $\{\frac{\partial \xi}{\partial \boldsymbol{\beta}}\}$ each satisfy the Uniform Weak Law of Large Numbers (UWLLN) conditions at γ_0 (e.g. Satten *et al.*, 1998).

(C3) The expectation matrix, $\mathbf{A}$, of the Jacobian matrix of the score equations (12) is invertible, where $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ and

$$\mathbf{A}_{11} = \mathbf{Q}(\boldsymbol{\beta}_0) - \mathrm{E}\{\boldsymbol{\xi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)\mathbf{U}_{x_i,\boldsymbol{\beta}}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)'\},$$

$$\mathbf{A}_{12} = \mathrm{E}\{\boldsymbol{\xi}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)\mathbf{U}_{x_i,\boldsymbol{\eta}}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)'\},$$

$$\mathbf{A}_{21} = \mathrm{E}\left\{\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)'\right\},$$

$$\mathbf{A}_{22} = \mathrm{E}\left\{\frac{\partial}{\partial \boldsymbol{\eta}}\mathbf{U}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma_0)'\right\},$$

$$\mathbf{Q}(\boldsymbol{\beta}_0) = \int \left\{\frac{s^{(2)}(t; \boldsymbol{\beta}_0)}{s^{(0)}(t; \boldsymbol{\beta}_0)} - \frac{s^{(1)}(t; \boldsymbol{\beta}_0)^{\otimes 2}}{s^{(0)}(t; \boldsymbol{\beta}_0)^2}\right\} \mathrm{d}G(t),$$

and all the expectations involved are taken under the true parameter γ_0.

Proofs of Theorems 1 and 2. Along the line of Datta *et al*. (2000) and Li *et al*. (2003), we first give two lemmas leading to the proofs of theorems 1 and 2.

Lemma 1

As $m \to \infty$, for each $K > 0$, $m^{-1/2}\mathbb{S}(\gamma) = m^{-1/2}\tilde{\mathbb{S}}(\gamma) + o_p(1)$ uniformly in $\mathcal{N}_{Km^{-\frac{1}{2}}}(\gamma_0)$, where $\tilde{\mathbb{S}}(\gamma) = \sum_{i=1}^m \boldsymbol{\psi}(T_i, \delta_i, \mathbf{W}_i, \mathbf{Z}_i; \gamma)$.

Proof. Consider $R_m(\boldsymbol{\beta}) = m^{-1/2}\{\mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) - \tilde{\mathbf{S}}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})\}$, where

$$\tilde{\mathbf{S}}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) = \sum_{i=1}^{m} \boldsymbol{\xi}(T_i, \delta_i, \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\beta}).$$

Then

$$\begin{aligned} m^{-1/2}\{\mathbb{S}(\gamma) - \tilde{\mathbb{S}}(\gamma)\} &= \int R_m(\boldsymbol{\beta})\, \mathrm{d}F(\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma) \\ &= \int R_m(\boldsymbol{\beta}) \exp\{\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma^*)'(\gamma - \gamma_0)\}\, \mathrm{d}F(\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma_0), \end{aligned}$$

where γ^* lies on the line segment connecting γ and γ_0. For an arbitrary function $H(\mathbf{x})$: $R^{r+q} \to R$, denote its suprema in $\mathcal{N}_{Km^{-1/2}}(\gamma_0)$ by $\bigvee H(\gamma) = \sup_{\gamma \in \mathcal{N}_{Km^{-1/2}}(\gamma_0)} H(\gamma)$.

Then $\bigvee |m^{-1/2}\{\mathbb{S}(\gamma) - \tilde{\mathbb{S}}(\gamma)\}| \leq C_1 C_2$, where

$$C_1 = \exp\{K \cdot \bigvee ||m^{-1/2}\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)||\} \tag{22}$$

and $C_2 = \int \bigvee\{|R_m(\boldsymbol{\beta})|\}\, \mathrm{d}F(\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \gamma_0)$.

Notice that

$$\begin{aligned} \bigvee ||m^{-1/2}\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)|| &\leq ||m^{-1/2}\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma_0)|| \\ &\quad + K \cdot \bigvee ||m^{-1}\frac{\partial}{\partial \gamma}\mathbf{U}_{x,\gamma}(\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}, \mathbf{X}; \gamma)||. \end{aligned}$$

The first term is $O_p(1)$ by applying the central limit theorem, while the second term on the right-hand side of the inequality converges to $K||\mathrm{E}\left\{\frac{\partial}{\partial \gamma}\mathbf{U}_{x,\gamma}(T_i, \delta_i, \bar{\mathbf{W}}_i, \mathbf{Z}_i, \mathbf{X}_i; \gamma)|_{\gamma=\gamma_0}\right\}||$ by the UWLLN condition; hence $C_1 = O_p(1)$.

