Statistica Sinica **9**(1999), 953-969

CONSISTENT ESTIMATION IN COX PROPORTIONAL HAZARDS MODEL WITH COVARIATE MEASUREMENT ERRORS

Fan Hui Kong and Minggao Gu

Westat and McGill University

Abstract: The regular maximum partial likelihood estimator is biased when the covariates in the Cox proportional hazards model are measured with error, unless the measurement errors tend to zero. Although several alternative estimators have been proposed, theoretical justifications for them are lacking. We try to fill this gap by showing that the corrected maximum partial likelihood estimator proposed by Nakamura (1992) is consistent and has an asymptotic normal distribution. A consistent estimator of its variance is derived as well. We also show that the corrected baseline hazard estimator proposed by Kong, Huang and Li (1998) is consistent and converges to a Gaussian process. Furthermore, we obtain the estimators for more general measurement error models where the errors are not normally distributed. Simulations are performed to show the accuracy of the variance estimator.

Key words and phrases: Cox proportional hazards model, estimating equation, hazard function, measurement error model, partial likelihood, unbiased score.

1. Introduction

The Cox proportional hazards model is the most popular model describing the relationship between risk factors and survival time. Yet it is quite common that some or even all of the covariates are measured with error or are misspecified. A common consequence of such measurement error or misspecification is that the parameter estimations are attenuated, which means that the estimates shrink toward zero. A more serious consequence is that such error could violate the correct model relationship. In an example studying the survival time of women with breast cancer in the San Francisco Bay Area, Gong, Whittemore and Grosser (1990) have pointed out that a valid proportional hazards model between survival time and cancer stage was violated because of cancer stage misspecification.

A typical approach to correct the parameter estimation is based on induced hazard rates, which are defined as the conditional hazards given the observed covariates and the condition that the event has not occurred (Prentice (1982)). Such a procedure is believed to give consistent induced maximum partial likelihood estimators (induced MPLEs) and was further explored by Pepe, Self and Prentice (1989), Tsiatis, DeGruttola and Wulfsohn (1995), Wulfson and Tsiatis

(1997) and Zhou and Pepe (1995). However, so far there is no effective way to derive this estimator.

Another approach, which comes from a general strategy developed by Stefanski (1989) and Nakamura (1990), is to construct an unbiased score function. Knowing that the naive score is biased, these authors constructed a corrected score such that, given the true covariates, its expectation under the measurement error distribution is the same as the true score; thus, a corrected maximum likelihood estimator is derived. However, as Stefanski (1989) has argued, such a corrected score does not exist for the partial likelihood score. Therefore, Nakamura (1992) introduced an approximately corrected partial likelihood score and derived an estimator that was less biased than the naive MPLE, as shown through simulations. A similar approach was taken by Buzas (1996).

The regression calibration approach was introduced to failure time analysis by Wang, Hsu, Feng and Prentice (1997). As in logistic regression (Rosner, Willett and Spiegelman (1989)), this method generates a small bias in estimating the parameters. However, it can be applied in quite general circumstances and is easy to implement.

Also as a general approach, Kong (1999) modified the naive MPLE by deducting its estimated bias. This approach produces less biased estimators with existing software, and the new estimates can be further improved. At the same time, Kong, Huang and Li (1998) provided a corrected cumulative baseline hazard estimator by modifying the Breslow estimator.

In this paper we show that Nakamura's approach produces a consistent estimator as sample size increases, without requiring that the error magnitude be small. Under quite general conditions, this estimator also has an asymptotic normal distribution, with the variance consistently estimated by a sandwich estimator (Theorem 3.1). Examples of sandwich estimators can be found in Liang and Zeger (1986), Lin and Wei (1989) and White (1982). At the same time, we show that the corrected cumulative baseline hazard estimator is consistent and follows the Central Limit Theorem under general conditions (Theorem 4.1).

Suppose that the hazard rate $\lambda(t)$ for survival time T follows the Cox (1972) proportional hazards model

$$\lambda\{t; z(t)\} = \lambda_0(t) \exp\{\beta' z(t)\}, \tag{1.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard, z is a possibly time-varying p-vector covariate and β is a p-vector regression parameter. Usually, the observed data for n subjects are n triplets $\{U_i, Z_i(\cdot), \delta_i\}$, where $U_i = \min(T_i, C_i)$ is the event time, C_i is the censoring time, $Z_i(\cdot)$ is the true covariate process and $\delta_i = I(T_i \leq C_i)$, with $I(\cdot)$ being an indicator function. When covariate measurement error is involved, covariate $Z_i(\cdot)$ becomes unobservable. The observed covariate is actually the true covariate plus a measurement error. We assume an additive measurement error of the form:

$$X_i(t) = Z_i(t) + \epsilon_i(t), \quad \text{for} \quad i = 1, \ldots, n, \tag{1.2}$$

for a stationary Gaussian process $\epsilon_i(t)$ independent of $Z_i(t)$. The main assumption on $\epsilon_i(t)$ is that it has mean zero and a variance-covariance matrix D that is not dependent on time t. It is not necessary for $\epsilon_i(t)$ to be normally distributed. But for ease of notation, we derive the theorems for normally distributed $\epsilon_i(t)$ and give the general formulas in Section 5. As a necessary requirement, a consistent estimator $\hat{D}$ of D is assumed to exist. Usually, such estimators can be derived using validation data (e.g. Carroll, Ruppert and Stefanski (1995), Chapter 3).

