Survival Analysis With Heterogeneous Covariate Measurement Error

Yi LI and Louise RYAN

This article is motivated by a time-to-event analysis where the covariate of interest was measured at the wrong time. We show that the problem can be formulated as a special case of survival analysis with heterogeneous covariate measurement error and develop a general analytic framework. We study the asymptotic behavior of the naive partial likelihood estimates and analytically demonstrate that under the heterogeneous measurement error structure and the assumption that all components of the covariate vector and the measurement error vector combined are mutually independent, these naive estimates will shrink toward 0, and that the degree of attenuation increases as the measurement error increases. We also give counterexamples for reverse attenuation when the independence conditions are violated. We use our analytical results to derive a simple bias-correcting estimator that performs well in simulations for small and moderate amounts of measurement error. Our framework can be used to provide insight into the behavior of the commonly used partial likelihood score test for testing no association between a failure outcome and an exposure, for example, in the presence of measurement error or mistiming error. In particular, we derive the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives and discuss the asymptotic relative efficiency. As a result, a simple sample size formula to account for the contamination of covariates is obtained.

KEY WORDS: Asymptotic bias analysis; Asymptotic relative efficiency; Corrected estimator; Cox model; Mistimed covariate; Schoenfeld's sample size formula.

1. INTRODUCTION

One serious problem in survival analysis is that major time-varying covariates of interest are often mistimed. For example, in clinical trials, some baseline biometric measurements (e.g., platelet count and serum creatinine) may not be available at time of entry but may be assessed during the early part of treatment. The ongoing Home Allergen Study (Gold et al. 1999), which motivates this work, was designed to assess environmental effects, such as bacterial endotoxin exposure at birth, on immunological function, allergy, and asthma in infants and young children. The scientific hypothesis is that immunological response to endotoxin exposure in infancy helps to "prime" the infant's ability to respond to environmental triggers, and hence exposure to bacterial endotoxins early in life may confer protection against the development of allergy and asthma in later childhood. However, most noticeably in the design of this study, except for some new born infants, the endotoxin exposure levels were not assessed at birth and were often substituted with measurements made later in the same household. Issues of mistimed covariates have been addressed in a heuristic way by Keiding (1992) and in the context of a pharmacokinetic study by Higgins, Davidian, and Giltinan (1997), but a detailed study in the context of survival analysis has not been conducted.

Of course, failure time regression subject to covariate measurement errors or missing covariates has aroused much interest over the past two decades. Prentice (1982) has demonstrated the impact of measurement error by deriving the induced hazard function in the presence of covariate measurement error and advocated a regression calibration method to draw inference. Zhou and Pepe (1995) and Zhou and Wang (2000) have discussed the use of the calibration approach when some covariates are missing. Xie, Wang, and Prentice (2001) and Wang, Xie, and Prentice (2001) have applied a risk set calibration procedure in a measurement error setting. Estimating equation approaches have been proposed by Huang and Wang (2000), Tsiatis and Davidian (2001), and Hu and Lin (2002) in the context of survival measurement error models and by Lin and Ying (1993), Lipsitz and Ibrahim (1998), Leong, Lipsitz, and Ibrahim (2001), and Chen (2002) in the context of missing-covariate models. Nonparametric maximum likelihood approaches have been adopted by Zhong, Sen, and Cai (1996), and Hu, Tsiatis, and Davidian (1998), Chen and Little (1999), Martinussen (1999), Herring and Ibrahim (2001), and Pons (2002), and maximum partial likelihood estimators have been suggested by Paik and Tsai (1997). But none of these authors has derived expressions describing the asymptotic behavior of the standard Cox proportional hazards model when the covariates are contaminated. Furthermore, none has considered the setting of heterogeneous measurement error, that is, when the measurement error variance changes across individuals. In a different context, the issue of heterogeneous measurement error has been addressed by Carroll and Stefanski (1990) for generalized linear models, but not for censored data. Our first objective, motivated by the work of Hughes (1993) on regression dilution in the naive partial likelihood estimates under a specific measurement error model, is to provide additional insight into the asymptotic behavior of these naive estimates in a broader context. Specifically, we recast the original mistiming problem in a more general framework of measurement error models, and discuss the asymptotic properties of the naive maximum partial likelihood estimates. We prove that under the assumption that all components of the covariate vector and the measurement error vector combined are mutually independent, the naive estimates shrink toward 0, and the degree of attenuation increases as the measurement error becomes more severe. We also give counterexamples for reverse attenuation when the independence condition is violated. Moreover, based on our asymptotic results, we derive a simple bias-correcting estimator, which can be obtained with output from Cox regression analyses in standard software. Simulations indicate that this simple estimator performs well in simulations for small and moderate measurement errors. The second part of this article focuses on the behavior of the commonly used partial likelihood score test for