To estimate the magnitude of C_2, we consider the integrand $\bigvee|R_m(\boldsymbol{\beta})|$. We observe that

$$\bigvee |R_m(\boldsymbol{\beta})| = \bigvee m^{-1/2}|\mathbf{S}(\mathbf{T},\boldsymbol{\Delta},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}) - \tilde{\mathbf{S}}(\mathbf{T},\boldsymbol{\Delta},\mathbf{X},\mathbf{Z};\boldsymbol{\beta})|$$
$$\leq |R_m(\boldsymbol{\beta}_0)| + K\cdot\bigvee ||m^{-1}\mathcal{I}(\mathbf{T},\boldsymbol{\Delta},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}) - \mathbf{Q}(\boldsymbol{\beta}_0)||$$
$$+ K\cdot\bigvee || - m^{-1}\frac{\partial}{\partial\boldsymbol{\beta}}\xi(T_i,\delta_i,\mathbf{X}_i,\mathbf{Z}_i;\boldsymbol{\beta}) - \mathbf{Q}(\boldsymbol{\beta}_0)||. \tag{23}$$

Following the decomposition of the partial likelihood score (Lin & Wei, 1989), we may find $R_m(\boldsymbol{\beta}_0) = o_p(1)$ uniformly with respect to $\mathbf{X}$. Recalling that $\mathcal{I}(\mathbf{T},\ \boldsymbol{\Delta},\ \mathbf{X},\ \mathbf{Z};\ \boldsymbol{\beta})$ is the partial likelihood information (given the true $\mathbf{X}$) and applying theorems 3.2 and 4.2 of Andersen & Gill (1982), one may show that the second term in (23) is $o_p(1)$ uniformly with respect to $\mathbf{X}$. With the UWLLN conditions on $\partial\boldsymbol{\xi}/\partial\boldsymbol{\beta}$, the third term in (23) is $o_p(1)$ uniformly with respect to $\mathbf{X}$ as well. Thus, we have been able to show that $C_2 = o_p(1)$. Therefore, $C_1C_2 = o_p(1)$, which proves the lemma.

We next show in lemma 2, that the $\mathbf{A}_{ij}^m(\gamma)$, the components of the Jocobian matrix of the score equations (12), defined in (28), are consistent estimators of the $\mathbf{A}_{ij}$ in a small neighbourhood of γ_0.

Lemma 2

As $m\to\infty$, for each $K>0$,

$$\sup_{\gamma\in\mathcal{N}_{Km^{-\frac{1}{2}}}(\gamma_0)}||\mathbf{A}_{ij}^m(\gamma) - \mathbf{A}_{ij}|| = o_p(1).$$

Proof. As shown in lemma 1,

$$\bigvee ||m^{-1}\mathcal{I}(\mathbf{T},\boldsymbol{\Delta},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}) - \mathbf{Q}(\boldsymbol{\beta}_0)|| = o_p(1).$$

By a similar calculation as before,

$$Z_m = \bigvee ||\int \{m^{-1}\mathcal{I}(\mathbf{T},\boldsymbol{\Delta},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}) - \mathbf{Q}(\boldsymbol{\beta}_0)\}\ \mathrm{d}F(\mathbf{X}|\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z};\gamma)||$$
$$\leq C_1\int D_m\,\mathrm{d}F(\mathbf{X}|\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z};\gamma_0), \tag{24}$$

where C_1 is as in (22) and $D_m = \bigvee||m^{-1}\mathcal{I}(\mathbf{T},\ \boldsymbol{\Delta},\ \mathbf{X},\ \mathbf{Z};\ \boldsymbol{\beta}) - \mathbf{Q}(\boldsymbol{\beta}_0)||$. Hence, $D_m = o_p(1)$. As D_m is bounded, by the dominated convergence theorem, $\int D_m\,\mathrm{d}F(\mathbf{X}|\mathbf{T},\ \boldsymbol{\Delta},\ \bar{\mathbf{W}},\ \mathbf{Z};\ \gamma_0)\to 0$. Hence, $Z_m = o_p(1)$. In a way analogous to the proof in lemma 1, one can also establish that

$$\bigvee ||\int R_m(\boldsymbol{\beta})\mathbf{U}_{x,\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z},\mathbf{X};\boldsymbol{\beta},\boldsymbol{\eta})\ \mathrm{d}F(\mathbf{X}|\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z};\gamma)|| = o_p(1). \tag{25}$$