2. Notation

The arguments in this paper apply mainly to functional models but can easily be modified for structural models. For functional models, we assume that the regularity conditions in Andersen and Gill (1982) hold. For structural models where $\{T_i, C_i, Z_i(\cdot), \epsilon_i(\cdot)\}$, $i = 1, \ldots, n$, are i.i.d. random vectors, we make necessary assumptions so that the Laws of Large Numbers holds. We also assume that T_i, C_i are independent given the true covariate $Z_i(\cdot)$ for $i = 1, \ldots, n$. Denoting $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa'$ for any column vector a, we introduce the following notations:

$$S^{(r)}(\beta, t, Z) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) Z_i(t)^{\otimes r} e^{\beta' Z_i(t)}, \quad r = 0, 1, 2, \quad \text{for } Y_i(t) = I(U_i \geq t),$$

$$E(\beta, t, Z) = \frac{S^{(1)}(\beta, t, Z)}{S^{(0)}(\beta, t, Z)}, \quad \text{and} \quad V(\beta, t, Z) = \frac{S^{(2)}(\beta, t, Z)}{S^{(0)}(\beta, t, Z)} - E(\beta, t, Z)^{\otimes 2}.$$

Also let

$$s^{(r)}(\beta, t) = \mathrm{E}\{S^{(r)}(\beta, t, Z)\} \quad \text{for} \quad r = 0, 1, 2,$$

where the expectation is taken with respect to the true distribution of (T, C, Z). Under the above assumptions for both functional and structural models, as n increases, $S^{(r)}(\beta, t, Z) \to s^{(r)}(\beta, t)$ for $r = 0, 1, 2$, so that $E(\beta, t, Z) \to e(\beta, t)$ and $V(\beta, t, Z) \to v(\beta, t)$, where

$$e(\beta, t) = \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \quad \text{and} \quad v(\beta, t) = \frac{s^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - e(\beta, t)^{\otimes 2}.$$

Here and later, all assertions of consistency are in probability. For the true parameter β_0 and true covariate $Z_i(t)$, define the counting process as $N_i(t) = \delta_i I(U_i \leq t)$ and the martingale process as

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp\{\beta_0' Z_i(s)\} \lambda_0(s) ds, \quad \text{for} \quad i = 1, \ldots, n. \tag{2.1}$$

With $\bar{N}(t) = \sum_1^n N_i(t)$, the score statistic based on true covariates is defined as

$$\mathcal{U}_n(\beta, t, Z) = \sum_{i=1}^{n} \int_0^t \{Z_i(s) - E(\beta, s, Z)\} dN_i(s). \tag{2.2}$$

For observed covariates $X_i(t), i = 1, \ldots, n$, in (1.2), $S^{(r)}(\beta, t, X)$, $r = 0, 1, 2$, $E(\beta, t, X)$, $V(\beta, t, X)$ and $\mathcal{U}_n(\beta, t, X)$ are defined simply by replacing $Z(t)$ with $X(t)$ in the above definitions. Under observed values of $\{X_i, U_i, \delta_i\}$, $i = 1, \ldots, n$, the observed score $\mathcal{U}_n(\beta, t, X)$ is biased, which means that $\mathcal{U}_n(\beta, t, X)$ does not have mean zero at β_0. By defining the corrected score as

$$\tilde{\mathcal{U}}_n(\beta, t, X) = \sum_{i=1}^{n} \int_0^t \{X_i(s) - E(\beta, s, X) + D\beta\} dN_i(s) \tag{2.3}$$

for $t \leq \tau \leq \infty$, Nakamura's corrected MPLE is the $\hat{\beta}$ that satisfies

$$\tilde{\mathcal{U}}_n(\hat{\beta}, \tau, X) = 0. \tag{2.4}$$

The corrected information matrix is

$$\tilde{I}_n(\hat{\beta}, X) = -\nabla_\beta \tilde{\mathcal{U}}_n(\hat{\beta}, \tau, X), \tag{2.5}$$

where ∇_β is the derivative with respect to β.

To estimate the cumulative baseline hazard function $\Lambda_0(t)$, Kong, Huang and Li (1998) proposed a corrected cumulative baseline hazard estimator of the form

$$\hat{\Lambda}_0(t, X) = \exp\{\frac{1}{2}\hat{\beta}'\hat{D}\hat{\beta}\} \int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\hat{\beta}, s, X)}, \tag{2.6}$$

where $\hat{\beta}$ is an estimator of β and $\hat{D}$ is an estimator of the measurement error variance-covariance matrix derived from validation data or repeated measurements (e.g. Carroll, Ruppert and Stefanski (1995), Chapter 3). To derive a corrected cumulative baseline hazard estimator, $\hat{\beta}$ in (2.6) can be any estimator rather than just Nakamura's corrected MPLE. But to show the consistency of $\hat{\Lambda}_0(t, X)$, we require that both $\hat{\beta}$ and $\hat{D}$ be consistent. To derive the asymptotic normality of $\hat{\Lambda}_0(t, X)$, we require that both $\hat{D}$ and $\hat{\beta}$ be consistent and that $\hat{\beta}$ follow the Central Limit Theorem. However, to avoid confusion in this paper, we assume that $\hat{\beta}$ is Nakamura's corrected MPLE, as defined in (2.4). It is consistent as well as asymptotically normally distributed, as shown in the next section.

3. Corrected Maximum Partial Likelihood Estimator

The main result for the corrected MPLE, $\hat{\beta}$, is given in Theorem 3.1. However, for the sake of convenience, we begin with a lemma.

Lemma 3.1. *For $Z_i(t), \epsilon_i(t), X_i(t)$ and $Y_i(t)$, $i = 1, \dots, n$, as defined in Section 2, let*

$$Q_0(\beta, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)} \{ e^{\beta' \epsilon_i(t) - \frac{1}{2}\beta' D \beta} - 1 \}$$

and

$$Q_1(\beta, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)} \Big[X_i(t) e^{\beta' \epsilon_i(t) - \frac{1}{2}\beta' D \beta} - \{ Z_i(t) + D\beta \} \Big],$$

for $t \le \tau$ and measurement error covariance matrix D. Then for $S^{(r)}(\beta, t, X)$ and $S^{(r)}(\beta, t, Z)$, $r = 0, 1$, as defined in Section 2, the following relationships hold:

$$S^{(0)}(\beta, t, X) = e^{\frac{1}{2}\beta' D \beta} \Big\{ S^{(0)}(\beta, t, Z) + \frac{1}{\sqrt{n}} Q_0(\beta, t) \Big\} \tag{3.1}$$

and

$$S^{(1)}(\beta, t, X) = e^{\frac{1}{2}\beta' D \beta} \Big\{ S^{(1)}(\beta, t, Z) + D\beta S^{(0)}(\beta, t, Z) + \frac{1}{\sqrt{n}} Q_1(\beta, t) \Big\}. \tag{3.2}$$

Proof. Because the proofs for (3.1) and (3.2) are similar, we prove only (3.1). In fact, using (1.2), $S^{(0)}(\beta, t, X)$ can be written as

$$S^{(0)}(\beta, t, X) = e^{\frac{1}{2}\beta' D \beta} S^{(0)}(\beta, t, Z) + \frac{1}{n} \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)} \{ e^{\beta' \epsilon_i(t)} - e^{\frac{1}{2}\beta' D \beta} \}, \tag{3.3}$$

which is (3.1).