Yi Li is Assistant Professor (E-mail: *yili@jimmy.harvard.edu*) and Louise Ryan is Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115. This work was supported in part by U.S. National Cancer Institute grants R01 CA95747 (to Li) and R01 CA48061 (to Ryan). The authors thank the editor, an associate editor, and three referees for helpful comments.

Journal of the American Statistical Association
September 2004, Vol. 99, No. 467, Theory and Methods
DOI 10.1198/016214504000001079

hypothesis testing in the presence of general covariate measurement error. In particular, we derive the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives and compute its asymptotic relative efficiency. Our results yield an appealing sample size formula, useful for designing an observational study to compensate for the efficiency loss due to covariate contamination in hypothesis testing.

2. MODELS AND NOTATION

2.1 Mistimed Covariates Model

Assume that the survival times of m independent individuals are subject to right censoring, and assume throughout that censoring is noninformative. Let $T_i = \min(\tilde{T}_i, \tilde{C}_i)$ be the observed survival for subject i $(i = 1, \dots, m)$, where $\tilde{T}_i$ is the true survival time and $\tilde{C}_i$ is the potential censoring time. The individual-specific time origin is the date of a beginning event (e.g., birth) which effectively excludes the possibility of left censoring. Let $\delta_i = I(\tilde{T}_i \le \tilde{C}_i)$ be the noncensoring indicator, which takes value 1 if a failure was observed and 0 otherwise. For subject i, let $\mathbf{X}_i(\cdot)$ be an l_1-dimensional covariate process, which is actually continuous in time but may be observed only at a few time points. In our motivating home allergen study, for example, $\mathbf{X}_i(t)$ is the household endotoxin exposure for subject i at time t. Suppose that for each subject we are interested in relating the outcome T_i to the covariate process $\mathbf{X}_i(\cdot)$ measured at time 0, controlled for the effects of $\mathbf{Z}_i$, an $l_2 \times 1$ accurately measured covariate vector. That is, the outcomes $(T_i, \delta_i), i = 1, \dots, n$, are independent, with the hazard function being linked to the covariates through the following model:

$$\lim_{dt\to 0} (dt)^{-1} P\big(t \le \tilde{T}_i \le t + dt \,|\, \tilde{T}_i \ge t, \mathbf{X}_i(\cdot), \mathbf{Z}_i\big) = \lambda\{t, \mathbf{X}_i(0), \mathbf{Z}_i\}. \quad (1)$$

By noninformative censoring, we mean that $\tilde{C}_i$ is independent of $\tilde{T}_i$ conditional on $\mathbf{X}_i(0)$ and $\mathbf{Z}_i$. A common choice of (1) is the proportional hazards model (Cox 1972),

$$\lambda\{t, \mathbf{X}_i(0), \mathbf{Z}_i\} = \lambda_0(t) \exp\{\boldsymbol{\beta}_x' \mathbf{X}_i(0) + \boldsymbol{\beta}_z' \mathbf{Z}_i\}, \quad (2)$$

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

Model (1) indicates that, conditional on $\mathbf{X}_i(0)$ and $\mathbf{Z}_i$, the measurements of $\mathbf{X}_i(\cdot)$ at other time points provide no extra information regarding the outcome, an analog to the classical assumption of nondifferential measurement error (Carroll, Ruppert, and Stefanski 1995). The statistical challenge is that $\mathbf{X}_i(\cdot)$ is not measured at time 0 but is available at a later time, say $\mathcal{T}_i$ (often prearranged in practice), which is assumed to be independent of the outcome and the concerned covariates. It is, however, customary in data analysis to fit a proportional hazards model (2) by ignoring the mistiming error and directly replacing the unobserved $\mathbf{X}_i(0)$ with the observed $\mathbf{X}_i(\mathcal{T}_i)$. Hence, it will be imperative to analyze the resulting biases.