On the other hand,

$$m^{-1}\int\sum_{i=1}^m \xi(T_i,\delta_i,\mathbf{X}_i,\mathbf{Z}_i;\gamma)\mathbf{U}_{x,\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z},\mathbf{X};\boldsymbol{\beta},\boldsymbol{\eta})\ \mathrm{d}F(\mathbf{X}|\mathbf{T},\boldsymbol{\Delta},\bar{\mathbf{W}},\mathbf{Z};\gamma)$$
$$= m^{-1}\sum_{i=1}^m\int \xi(T_i,\delta_i,\mathbf{X}_i,\mathbf{Z}_i;\gamma)\mathbf{U}_{x_i,\boldsymbol{\beta}}(T_i,\delta_i,\bar{\mathbf{W}}_i,\mathbf{Z}_i,\mathbf{X}_i;\gamma)\ \mathrm{d}F(x_i|T_i,\delta_i,\bar{\mathbf{W}}_i,\mathbf{Z}_i;\gamma)$$
$$\xrightarrow{p} \mathrm{E}\{\boldsymbol{\xi}(T_i,\delta_i,\mathbf{X}_i,\mathbf{Z}_i;\gamma)\mathbf{U}_{x_i,\boldsymbol{\beta}}(T_i,\delta_i,\bar{\mathbf{W}}_i,\mathbf{Z}_i,\mathbf{X}_i;\gamma)\}, \tag{26}$$

uniformly in $\mathcal{N}_{Km^{-1/2}}(\gamma_0)$ by the UWLLN conditions. Thus, combining (24)–(26), we finish the proof of the uniform convergence of the $\mathbf{A}_{11}^m$. Similarly, we can obtain the convergence for $\mathbf{A}_{12}^m$.

Convergence of $\mathbf{A}_{21}^m$ and $\mathbf{A}_{22}^m$ follows from the standard maximum likelihood score argument and the UWLLN conditions.

With lemmas 1 and 2 established, we can prove consistency and asymptotic normality of the estimators.

Proof of Theorem 1. Let $\mathbf{P}(\gamma) = \{\mathbb{S}(\gamma), \mathbb{U}(\gamma)\}$ and assume that $\mathbf{A}$ is positive definite, otherwise we can replace $\mathbf{P}(\gamma)$ with $\mathbf{A}'\mathbf{P}(\gamma)$. A standard Taylor expansion gives that

$$m^{-1/2}\mathbf{P}(\gamma) = m^{-1/2}\mathbf{P}(\gamma_0) - \mathbf{A}_m(\gamma^*)m^{1/2}(\gamma - \gamma_0),$$

where γ^* lies between γ_0 and γ.

By lemma 1 and the central limit theorem, $m^{-1/2}\mathbf{P}(\gamma_0)$ converges to a mean 0 random normal variable. Hence $m^{-1/2}\mathbf{P}(\gamma_0) = O_p(1)$. Let $\epsilon > 0$ be arbitrary. Then for sufficiently large m_{01}, when $m > m_{01}$, on a set with probability $1 - (1/2)\epsilon$, $||m^{-1/2}\mathbf{P}(\gamma_0)|| < J$, where $J < \infty$. By lemma 2, there exists an $m_{02} > 0$ such that when $m > m_{02}$, on a set with probability $1 - (1/2)\epsilon$, $\mathbf{A}_m(\gamma)$ converges uniformly to $\mathbf{A}$ in $\gamma \in \mathcal{N}_{Km^{-1/2}}(\gamma_0)$, where K is any positive numbers. Let $m_0 = \max(m_{01}, m_{02})$. We then work on the intersection of the two random sets (with probability at least $1 - \epsilon$. Now we fix any $m > m_0$. Denote by $\lambda_{\min}$ the minimum eigenvalue of $\mathbf{A}$. Then for $K_0 = 2J/\lambda_{\min}$, one can show $||(\gamma - \gamma_0)'\mathbf{P}(\gamma)|| \geq m^{-1/2}(\lambda_{\min}K_0^2 - JK_0) > 0$ for $||\gamma - \gamma_0|| = K_0 m^{-1/2} = 2(J/\lambda_{\min})m^{-(1/2)}$. Since $\mathbf{P}(\gamma)$ is continuous in γ, by the fixed point theorem of Aitchison & Silvey (1958, lemma 2), $\mathbf{P}(\gamma)$ has a solution in $||\gamma - \gamma_0|| < 2(J/\lambda_{\min})m^{-1/2}$.