Because $Q_0(\beta, t)$ and $Q_1(\beta, t)$ are sums of independent random variables with mean zero, both are asymptotically normally distributed under the Lindeberg condition. We assume that the Lindeberg condition holds here and for other sums of independent random variables given throughout this paper. To derive the asymptotic variance of the corrected MPLE, let

$$\begin{aligned} C_{1i}(\hat\beta, X) = & \int_0^\tau \Big\{ X_i(s) - \frac{S^{(1)}(\hat\beta, s, X)}{S^{(0)}(\hat\beta, s, X)} + D\hat\beta \Big\} dN_i(s) \\ & - \int_0^\tau \Big\{ X_i(s) - \frac{S^{(1)}(\hat\beta, s, X)}{S^{(0)}(\hat\beta, s, X)} \Big\} \frac{Y_i(s) e^{\hat\beta' X_i(s)}}{S^{(0)}(\hat\beta, s, X)} \frac{d\bar N(s)}{n} \end{aligned} \tag{3.4}$$

and

$$\tilde J_n(\hat\beta, X) = \frac{1}{n} \sum_{i=1}^{n} C_{1i}(\hat\beta, X)^{\otimes 2}.$$

Then Theorem 3.1 follows.

Theorem 3.1. *For the Cox proportional hazards model* (1.1) *and the additive measurement error model* (1.2)*, assume that the regularity conditions in Anderson and Gill* (1982) *hold. Then, the corrected MPLE, $\hat{\beta}$, is a consistent estimator of β_0 as $n \to \infty$. At the same time, $n^{1/2}(\hat{\beta} - \beta_0)$ has an asymptotic normal distribution with mean zero and a variance-covariance matrix consistently estimated by*

$$\tilde{I}_n(\hat{\beta}, X)^{-1}\tilde{J}_n(\hat{\beta}, X)\tilde{I}_n(\hat{\beta}, X)^{-1},$$

where $\tilde{I}_n(\hat{\beta}, X)$ is the corrected information matrix defined in (2.5)*, and $\tilde{J}_n(\hat{\beta}, X)$ is defined by* (3.4).

Comment. For the special case of no measurement error, $\tilde{\mathcal{U}}_n(\beta, t, X) = \mathcal{U}_n(\beta, t, Z)$, so that Nakamura's corrected MPLE becomes a regular MPLE and $\tilde{I}_n(\hat{\beta}, X)$ becomes the observed Fisher information matrix. The variance estimator is not reduced to the inverse of the observed Fisher information. However, it is still a consistent estimator of the variance of $\hat{\beta}$. In fact, it becomes the robust variance estimator derived by Lin and Wei (1989), given that the Cox model is not misspecified.

Proof. In order to show the consistency of $\hat{\beta}$, we must derive the limit of $\tilde{\mathcal{U}}_n(\beta, \tau, X)/n$ for any β as in Andersen and Gill (1982). Note that for any $0 \le t \le \tau$, by (2.3),

$$\begin{aligned}\frac{1}{n}\tilde{\mathcal{U}}_n(\beta, t, X) = &\frac{1}{n}\sum_{i=1}^{n}\int_0^t \{X_i(s) - E(\beta, s, X) + D\beta\} dM_i(s)\\ &+\frac{1}{n}\sum_{i=1}^{n}\int_0^t \{X_i(s) - E(\beta, s, X) + D\beta\} Y_i(s) e^{\beta_0' Z_i(s)}\lambda_0(s)ds, \quad (3.5)\end{aligned}$$

where β_0 is the true parameter and $M_i(t)$ is a local-square integrable martingale as defined in (2.1). The first term of (3.5) is a local square-integrable martingale. Dividing (3.2) by (3.1) and noting that $\frac{1}{\sqrt{n}}Q_r(\beta, t) \to 0$ for $r = 0, 1$ as $n \to \infty$, we find

$$E(\beta, t, X) = E(\beta, t, Z) + D\beta + o_p(1). \quad (3.6)$$

By calculating the variance function, we can show that the first term of (3.5) converges to zero. Therefore, with $X_i(t)$ replaced by $Z_i(t) + \epsilon_i(t)$, (3.5) becomes

$$\begin{aligned}\frac{1}{n}\tilde{\mathcal{U}}_n(\beta, \tau, X) = &\int_0^\tau S^{(1)}(\beta_0, s, Z)\lambda_0(s)ds + \int_0^\tau \frac{1}{n}\sum_{i=1}^{n} Y_i(s)\epsilon_i(s) e^{\beta_0' Z_i(s)}\lambda_0(s)ds\\ &-\int_0^\tau E(\beta, s, Z) S^{(0)}(\beta_0, s, Z)\lambda_0(s)ds + o_p(1).\end{aligned}$$

The second term is a sum of independent random variables with mean zero and a variance that can be easily derived. Therefore, as $n \to \infty$, under regularity conditions the second term goes to zero and the remainder converges to

$$\int_0^\tau \Big[s^{(1)}(\beta_0, s) - e(\beta, s)s^{(0)}(\beta_0, s)\Big]\lambda_0(s)ds,$$

for $s^{(r)}(\beta, s)$, $r = 0, 1$, and $e(\beta, s)$ as defined in Section 2. Under the regularity conditions, this limiting function becomes zero at $\beta = \beta_0$, and its derivative is a negative function of β. Therefore, using the argument in Andersen and Gill (1982), we have shown the consistency of $\hat{\beta}$.