The mistimed covariate model is completed by specifying the covariate process $\mathbf{X}_i(\cdot)$. We assume that $\mathbf{X}_i(t)$ are iid stochastic processes,

$$\mathbf{X}_i(t) = \mathbf{X}_i(0) + \sigma(t)\mathbf{B}_i(t), \quad (3)$$

where $\sigma(t)$ controls the magnitude of perturbation and the $\mathbf{B}_i(t)$ are independent mean 0 multidimensional stochastic processes with variance–covariance matrix $\mathbf{D}(t)$, which may depend on time t (see, e.g., Kakihara 1997). But for simplicity (and identifiability), we assume that $\mathbf{D}(t)$ is independent of t and hence write it as $\mathbf{D}$. In a setting of mistimed covariates, we typically assume that the scale variance function $\sigma(t)$ is a nondecreasing function of t, exemplifying that the departure from the true covariate tends to amplify with time. Choices of $\sigma(t)$, for example, include $\sigma(t) \equiv \textit{constant}$ or $\sigma(t) = t$ or $\sigma(t) = \exp(t) - 1$, which we use in our numerical studies. Often a moment-based method can be used to consistently estimate $\sigma(t)$ and $\mathbf{D}$ when the process $\mathbf{X}_i(\cdot)$ can be observed at multiple time points (see, e.g., Carroll, Gail, and Lubin 1993; Carroll et al. 1995). In a structural model, the true covariates $\mathbf{X}_i(0)$ are considered to be independently generated from a parameterized distribution (e.g., a multivariate normal distribution), whereas in a functional model, they are considered fixed but unobserved constants.

The naive estimator is the one under model (2) that ignores the mistiming error with directly replacing $\mathbf{X}_i \stackrel{\text{def}}{\equiv} \mathbf{X}_i(0)$ with $\mathbf{W}_i \stackrel{\text{def}}{\equiv} \mathbf{X}_i(\mathcal{T}_i)$. Thus the naively specified hazard function is

$$\lambda_{i,\text{naive}}(t|\mathbf{W}_i, \mathbf{Z}_i) = \lambda_{0,\text{naive}}(t) \exp\{\mathbf{W}_i' \boldsymbol{\beta}_{x,\text{naive}} + \mathbf{Z}_i' \boldsymbol{\beta}_{z,\text{naive}}\}. \quad (4)$$

We would expect that ignoring the mistiming error and directly using the common tools for survival analysis, such as the Cox partial likelihood score approach (Cox 1972), will produce biased results. We investigate these issues analytically in Section 4. First, however, we embed our mistimed covariates model into a broader framework in the next section.

2.2 Heterogeneous Measurement Error Model

Model (3) differs from the classical measurement error model by involving a heterogeneous measurement error variance that depends on other available information—in our motivating example, the assessment time of covariates. This motivates a more general measurement error structure for the unobserved covariates,

$$\mathbf{W}_i = \mathbf{X}_i + \mathbf{u}_i, \quad (5)$$

where, conditional on a subject-specific nonnegative random variable τ_i, the measurement error $\mathbf{u}_i$ are independently normally distributed with mean 0 and variance–covariance matrix $\tau_i \mathbf{D}$ and are independent of $\mathbf{X}_i$, $\tilde{T}_i$, and $\tilde{C}_i$. Here $\mathbf{D}$ is a nonrandom positive definite matrix (which may depend on observed quantities), and τ_i reflects the magnitude of measurement error. We further assume that the τ_i's are independent random variables with mean σ^2 and a finite moment-generating function $M_\tau(v, \sigma^2)$ (in a neighborhood of 0), where