Proof of Theorem 2. With $\mathbf{P}(\widehat{\gamma}) = 0$, expanding it about γ_0 gives that

$$\mathbf{A}_m(\gamma^*)m^{1/2}(\widehat{\gamma} - \gamma_0) = m^{-1/2}\mathbf{P}(\gamma_0), \tag{27}$$

where γ^* lies between $\widehat{\gamma}$ and γ_0. By the proof of theorem 1, for any $\epsilon > 0$, there exists a $K_0 > 0$ and m such that the event $\{||\widehat{\gamma} - \gamma_0|| < K_0 m^{-1/2}\}$ has measure at least $1 - \epsilon$. Hence, by lemma 2 $\mathbf{A}_m(\gamma^*) \xrightarrow{p} \mathbf{A}$. Using lemma 1 and a central limit theorem, one obtains that

$$m^{-1/2}\mathbf{P}(\gamma_0) \xrightarrow{d} N(0, \mathbf{\Psi}).$$

Hence, theorem 2 follows from (27) by the Slutsky theorem.

Variance Estimator. We can show $\mathbf{A}$ can be consistently estimated by $\mathbf{A}_m(\widehat{\gamma})$, where $\widehat{\gamma}$ is the solution to (12) and $\mathbf{A}_m(\gamma)$ is the Jacobian matrix of the score equations (12). It is given by

$$\mathbf{A}_m(\boldsymbol{\beta}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{A}_{11}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) & \mathbf{A}_{12}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) \\ \mathbf{A}_{21}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) & \mathbf{A}_{22}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) \end{pmatrix} = -\frac{1}{m}\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}}\mathbb{S}(\boldsymbol{\beta}, \boldsymbol{\eta}) & \frac{\partial}{\partial \boldsymbol{\eta}}\mathbb{S}(\boldsymbol{\beta}, \boldsymbol{\eta}) \\ \frac{\partial}{\partial \boldsymbol{\beta}}\mathbb{U}(\boldsymbol{\beta}, \boldsymbol{\eta}) & \frac{\partial}{\partial \boldsymbol{\eta}}\mathbb{U}(\boldsymbol{\beta}, \boldsymbol{\eta}) \end{pmatrix}, \tag{28}$$

where $\mathbf{A}_{21}^m$ and $\mathbf{A}_{22}^m$ are easily obtained by differentiating (11) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, while $\mathbf{A}_{11}^m$ and $\mathbf{A}_{12}^m$ are given by

$$\mathbf{A}_{11}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{1}{m}\int \left\{\mathcal{I}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) - \mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})\mathbf{U}_{x,\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta})\right\} \mathrm{d}F(\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta})$$

and

$$\mathbf{A}_{12}^m(\boldsymbol{\beta}, \boldsymbol{\eta}) = -\frac{1}{m}\int \mathbf{S}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta})\mathbf{U}_{x,\boldsymbol{\eta}}(\mathbf{T}, \boldsymbol{\Delta}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta})\, \mathrm{d}F(\mathbf{X}|\mathbf{T}, \boldsymbol{\Delta}, \bar{\mathbf{W}}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\eta}).$$

To develop a consistent estimator for $\mathbf{\Psi}$, we begin with $\boldsymbol{\xi}(T_i, \delta_i, \mathbf{X}_i, \mathbf{Z}_i; \gamma)$. For each $t \in (0, \tau)$, we consider consistent estimates for $s^{(0)}(t; \boldsymbol{\beta}) = \mathrm{E}(Y_i(t)\, \mathrm{e}^{\boldsymbol{\beta}_x'\mathbf{X}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i})$, $\mathbf{s}^{(1)}(t; \boldsymbol{\beta}) = \{\mathbf{s}_x^{(1)}(t; \boldsymbol{\beta}), \mathbf{s}_z^{(1)}(t; \boldsymbol{\beta})\} = \mathrm{E}\{Y_i(t)\mathbf{X}_i\, \mathrm{e}^{\boldsymbol{\beta}_x'\mathbf{X}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i}, Y_i(t)\mathbf{Z}_i\, \mathrm{e}^{\boldsymbol{\beta}_x'\mathbf{X}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i}\}$. To proceed, denote by $S^{(0)} = (1/m)\sum Y_i(t)\, \mathrm{e}^{\boldsymbol{\beta}_x'\bar{\mathbf{W}}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i}$, $\mathbf{S}_w^{(1)} = (1/m)\sum Y_i(t)\bar{\mathbf{W}}_i\, \mathrm{e}^{\boldsymbol{\beta}_x'\bar{\mathbf{W}}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i}$ and $\mathbf{S}_z^{(1)} = (1/m)\sum Y_i(t)\bar{\mathbf{Z}}_i\, \mathrm{e}^{\boldsymbol{\beta}_x'\bar{\mathbf{W}}_i + \boldsymbol{\beta}_z'\mathbf{Z}_i}$. Note