To derive the limiting distribution of $\hat{\beta}$, we apply the Mean Value Theorem to $\tilde{\mathcal{U}}_n(\beta, \tau, X)$ and write

$$n^{1/2}(\hat{\beta} - \beta_0) = \Big[\frac{1}{n}\tilde{I}_n(\beta^*, X)\Big]^{-1}\frac{1}{\sqrt{n}}\tilde{\mathcal{U}}_n(\beta_0, \tau, X) \tag{3.7}$$

for β^* between β_0 and $\hat{\beta}$. First consider $\frac{1}{\sqrt{n}}\tilde{\mathcal{U}}_n(\beta_0, t, X)$ for $t \le \tau$, which can be written as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\tilde{\mathcal{U}}_n(\beta_0, t, X) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \{X_i(s) - E(\beta_0, s, X) + D\beta_0\}dM_i(s) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \{X_i(s) - E(\beta_0, s, X) + D\beta_0\}Y_i(s)e^{\beta_0' Z_i(s)}\lambda_0(s)ds \\
&= A_1 + A_2.
\end{aligned}$$

With the aid of expressions (3.1), (3.2) and (3.6), $E(\beta, t, X)$ can be written as

$$\begin{aligned}
E(\beta, t, X) &= \frac{E(\beta, t, Z) + D\beta + \frac{1}{\sqrt{n}}Q_1(\beta, t)/S^{(0)}(\beta, t, Z)}{1 + \frac{1}{\sqrt{n}}Q_0(\beta, t)/S^{(0)}(\beta, t, Z)} \\
&= E(\beta, t, Z) + D\beta + \frac{1}{\sqrt{n}}Q_1(\beta, t)/S^{(0)}(\beta, t, Z) \\
&\quad -\{E(\beta, t, Z) + D\beta\}\frac{1}{\sqrt{n}}Q_0(\beta, t)/S^{(0)}(\beta, t, Z) + o_p(\frac{1}{\sqrt{n}}). \qquad (3.8)
\end{aligned}$$

With $E(\beta_0, t, Z)$ approximated by $e(\beta_0, t)$, A_1 becomes

$$\begin{aligned}
A_1 =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \{X_i(s) - e(\beta_0, s)\}dM_i(s) \\
& - \int_0^t \frac{1}{S^{(0)}(\beta_0, s, Z)}\Big[Q_1(\beta_0, s) - \{E(\beta_0, s, Z) + D\beta_0\}Q_0(\beta_0, s)\Big]\frac{d\bar{M}(s)}{n} + o_p(1),
\end{aligned}$$

with $\bar{M}(t) = \sum_1^n M_i(t)$. Because both $Q_0(\beta, t)$ and $Q_1(\beta, t)$ have finite limiting distributions and both $E(\beta_0, t, Z)$ and $S^{(0)}(\beta_0, t, Z)$ have finite limits, by calculating the variance function we can show that the second term converges to zero. Similarly, substituting (3.8) into A_2, we find that

$$\begin{aligned}
A_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \Big\{X_i(s) - E(\beta_0, s, Z)\Big\} Y_i(s) e^{\beta_0' Z_i(s)} \lambda_0(s) ds \\
&\quad - \int_0^t [Q_1(\beta_0, s) - \{E(\beta_0, s, Z) + D\beta_0\} Q_0(\beta_0, s)] \lambda_0(s) ds + o_p(1) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \Big[\Big\{X_i(s) - E(\beta_0, s, Z) - D\beta_0\Big\} e^{\beta_0' \epsilon_i(s) - \frac{1}{2}\beta_0' D \beta_0} \\
&\quad - \{X_i(s) - E(\beta_0, s, Z)\}\Big] Y_i(s) e^{\beta_0' Z_i(s)} \lambda_0(s) ds + o_p(1) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \Big[\Big\{X_i(s) - e(\beta_0, s) - D\beta_0\Big\} e^{\beta_0' \epsilon_i(s) - \frac{1}{2}\beta_0' D \beta_0} \\
&\quad - \{X_i(s) - e(\beta_0, s)\}\Big] Y_i(s) e^{\beta_0' Z_i(s)} \lambda_0(s) ds + o_p(1).
\end{aligned}$$

Notice that the replacement of $E(\beta_0, s, Z)$ by $e(\beta_0, s)$ in the last equation causes an error of $o_p(1)$. Combining A_1 and A_2 yields

$$\begin{aligned}
\frac{1}{\sqrt{n}} \tilde{\mathcal{U}}_n(\beta_0, \tau, X) = &\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{X_i(s) - e(\beta_0, s)\} dN_i(s) \\
&- \frac{e^{-\frac{1}{2}\beta_0' D \beta_0}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \Big\{X_i(s) - e(\beta_0, s) - D\beta_0\Big\} Y_i(s) e^{\beta_0' X_i(s)} \lambda_0(s) ds \\
&+ o_p(1),
\end{aligned} \tag{3.9}$$

which is a sum of independent random variables, each with mean zero. Therefore, under previous assumptions, the Central Limit Theorem applies. At the same time, $\frac{1}{n}\tilde{I}_n(\beta, X)$ can be written as follows:

$$\frac{1}{n}\tilde{I}_n(\beta, X) = -\int_0^\tau \Big\{D - \frac{S^{(2)}(\beta, s, X)}{S^{(0)}(\beta, s, X)} + E(\beta, s, X)^{\otimes 2}\Big\} \frac{d\bar{N}(s)}{n}. \tag{3.10}$$

Replacing $X_i(t)$ with $Z_i(t) + \epsilon_i(t)$ in both $S^{(0)}(\beta, t, X)$ and $S^{(2)}(\beta, t, X)$ for $i = 1, \ldots, n$, the former becomes

$$S^{(0)}(\beta, t, X) = e^{\frac{1}{2}\beta' D \beta} S^{(0)}(\beta, t, Z) + o_p(1) \tag{3.11}$$

and the latter becomes

$$S^{(2)}(\beta, t, X) = \frac{1}{n} \sum_{i=1}^n Y_i(t) Z_i(t)^{\otimes 2} e^{\beta' Z_i(t) + \beta' \epsilon_i(t)} + \frac{1}{n} \sum_{i=1}^n Y_i(t) \epsilon_i(t)^{\otimes 2} e^{\beta' Z_i(t) + \beta' \epsilon_i(t)}$$

$$+\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\{Z_i(t)\epsilon_i(t)' + \epsilon_i(t)Z_i(t)'\}e^{\beta' Z_i(t)+\beta'\epsilon_i(t)}. \tag{3.12}$$