$$M_\tau(v, \sigma^2) = E(e^{v\tau_i}; \sigma^2).$$

Finally, we assume that $\mathbf{X}_i$ has variance–covariance $\boldsymbol{\Sigma}_x$, but that its distribution function is left unspecified for the time being. This new class of measurement error models, allowing the measurement error variance to vary across subjects, is general and encompasses the classical additive measurement error structure (Carroll et al. 1995) and the heteroscedastic measurement error structure (Carroll and Stefanski 1990). For example, it reduces to the classical additive model with $\tau_i = \sigma^2$ almost surely, the heteroscedastic measurement error with $\tau_i = f(\mathbf{X}_i)$

for a given function $f(\cdot)$, and the mistimed covariate model (3) with $\tau_i = \sigma(\mathcal{T}_i)$, where $\mathcal{T}_i$ is the actual time when the error-prone covariate was assessed. Measurement error models with nonconstant numbers of replicated samples (see, e.g., Hu et al. 1998; Xie et al. 2001) also fall into (5).

We turn now to an assessment of the asymptotic behavior of the naive estimates based on the standard Cox partial likelihood approach in the presence of heterogeneous covariate measurement errors. We proceed by introducing the standard counting process notation. Let $N_i = I(T_i \le t, \delta_i = 1), i = 1, \ldots, m$, be a right-continuous process for individual i that documents the number of observed failures, 0 or 1, in an interval $[0, t]$, and $Y_i(t) = I(T_i \ge t)$, a predictable process indicating whether a subject is still at risk at time t. We assume that $(N_i, Y_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i, \tau_i)$ are iid copies of $(N, Y, \mathbf{X}, \mathbf{Z}, \mathbf{W}, \tau)$.

On a probability space, say $(\Omega, \mathcal{F}, P)$, let $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), 0 \le u < t, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i, \tau_i, i = 1, \ldots, m\}$ denote the increasing filtrations that contain the information about survival and covariates up to time t. Suppose, with respect to $\mathcal{F}_t$, that N_i has an intensity function

$$Y_i(t)\lambda(t; \mathbf{X}_i, \mathbf{Z}_i), \tag{6}$$

which may not necessarily be of the proportional form (2). Hence $M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda(u; \mathbf{X}_i, \mathbf{Z}_i)\,du$ are $\mathcal{F}_t$-adapted independent local square-integrable martingales with the variation processes being $\langle M_i, M_j\rangle(t) = 0$ if $i \ne j$ and $\langle M_i, M_i\rangle(t) = \int_0^t Y_i(u)\lambda(u; X_i, \mathbf{Z}_i)\,du$ (see, e.g., Fleming and Harrington 1991).

Now consider the situation where the intensity function for N_i is misspecified by (4). Denote by $\mathbf{W}_{i*} = (\mathbf{W}_i, \mathbf{Z}_i)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_x', \boldsymbol{\beta}_z')'$, and introduce $\mathbf{S}^{(j)}(t) = m^{-1}\sum_i \mathbf{W}_{i*}^{\otimes j} Y_i(t)\lambda(t; \mathbf{X}_i, \mathbf{Z}_i)$, $\mathbf{s}^{(j)}(t) = E\{\mathbf{S}^{(j)}(t)\}$, $\mathbf{S}^{(j)}(\boldsymbol{\beta}, t) = m^{-1}\sum_i \mathbf{W}_{i*}^{\otimes j} Y_i(t) \times \exp(\boldsymbol{\beta}'\mathbf{W}_{i*})$, and $\mathbf{s}^{(j)}(\boldsymbol{\beta}, t) = E\{\mathbf{S}^{(j)}(\boldsymbol{\beta}, t)\}$, where $j = 0, 1, 2$, and the expectations are taken with respect to the true distributions of $N, Y, \mathbf{X}, \mathbf{W}$, and $\mathbf{Z}$ based on models (6) and (5) and for a vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$.