$$\mathrm{E}\Big\{Y_i(t)\,\mathrm{e}^{\boldsymbol{\beta}_x'\bar{\mathbf{W}}_i\boldsymbol{\beta}_z'\mathbf{Z}_i}\Big\} = \mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x s^{(0)}(t;\boldsymbol{\beta})},$$

$$\mathrm{E}\Big\{Y_i(t)\mathbf{W}_i\,\mathrm{e}^{\boldsymbol{\beta}_x'\bar{\mathbf{W}}_i\boldsymbol{\beta}_z'\mathbf{Z}_i}\Big\} = \mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x}\mathbf{s}_x^{(1)}(t;\boldsymbol{\beta}) + n_i^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x\,\mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x}s^{(0)}(t;\boldsymbol{\beta})$$

$$\mathrm{E}\Big\{Y_i(t)\mathbf{Z}_i\,\mathrm{e}^{\boldsymbol{\beta}_x'\mathbf{Z}_i\boldsymbol{\beta}_z'\mathbf{Z}_i}\Big\} = \mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x}\mathbf{s}_z^{(1)}(t;\boldsymbol{\beta}).$$

Hence, $s^{(0)}(t;\,\boldsymbol{\beta}), \mathbf{s}^{(1)}(t;\,\boldsymbol{\beta}) = \{\mathbf{s}_x^{(1)}(t;\,\boldsymbol{\beta}), \mathbf{s}_z^{(1)}(t;\,\boldsymbol{\beta})\}$ can be consistently estimated by $\hat{s}^{(0)}(t;\,\boldsymbol{\beta}) = S^{(0)}/M_0,\ \widehat{\mathbf{s}}^{(1)}(t;\,\boldsymbol{\beta}) = \{(\mathbf{S}_w^{(1)} - S^{(0)}\mathbf{M}_1)/M_0,\ \mathbf{S}_z^{(1)}/M_0\}$, where $M_0 = (1/m)\sum_{i=1}^m \times \mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x}$, $\mathbf{M}_1 = (1/m)\sum_{i=1}^m n_i^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x\,\mathrm{e}^{n_i^{-1}\frac{1}{2}\boldsymbol{\beta}_x'\boldsymbol{\Sigma}_u\boldsymbol{\beta}_x}$ and $\boldsymbol{\Sigma}_u$ can be replaced by a moment estimator $\hat{\boldsymbol{\Sigma}}_u$ in section 5.

It follows that, with the same argument in Lin & Wei (1989), $\boldsymbol{\xi}$ can be estimated by

$$\widehat{\boldsymbol{\xi}}(T_i,\delta_i,\mathbf{X}_i,\mathbf{Z}_i;\boldsymbol{\beta},\boldsymbol{\eta}) = \delta_i\left\{\mathbf{Z}_i^* - \frac{\widehat{s}^{(1)}(T_i;\boldsymbol{\beta})}{\widehat{s}^{(0)}(T_i;\boldsymbol{\beta})}\right\} - \sum_{i'}\frac{N_{i'}(T_i)\,\mathrm{e}^{\mathbf{Z}_i^*\boldsymbol{\beta}}}{m\cdot\widehat{s}^{(0)}(T_{i'};\boldsymbol{\beta})}\left\{\mathbf{Z}_i^* - \frac{\widehat{s}^{(1)}(T_{i'};\boldsymbol{\beta})}{\widehat{s}^{(0)}(T_{i'};\boldsymbol{\beta})}\right\}.$$

Note that $\widehat{\boldsymbol{\xi}}(\cdot)$ resembles the influence function for the 'complete' data proportional hazards model (Reid & Crépeau, 1985).