To derive an approximation for $S^{(2)}(\beta, t, X)$, we consider the above three terms separately. The first term can be written as

$$\begin{aligned}&e^{\frac{1}{2}\beta' D\beta}\Big[S^{(2)}(\beta, t, Z) + \frac{1}{n}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes 2}e^{\beta' Z_i(t)}\{e^{\beta'\epsilon_i(t)-\frac{1}{2}\beta' D\beta} - 1\}\Big]\\ &= e^{\frac{1}{2}\beta' D\beta}S^{(2)}(\beta, t, Z) + o_p(1)\end{aligned}$$

since the n summands are independent variables, each with mean zero. Therefore, the summation converges to zero in probability. With the same technique, the second term of (3.12) becomes

$$e^{\frac{1}{2}\beta' D\beta}S^{(0)}(\beta, t, Z)(D + D\beta\beta' D) + o_p(1)$$

and the last term becomes

$$e^{\frac{1}{2}\beta' D\beta}\Big[S^{(1)}(\beta, t, Z)\beta' D + D\beta S^{(1)}(\beta, t, Z)'\Big] + o_p(1).$$

Combining these expansions of $S^{(2)}(\beta, t, X)$, we have the following expansion:

$$\frac{S^{(2)}(\beta, t, X)}{S^{(0)}(\beta, t, X)} = \frac{S^{(2)}(\beta, t, Z)}{S^{(0)}(\beta, t, Z)} + E(\beta, t, Z)\beta' D + D\beta E(\beta, t, Z)' + D + D\beta\beta' D + o_p(1)$$

for any value of β. Substituting this expression into (3.10) and using (3.6) for $E(\beta, t, X)$, it follows that

$$\begin{aligned}\frac{1}{n}\tilde{I}_n(\beta, X) &= -\int_0^{\tau}\Big\{-\frac{S^{(2)}(\beta, s, Z)}{S^{(0)}(\beta, s, Z)} + E(\beta, s, Z)^{\otimes 2}\Big\}\frac{d\bar{N}(s)}{n} + o_p(1)\\ &= \Sigma + o_p(1),\end{aligned} \tag{3.13}$$

where Σ is the information matrix with the true covariates. Letting $n \to \infty$, $\hat{\beta} \to \beta_0$ in probability for true β_0, as does β^*. Therefore,

$$\frac{1}{n}\tilde{I}_n(\hat{\beta}, X) = \Big|\frac{1}{n}\tilde{I}_n(\hat{\beta}, X) - \frac{1}{n}\tilde{I}_n(\beta, X)\Big| + \frac{1}{n}\tilde{I}_n(\beta, X) \to_p \Sigma,$$

which in turn is a consistent estimator of Σ, as is $\frac{1}{n}\tilde{I}_n(\beta^*, X)$. Because (3.9) is a sum of n independent random variables, each with mean zero, its variance is the sum of the variances of these variables, which also equals the sum of the expectations of the squares of these variables. By the Laws of Large Numbers, for any fixed β_0 and $\lambda_0(t)$, this variance can be consistently estimated by the sum of the squares of these variables. By replacing β_0 with $\hat{\beta}$ and replacing $\lambda_0(t)$

with the corrected baseline hazard estimator, this sample sum is $\tilde{J}_n(\hat{\beta}, X)$, as defined in (3.4), and converges in probability to the variance of $\frac{1}{\sqrt{n}}\tilde{\mathcal{U}}_n(\beta_0, t, X)$. As a result, $n^{1/2}(\hat{\beta} - \beta_0)$, which can be expressed as in (3.7), has an asymptotic normal distribution with mean zero and a variance consistently estimated by

$$\tilde{I}_n(\hat{\beta}, X)^{-1}\tilde{J}_n(\hat{\beta}, X)\tilde{I}_n(\hat{\beta}, X)^{-1}.$$

Here, to estimate the variance of $\frac{1}{\sqrt{n}}\tilde{\mathcal{U}}_n(\beta_0, \tau, X)$, we replace $\lambda_0(t)$ with the modified Breslow estimator, the consistency of which is given in the next section.

4. Corrected Baseline Hazard Estimator

As we saw at the end of the last section, a consistent estimator of the baseline hazard $\lambda_0(t)$ is essential for deriving a consistent variance estimator for $\hat{\beta}$. In the following we shall show that the corrected baseline hazard estimator defined in (2.6) is such an estimator.

Because the true covariate Z is not observable, the Breslow estimator for the cumulative baseline hazard function $\Lambda_0(t)$ cannot be calculated. To estimate $\Lambda_0(t)$, Kong, Huang and Li (1998) proposed the modified Breslow estimator $\hat{\Lambda}_0(t, X)$. In the following discussion, we show that this estimator is consistent and converges to a Gaussian process as $n \to \infty$. To estimate the variance function of the limiting process of $n^{1/2}\{\hat{\Lambda}_0(t, X) - \Lambda_0(t)\}$, we introduce the following notation. Let

$$\hat{H}(\hat{\beta}, t) = -\int_0^t \frac{e^{\frac{1}{2}\hat{\beta}'\hat{D}\hat{\beta}}}{S^{(0)}(\hat{\beta}, s, X)}\Big\{\frac{S^{(1)}(\hat{\beta}, s, X)}{S^{(0)}(\hat{\beta}, s, X)} + \hat{D}\hat{\beta}\Big\}\frac{d\bar{N}(s)}{n} \tag{4.1}$$

and

$$C_{2i}(\hat{\beta}, X, t) = \int_0^t \frac{e^{-\frac{1}{2}\hat{\beta}'\hat{D}\hat{\beta}}}{S^{(0)}(\hat{\beta}, s, X)}\Big\{dN_i(s) - \frac{Y_i(s)e^{\hat{\beta}'X_i(s)}}{S^{(0)}(\hat{\beta}, s, X)}\frac{d\bar{N}(s)}{n}\Big\},$$

for $i = 1, \ldots, n$. Then Theorem 4.1 follows.

Theorem 4.1. *Under the assumptions of Theorem* 3.1, $\hat{\Lambda}_0(t, X)$, *as defined in* (2.6), *is a consistent estimator of* $\Lambda_0(t)$. *Also, on* $[0, \tau]$, $n^{1/2}\{\hat{\Lambda}_0(t, X) - \Lambda_0(t)\}$ *can be expressed as a sum of independent random processes; therefore, the large sample theory applies, and it converges to a Gaussian process with mean zero and a variance function that can be consistently estimated by*

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{n\hat{H}(\hat{\beta}, t)\tilde{I}(\hat{\beta}, X)^{-1}C_{1i}(\hat{\beta}, X) + C_{2i}(\hat{\beta}, X, t)\Big\}^{\otimes 2}$$

for $0 \le t \le \tau < \infty$ *and* $C_{1i}(\hat{\beta}, X)$ *defined in* (3.4).