Under the naive model (4), the regression coefficients $\boldsymbol{\beta}$ would typically be estimated by maximizing the log partial likelihood function (Cox 1972)

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^m \int_0^{T_o} \boldsymbol{\beta}'\mathbf{W}_{i*}\,dN_i - \int_0^{T_o} \log S^{(0)}(\boldsymbol{\beta}, t) \sum_{i=1}^m dN_i(t), \tag{7}$$

where T_o is a prespecified constant such that it is within the support of the observed failure time, that is, $P\{T_o < \tilde{C}_i \wedge \tilde{T}_i\} > 0$. In practice, T_o is often the observed maximal study duration for each individual.

The next theorem shows that the sequence of naive estimates $\hat{\boldsymbol{\beta}}_{\text{naive}}$ converges in probability to $\boldsymbol{\beta}^*$, the solution to

$$\mathbf{h}(\boldsymbol{\beta}) = \int_0^{T_o} \mathbf{s}^{(1)}(t)\,dt - \int_0^{T_o} \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} s^{(0)}(t)\,dt. \tag{8}$$

Theorem 1. Under regularity conditions (R.1) and (R.2) given in Appendix A.1, the maximum partial likelihood estimate $\hat{\boldsymbol{\beta}}_{\text{naive}}$ is a consistent estimator of $\boldsymbol{\beta}^*$.

Indeed, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{\text{naive}}$ can also be investigated along the line of Lin and Wei (1989), and a related theorem is documented in Appendix A.2.

3. PROPORTIONAL HAZARDS MODELS

Given the established asymptotic property of the naive estimates, we can evaluate the distortion of the covariate effects under various model misspecifications. In the following we focus on the setting where one correctly specifies the proportional hazards model but fails to account for mistiming errors or covariate measurement errors.

3.1 Asymptotic Bias

Suppose that the true hazard follows a proportional hazards model,

$$\lambda(t, \mathbf{X}, \mathbf{Z}) = \lambda_0(t)\exp\big\{\beta_{1x}^{(0)} X_1 + \cdots + \beta_{l_1 x}^{(0)} X_{l_1} + \beta_{1z}^{(0)} Z_1 + \cdots + \beta_{l_2 z}^{(0)} Z_{l_2}\big\}, \tag{9}$$

where $\mathbf{X} = (X_1, \ldots, X_{l_1})$ and $\mathbf{Z} = (Z_1, \ldots, Z_{l_2})$. Assume that $\mathbf{X}_*' = (\mathbf{X}', \mathbf{Z}')$, $\mathbf{W}_*' = (\mathbf{W}', \mathbf{Z}')$, and $\boldsymbol{\beta}_0' = \{\boldsymbol{\beta}_x^{(0)'}, \boldsymbol{\beta}_z^{(0)'}\} = \{\beta_{1x}^{(0)}, \ldots, \beta_{l_1 x}^{(0)}, \beta_{1z}^{(0)}, \ldots, \beta_{l_2 z}^{(0)}\}$. Let $G(t, \mathbf{X}_*)$, $g(t, \mathbf{X}_*)$, and $C(t, \mathbf{X}_*)$ be the survival function for the survival time, the density function for the survival time, and the survival function for the censoring time, where $G(t, \mathbf{X}_*) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{\beta}_0'\mathbf{X}_*)\}$, $g(t, \mathbf{X}_*) = -\frac{\partial}{\partial t} G(t, \mathbf{X}_*)$, and $C(t, \mathbf{X}_*)$ is left unspecified. As implied by Theorem 1, the asymptotic limit of the naive estimate $\hat{\boldsymbol{\beta}}_{\text{naive}}$, denoted by

$$\boldsymbol{\beta}^*(\sigma^2) = \big\{\boldsymbol{\beta}_x^*(\sigma^2)', \boldsymbol{\beta}_z^*(\sigma^2)'\big\}' = \big\{\beta_{1x}^*(\sigma^2), \ldots, \beta_{l_1 x}^*(\sigma^2), \beta_{1z}^*(\sigma^2), \ldots, \beta_{l_2 z}^*(\sigma^2)\big\}',$$