Because each $\boldsymbol{\Psi}(T_i,\,\delta_i,\,\bar{\mathbf{W}}_i,\,\mathbf{Z}_i;\,\gamma)$ is the expectation (with respect to $\mathbf{X}_i$ conditional on $T_i,\,\delta_i,\,\bar{\mathbf{W}}_i$ and $\mathbf{Z}_i$) of $\boldsymbol{\xi}(T_i,\,\delta_i,\,\mathbf{X}_i,\,\mathbf{Z}_i;\,\gamma)$, it can be consistently estimated by the conditional expectation of $\widehat{\boldsymbol{\xi}}(T_i,\,\delta_i,\,\mathbf{X}_ic,\,\mathbf{Z}_i;\,\gamma)$, which is

$$\widehat{\boldsymbol{\Psi}}(T_i,\delta_i,\bar{\mathbf{W}}_i,\mathbf{Z}_i;\gamma) = \delta_i\left\{\begin{pmatrix} g_i(\boldsymbol{\beta}_x;\gamma) \\ \mathbf{Z}_i\end{pmatrix} - \frac{\widehat{s}^{(1)}(T_i;\boldsymbol{\beta})}{\widehat{s}^{(0)}(T_i;\boldsymbol{\beta})}\right\} - g_i(\boldsymbol{\beta}_x;\gamma)\,\mathrm{e}^{\mathbf{Z}_i'\boldsymbol{\beta}_z}\left[\sum_{i'}\frac{N_{i'}(T_i)}{m\cdot\widehat{s}^{(0)}(T_{i'};\boldsymbol{\beta})}\left\{\mathbf{Z}_i^{**} - \frac{\widehat{s}^{(1)}(T_{i'};\boldsymbol{\beta})}{\widehat{s}^{(0)}(T_{i'};\boldsymbol{\beta})}\right\}\right],$$

where $g_i(\mathbf{s};\,\gamma) = \int \mathrm{e}^{\mathbf{s}'\mathbf{X}_i}\,\mathrm{d}F(\mathbf{X}_i|T_i,\,\delta_i,\,\bar{\mathbf{W}}_i,\,\mathbf{Z}_i;\,\gamma)$ and $g_i^{(1)}(\mathbf{s};\,\gamma)$ is its first derivative with respect to s, and $\mathbf{Z}_i^{**} = \left\{g_i^{(1)}(\boldsymbol{\beta}_x;\,\gamma)/g_i(\boldsymbol{\beta}_x;\,\gamma),\,\mathbf{Z}_i\right\}'$.

Since $\widehat{\boldsymbol{\Psi}}(T_i,\,\delta_i,\,\bar{\mathbf{W}}_i,\,\mathbf{Z}_i;\,\widehat{\gamma})$ is an estimate of the contribution of the ith subject to the score $\mathbb{S}(\widehat{\gamma})$, the matrix $\boldsymbol{\Psi}$ can be estimated by $\widehat{\boldsymbol{\Psi}} = \begin{pmatrix}\widehat{\boldsymbol{\Psi}}_{11} & \widehat{\boldsymbol{\Psi}}_{12} \\ \widehat{\boldsymbol{\Psi}}_{12}' & \widehat{\boldsymbol{\Psi}}_{22}\end{pmatrix}$, where

$$\widehat{\boldsymbol{\Psi}}_{11} = \frac{1}{m}\sum_{i=1}^m \widehat{\boldsymbol{\Psi}}_i\widehat{\boldsymbol{\Psi}}_i', \qquad \widehat{\boldsymbol{\Psi}}_{12} = \frac{1}{m}\sum_{i=1}^m \widehat{\boldsymbol{\Psi}}_i\widehat{\mathbf{U}}_i', \qquad \widehat{\boldsymbol{\Psi}}_{22} = \frac{1}{m}\sum_{i=1}^m \widehat{\mathbf{U}}_i\widehat{\mathbf{U}}_i',$$

and $\widehat{\boldsymbol{\Psi}}_i$ and $\widehat{\mathbf{U}}_i$ are the abbreviations of $\widehat{\boldsymbol{\Psi}}(T_i,\,\delta_i,\,\bar{\mathbf{W}}_i,\,\mathbf{Z}_i;\,\widehat{\gamma})$ and $\widehat{\mathbf{U}}(T_i,\,\delta_i,\,\bar{\mathbf{W}}_i,\,\mathbf{Z}_i;\,\widehat{\gamma})$, respectively. Hence, the asymptotic variance of $m^{1/2}(\widehat{\gamma} - \gamma_0)$, $\mathbf{V}$, can be estimated by (18).