Comment. If the variance D of the measurement error is 0, indicating no measurement error, $\hat{\Lambda}_0(t,X)$ is reduced to the regular Breslow estimator of $\Lambda_0(t)$. Although the variance function estimator has a different form from that given by Andersen and Gill (1982), it is still a consistent estimator of $\Lambda_0(t)$, as indicated below.

Proof. Replacing β with $\hat{\beta}$ in (3.11), the equation is still valid. Putting (3.11) into $\hat{\Lambda}_0(t,X)$, by the consistency of the Breslow estimator the consistency of $\hat{\Lambda}_0(t,X)$ is obtained.

To see the limiting distribution, note that

$$\begin{aligned}n^{1/2}\{\hat{\Lambda}_0(t,X)-\Lambda_0(t)\} &= n^{1/2}\Big\{\hat{\Lambda}_0(t,X)-e^{\frac{1}{2}\beta_0' D\beta_0}\int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta_0,s,X)}\Big\}\\ &\quad +n^{1/2}\Big\{e^{\frac{1}{2}\beta_0' D\beta_0}\int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta_0,s,X)}-\Lambda_0(t)\Big\}\\ &= B_1+B_2.\end{aligned}$$

Let $H(\beta,t)$ be given by

$$\begin{aligned}&\frac{\partial}{\partial\beta}\Big\{e^{\frac{1}{2}\beta' D\beta}\int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta,s,X)}\Big\}\\ &= D\beta e^{\frac{1}{2}\beta' D\beta}\int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta,s,X)} - e^{\frac{1}{2}\beta' D\beta}\int_0^t \frac{S^{(1)}(\beta,s,X)}{\{S^{(0)}(\beta,s,X)\}^2}d\bar{N}(s)/n.\end{aligned}$$

Then B_1 becomes

$$H(\beta^*,t)n^{1/2}(\hat{\beta}-\beta_0)$$

for β^* between β_0 and $\hat{\beta}$. When $\hat{\beta}\to\beta_0$, $\beta^*\to\beta_0$. With the consistency of the corrected cumulative baseline hazard estimator, as $\beta\to\beta_0$, $H(\beta,t)$ can be approximated as

$$\begin{aligned}H(\beta,t) &= D\beta\Lambda_0(t)-\int_0^t \frac{S^{(1)}(\beta,s,X)}{S^{(0)}(\beta,s,X)}\lambda_0(s)ds+o_p(1)\\ &= D\beta\Lambda_0(t)-\int_0^t \{e(\beta,s)+D\beta\}\lambda_0(s)ds+o_p(1)\\ &= -\int_0^t e(\beta,s)\lambda_0(s)ds+o_p(1). \qquad (4.2)\end{aligned}$$

Adopting expression (3.7) for $n^{1/2}(\hat{\beta}-\beta_0)$, with (3.9) and (3.13), B_1 becomes

$$B_1 = H(\beta^*,t)\Sigma^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \{X_i(s)-e(\beta_0,s)\}dN_i(s)$$

$$
\begin{aligned}
&-\frac{e^{-\frac{1}{2}\beta_0' D\beta_0}}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{X_i(s) - e(\beta_0, s) - D\beta_0\} Y_i(s) e^{\beta_0' X_i(s)} \lambda_0(s) ds\Big] + o_p(1) \\
&= \frac{1}{\sqrt{n}} H(\beta_0, t) \Sigma^{-1} \tilde{\mathcal{U}}_n(\beta_0, \tau, X) + o_p(1). \qquad (4.3)
\end{aligned}
$$

With $H(\beta_0, t)$ replaced by (4.2), B_1 becomes a sum of independent random processes, each with mean zero. With (3.1) in Lemma 3.1, the second term of $n^{1/2}\{\hat{\Lambda}_0(t, X) - \Lambda_0(t)\}$ can be expressed as

$$
\begin{aligned}
B_2 &= n^{1/2} \Big\{ e^{\frac{1}{2}\beta_0' D\beta_0} \int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta_0, s, X)} - \int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\beta_0, s, Z)} \Big\} + n^{1/2} \int_0^t \frac{d\bar{M}(s)/n}{S^{(0)}(\beta_0, s, Z)} \\
&= -\int_0^t \frac{Q_0(\beta_0, s)}{\{S^{(0)}(\beta_0, s, Z)\}^2} d\bar{N}(s)/n + \frac{1}{\sqrt{n}} \int_0^t \frac{d\bar{M}(s)}{S^{(0)}(\beta_0, s, Z)} + o_p(1).
\end{aligned}
$$

Following the standard martingale argument in Fleming and Harrington (1991), it follows that

$$
\begin{aligned}
B_2 &= -\int_0^t \frac{\lambda_0(s)}{s^{(0)}(\beta_0, s)} \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i(s) e^{\beta_0' Z_i(s)} \{ e^{\beta_0' \epsilon_i(s) - \frac{1}{2}\beta_0' D\beta_0} - 1 \} ds \\
&\quad + \frac{1}{\sqrt{n}} \int_0^t \frac{1}{s^{(0)}(\beta_0, s)} d\bar{M}(s) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \frac{1}{s^{(0)}(\beta_0, s)} \Big[dN_i(s) - Y_i(s) e^{\beta_0' X_i(s) - \frac{1}{2}\beta_0' D\beta_0} \lambda_0(s) ds \Big] + o_p(1). \qquad (4.4)
\end{aligned}
$$

Both B_1 and B_2 are the sums of independent mean zero random processes, as is $B_1 + B_2$. With the Central Limit Theorem, the limiting process of $B_1 + B_2$ is a Gaussian process with mean zero and a variance function that can be consistently estimated by the sample variance function of the terms in (4.3) and (4.4), with unknown values replaced by their estimators.

5. General Case

The previous results are derived for normally distributed measurement errors. Similar results are valid for more general measurement error distributions when the estimators are suitably modified.