is the solution to (8) or, more specifically, to

$$0 = \int_0^{T_o} E\{gC\mathbf{X}_*\} - E(gC) \times \left\{\frac{E(GCe^{\boldsymbol{\beta}'\mathbf{X}_*}\mathbf{X}_*)}{E(GCe^{\boldsymbol{\beta}'\mathbf{X}_*})} + \mathbf{F}\mathbf{D}\boldsymbol{\beta}_x\, p\left(\frac{1}{2}\boldsymbol{\beta}_x'\mathbf{D}\boldsymbol{\beta}_x, \sigma^2\right)\right\} dt, \tag{10}$$

where $p(v, \sigma^2) \overset{\text{def}}{\equiv} \frac{\partial}{\partial v}\log M_\tau(v, \sigma^2)$, $\mathbf{F} = (\mathbf{I}_{l_1}, \mathbf{0}_{l_1 \times l_2})'$, $\mathbf{I}_{l_1}$ is an $l_1 \times l_1$ identity matrix, and $\mathbf{0}_{l_1 \times l_2}$ is an $l_1 \times l_2$ matrix with all of its entries being 0. Here we write $G = G(t, \mathbf{X}_*)$, $g = g(t, \mathbf{X}_*)$, and $C = C(t, \mathbf{X}_*)$ for notational simplicity. That (8) equals (10) in the case of the proportional hazards model follows from the nondifferentiality assumption of measurement error and the double-expectation theorem.

With regularity condition (R.2) from Appendix A.1 and by the implicit function theorem, the solution to (10) exists and is a smooth function of σ^2, reflecting the average magnitude of the measurement errors $\mathbf{u}_i$. In particular, when $\sigma^2 = 0$ ($\tau_i = 0$ almost surely), $\boldsymbol{\beta}^*(0) = \boldsymbol{\beta}_0$ is the solution to (10). When $\sigma^2 \to \infty$, $\boldsymbol{\beta}_x^*(\sigma^2) \to 0$, and the contaminated covariate $\mathbf{W}$ provides no information about the outcome. In this case we essentially deal with the Cox proportional hazards model with omitted covariates, and the asymptotic limit $\boldsymbol{\beta}_z^*(\infty)$ is the same as that identified by Struthers and Kalbfleisch (1986), Bretagnolle and Huber-Carol (1988), and Schmoor and Schumacher (1997).

With the independence assumption on the true covariates and the measurement error [(C.3) in App. A.3], a more in-depth investigation leads to the following theorem, which establishes that $\boldsymbol{\beta}^*(\sigma^2)$ is indeed a monotone function of σ^2, the average measurement error variance in the setting of heterogeneous measurement error [cf. (5)]. That is, the attenuation in the naive estimates becomes more severe as σ^2 increases; see Appendix A.3 for a proof. We also provide counter examples when these independence assumptions are violated.

Theorem 2. Under conditions (R.1) and (R.2) in Appendix A.1 and (C.0)–(C.5) in Appendix A.3, the following results are derived:

- If $\boldsymbol{\beta}_z^{(0)} = \mathbf{0}$, then $\boldsymbol{\beta}_z^*(\sigma^2) \equiv \mathbf{0}$; for each $j = 1, \ldots, l_1$; if $\beta_{jx}^{(0)} > 0$, then $\beta_{jx}^*(\sigma^2) > 0$ and $\partial\beta_{jx}^*(\sigma^2)/\partial\sigma^2 < 0$; and if $\beta_{jx}^{(0)} < 0$, then $\beta_{jx}^*(\sigma^2) < 0$ and $\partial\beta_x^*(\sigma^2)/\partial\sigma^2 > 0$.
- If $\boldsymbol{\beta}_x^{(0)} = \mathbf{0}$, then $\beta_{jz}^*(\sigma^2) = \beta_{jz}^{(0)}$ for each $j = 1, \ldots, l_2$. Assume that there exists a $1 \le k \le l_1$ such that $\beta_{kx}^{(0)} \neq 0$; if $\beta_{jz}^{(0)} > 0$; then $\beta_{jz}^*(\sigma^2) > 0$ and $\partial\beta_{jz}^*(\sigma^2)/\partial\sigma^2 < 0$, and if $\beta_{jz}^{(0)} < 0$, then $\beta_{jz}^*(\sigma^2) < 0$ and $\partial\beta_{jz}^*(\sigma^2)/\partial\sigma^2 > 0$.