For measurement error process $\epsilon_i(t)$, $i = 1, \ldots, n$, suppose that its moment generating function $\phi(\beta) = E\{e^{\beta' \epsilon_i(t)}\}$ does not depend on time t and exists for all β. Then we define the corrected score function as

$$
\tilde{\mathcal{U}}_n(\beta, t, X) = \sum_{i=1}^n \int_0^t \Big\{ X_i(s) - E(\beta, s, X) + \frac{\nabla_\beta \phi(\beta)}{\phi(\beta)} \Big\} dN_i(s) \qquad (5.1)
$$

for $t \leq \tau \leq \infty$, where ∇_β is the derivative with respect to β. Nakamura's corrected MPLE and the corrected information matrix are defined as in Section 2, with the score function defined in (5.1). The corrected cumulative baseline hazard estimator for $\Lambda_0(t)$ is given by

$$\hat{\Lambda}_0(t, X) = \phi(\hat{\beta}) \int_0^t \frac{d\bar{N}(s)/n}{S^{(0)}(\hat{\beta}, s, X)},$$

where $\hat{\beta}$ is an estimator of β. Then, under regularity conditions, Theorems 3.1 and 4.1 are both valid, except that $D\hat{\beta}$ in the definition of $\tilde{J}_n(\hat{\beta}, X)$ in (3.5) and $\hat{H}(\hat{\beta}, t)$ in (4.1) must be replaced by $\nabla_\beta \phi(\hat{\beta})/\phi(\hat{\beta})$.

6. Simulation Results

Let $SD(Z)$ be the standard deviation of covariate Z, and σ the standard deviation of measurement error. For a fixed sample size and different values of β, $SD(Z)$ and σ, Nakamura (1992) simulated the performance of the corrected MPLE and his variance estimator. Now, knowing that his variance estimator is not consistent, we have performed a similar simulation to compare its accuracy with that of ours.

Table 1. Simulation for the corrected MPLE of β and its standard deviation estimators. M($\hat{\beta}$) is the mean of 500 independent β estimates. SD($\hat{\beta}$) is the standard deviation of these 500 β estimates. M_1(SD) is the square root of the average of 500 variance estimates using Nakamura's variance estimator. M_2(SD) is the square root of the average of 500 variance estimates using our variance estimator. In this table, $n = 400, m = 300$.

β	σ	M($\hat{\beta}$)	SD($\hat{\beta}$)	M_1(SD)	M_2(SD)
0.5	0.6	0.5210	0.0798	0.0735	0.0810
0.6	0.5	0.6197	0.0827	0.0724	0.0802
0.7	0.5	0.7137	0.0926	0.0762	0.0852
0.8	0.4	0.8034	0.0827	0.0748	0.0824
0.9	0.3	0.9130	0.0736	0.0748	0.0789
1.0	0.3	1.0019	0.0802	0.0776	0.0836
1.0	0.4	1.0075	0.0865	0.0829	0.0944
1.0	0.5	1.0162	0.1040	0.0897	0.1093
1.0	0.6	1.0297	0.1274	0.0969	0.1349
1.0	0.7	1.0413	0.1503	0.1147	0.1726
2.0	0.15	2.0230	0.1232	0.1122	0.1188
2.0	0.2	2.0375	0.1372	0.1176	0.1327
2.0	0.25	2.0561	0.1593	0.1261	0.1506
2.0	0.3	2.0545	0.1824	0.1327	0.1722
2.0	0.35	2.0659	0.2202	0.1484	0.2055

Let n and m be the sample size and the number of failures, respectively. We chose $n = 400$ and let $\mathbf{Z}$ be a set of 400 uniform random numbers in $(0, 12^{1/2})$ so that $SD(Z)$ =1. For a given β and each covariate value Z, a failure time Y is generated from the proportional hazards model with $\lambda_0(t) = 1$ and relative risk $e^{\beta Z}$. In the first part of the simulation, for $m = 100,\ 200$ and 300, we chose Type-II censoring so that all individuals after the mth failure are censored. For a fixed σ and a measurement error V generated from $N(0, 1)$, an observed covariate X is the sum of Z and σV.

Given 400 independent pairs of (X, Y), we derived a corrected MPLE and two variance estimates, one using Nakamura's method and the other using our formula. This procedure was repeated 500 times, and we derived an average of the corrected MPLEs, its sample standard deviation and the square root of the average of Nakamura's variance estimates and the square root of the average of our variance estimates. To save space, Table 1 presents only the results for $m = 100$. The results for $m = 200, 300$ are similar and are available from the authors.

Because the main results derived in this paper are based on the random censorship assumption, we performed a simulation for the same model with censoring time independently generated from a uniform distribution on $[0, d]$, where d is a positive number. For a fixed β, d can be chosen such that the percentage of censoring is a desired value. We performed a simulation for a similar sets of β, σ, as above, and the results are listed in Table 2.

The results indicate that Nakamura's corrected MPLE performs quite well for the given sample size. At the same time, for most of the situations we have simulated, our variance estimator performs better than that of Nakamura. Especially in situations where $\beta\sigma$ is large and Nakamura's estimator leads to large biases, our estimator still gives quite good results (see Table 1). When the censoring percentage increases, both variance estimators become less accurate. The improvement of our estimator also becomes less dramatic. It is interesting to note from the simulation that although the theory is developed for the Cox proportional hazards model under random censoring, Nakamura's corrected MPLE and our variance estimator both perform as well for the model under Type-II censoring.

The performance of the corrected baseline hazard estimator $\hat{\Lambda}_0(t, X)$ has been simulated and presented in Kong, Huang and Li (1998). Those results indicate that $\hat{\Lambda}_0(t, X)$ is a good estimator of $\Lambda_0(t)$. Since the variance estimator of $\hat{\beta}$ depends on $\hat{\Lambda}_0(t, X)$, the performance of the former also reflects the performance of the latter. Therefore, the fact that our variance estimator of $\hat{\beta}$ performs well in the simulation gives us a positive indication about $\hat{\Lambda}_0(t, X)$.

Table 2. Simulation for the corrected MPLE of β and its standard deviation estimators under random censoring. M($\hat{\beta}$) is the mean of 500 independent β estimates. SD($\hat{\beta}$) is the standard deviation of these 500 β estimates. M_1(SD) is the square root of the average of 500 variance estimates using Nakamura's variance estimator. M_2(SD) is the square root of the average of 500 variance estimates using our variance estimator. In this table, $n = 400$ and the percentage of failure is listed in the table.

β	σ	% of failure	M($\hat{\beta}$)	SD($\hat{\beta}$)	M_1(SD)	M_2(SD)
0.5	0.6	75%	0.5094	0.0840	0.0747	0.0819
0.5	0.5	25%	0.5115	0.1326	0.1253	0.1280
0.6	0.5	75%	0.6147	0.0899	0.0743	0.0812
0.7	0.5	50%	0.7057	0.1069	0.0937	0.1024
0.8	0.4	50%	0.8042	0.0996	0.0933	0.0990
0.9	0.3	25%	0.8909	0.1621	0.1314	0.1337
1.0	0.3	75%	1.0095	0.0979	0.0782	0.0861
1.0	0.4	75%	1.0118	0.1184	0.0836	0.0976
1.0	0.5	50%	1.0093	0.1368	0.1105	0.1281
1.0	0.6	50%	1.0257	0.1940	0.1324	0.1628
2.0	0.15	75%	2.0385	0.1364	0.1158	0.1231
2.0	0.2	75%	2.0457	0.1593	0.1216	0.1339
2.0	0.25	63%	2.0436	0.2346	0.1883	0.2109
2.0	0.3	63%	2.0409	0.2857	0.2101	0.2459
2.0	0.35	75%	2.1046	0.2520	0.1650	0.2210

7. Remarks

In this paper, we have shown that both Nakamura's corrected MPLE and our corrected cumulative baseline hazard estimator for the Cox proportional hazards model under additive measurement error are consistent and obey the Central Limit Theorem. Consistent estimators for their respective variance and variance functions have also been derived. The corrected MPLE and the corrected cumulative baseline hazard estimator are both easy to implement given the measurement error variance, which is usually estimated from validation data. But the corrected MPLE may not always exist because the derivative of the corrected score function is not always negative at a certain neighborhood of true β. This usually happens when the measurement error is large. In our simulation, if $|\sigma\beta| \geq 0.8$, the corrected MPLE often fails to converge. To derive the modified Breslow estimator, any β estimator can be used. However, the resulting hazard estimator may not be consistent unless a consistent β estimator is applied.

The formulas for estimating the variance and variance functions of the corrected MPLE and the corrected cumulative baseline hazard estimators are complicated. Especially in the latter case, in which several unknown values are replaced by their estimators, such replacements may increase the variation of the estimators and make them less accurate. However, both estimators are consistent. So if the sample size is large enough, these formulas will produce accurate estimates. In our simulation where the sample size is 400 and censoring changes from 100 to 300, our variance estimator significantly improves the accuracy of the estimator suggested by Nakamura (1992). This is especially true when Nakamura's estimator gives a large bias.

Given the complexity of deriving the variance function estimator of the corrected cumulative baseline hazard estimator, one can consider a few alternatives. In the case where the measurement error variance is small, the estimator suggested by Kong, Huang and Li (1998) can be applied. When the sample size is small and only the variances at a small number of time points are of interest, the Bootstrap (Efron and Tibshirani (1993)) method can be applied.

The estimators discussed in this paper can be applied only to additive measurement error models. In this situation, the validation data are used only for estimating the measurement error variance. For more general measurement error models, one can apply the regression calibration method (Wang, Hsu, Feng and Prentice (1997)) or the likelihood method proposed by Wulfsohn and Tsiatis (1997). In those cases, more information is needed from the validation data. The regression calibration method is easy to implement, and its limiting properties are discussed in Wang, Hsu, Feng and Prentice (1997). The implementation of the likelihood method requires a much more complicated procedure, and it is not yet clear whether it is in fact consistent.

Acknowledgements

The authors are grateful to Ms. Li Ding for performing most of the simulation in this study.

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.

Bie, O., Borgan, Ø. and Liestøl, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scand. J. Statist.* **14**, 221-233.

Breslow, N. (1972). Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. B* **34**, 187-220.

Buzas, J. S. (1996). Unbiased scores in Cox regression with covariate measurement error. Manuscript.

Carroll, R. J. (1989). Covariance analysis in generalized linear measurement error models. *Statist. Med.* **8**, 1075-1095.

Carroll, R. J., Ruppert, D. and 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. B* **34**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Efron, B. and Tibshirani, R. (1993). *An Introduction to Bootstrap.* Chapman and Hall, London.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Gong, G., Whittemore, A. S. and Grosser, S. (1990). Censored survival data with misclassified covariate: a case study of breast-cancer mortality. *J. Amer. Statist. Assoc.* **85**, 20-28.

Kong, F. H., Huang, W. H. and Li, X. M. (1998). Estimating survival curves under proportional hazards model with covariate measurement errors. *Scand. J. Statist.* **25**, 573-587.

Kong, F. H. (1999). Adjusting regression attenuation in the Cox proportional hazards model. *J. Statist. Plann. Inference* **79**, 31-44.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84**, 1074-1078.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika* **77**, 127-137.

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

Pepe, M., Self, S. G. and Prentice, R. D. (1989). Further results on covariate measurement errors in cohort studies with time to response data. *Statist. Med.* **8**, 1167-1178.

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

Rosner, B., Willett, W. C. and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statist. Med.* **8**, 1051-1070.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. in Statist. A* **18**, 4335-4358.

Tsiatis, T. T. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108.

Tsiatis, T., T., DeGruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Amer. Statist. Assoc.* **90**, 27-37.

Wang, C. Y., Hsu, L., Feng, Z. D. and Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53**, 131-145.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrics* **50**, 1-25.

Whittemore, A. S. and Keller, J. B. (1988). Approximation for errors in variables regression. *J. Amer. Statist. Assoc.* **83**, 1057-1066.

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

Zhou, H. and Pepe, M. S. (1995). Auxiliary covariate data in failure time regession analysis. *Biometrika* **82**, 139-149.

Biostatistics Group, Westat, 1441 West Montgomery Avenue, Rockville, MD 20850, U.S.A.

Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke St. W., Montreal, QC, H3A 2K6, Canada.

(Received July 1997; accepted September 1998)