Remark 1. Although Prentice (1982) and Hughes (1993) demonstrated regression dilution in the naive partial likelihood estimates under a specific measurement error model, this theorem analytically elucidates the attenuation phenomenon under a more general measurement error model. Specifically, under the assumption of independence among the components of the true covariates and the measurement error [(C.3) in App. A.3], the presence of measurement error will attenuate the covariate effects, whereas the direction of impact will be preserved. In particular, the naive estimate of the effect for a correctly measured covariate, albeit independent of the contaminated covariates, is attenuated as well, and the degree of such attenuation increases with the magnitude of measure error. This is in contrast to the usual linear regression, where the naive estimates for the effects of correctly measured covariates are consistent as long as they are independent of the contaminated covariates (Carroll et al. 1995).

We performed numerical studies under the survival model (9) and the mistimed covariate model (3) to illustrate our theoretical results, using the following parameter values: $l_1 = l_2 = 1$; baseline hazard $\lambda_0 = 1$; the true regression coefficients $\beta_x^{(0)} = 1$, and $\beta_z^{(0)} = 1$; the true X and Z, independent and assumed to follow the standard normal distribution; the conditional variance of measurement error $\tau_i = \sigma_0^2\sigma(\mathcal{T}_i)$ in (5); the censoring time, to be independent of X and survival time and to follow an exponential distribution, that is, $C(t) = \exp(-at)$; and the maximal study duration for each individual $T_o = 2$. We assumed that the measurement time $\mathcal{T}_i$ for the error-prone covariate follows a uniform U[0, 1]. Denote by $\rho = \sigma_x/(\sigma_x^2 + \sigma^2)^{1/2}$ the marginal correlation between the observed value W and the true underlying covariate X. By varying σ_0^2, we let ρ range from 1 to 0, with $\rho = 1$ corresponding to no mistiming error and $\rho = 0$ indicating the worst situation where the observed covariate W contains no information about the true underlying covariate X. We chose a to be .13, .47, 1.7, 5.25 to obtain censoring proportions roughly equal to 10%, 30%, 60%, and 80%. We calculated the asymptotic relative bias, defined by the difference between the estimate and the true value divided by the true value, in the naive estimates $\hat{\beta}_{x,\text{naive}}$ and $\hat{\beta}_{z,\text{naive}}$, under measurement error variance functions $\sigma(t) = 1$, $\sigma(t) = t$, and $\sigma(t) = \exp(t) - 1$. An omitted plot shows that the biases increase as the correlation between the true and the observed covariates becomes weaker, which coincides with our theoretical results. In addition, we notice that the biases in the estimates of β_x^* and β_z^* increase when more observations are censored.

Remark 2. Although it may be a common conception that measurement error causes attenuation in the estimation of regression coefficients, reverse-attenuation examples can be found if the independence assumptions on the covariates and the measurement error are violated. To see this, assume the true covariate vector $\mathbf{X}_* \sim \text{MVN}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ and the measurement error $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{D}_*)$, where $\mathbf{D}_*$ is an $(l_1+l_2) \times (l_1+l_2)$ matrix whose first $l_1 \times l_1$ block is $\sigma^2\mathbf{D}$ and the rest 0. Then, under the rare event assumption (i.e., the event occurs with a negligible probability), for any $t \le T_o < \infty$,

$$\lambda(t|\mathbf{W}_*) = \text{E}\{\lambda(t|\mathbf{X}_*)|\tilde{T} \ge t, \mathbf{W}_*\} \doteq \text{E}\{\lambda(t|\mathbf{X}_*)|\mathbf{W}_*\},$$

where the last approximation is due to $P(\tilde{T} \ge t) \doteq 1$ (see Carroll et al. 1995). Using the normality of $\mathbf{X}_*|\mathbf{W}_*$, we obtain

$$\lambda(t|\mathbf{W}_*) \doteq \lambda_0^*(t)\exp\{\boldsymbol{\beta}_0'\boldsymbol{\Sigma}_*(\boldsymbol{\Sigma}_* + \mathbf{D}_*)^{-1}\mathbf{W}_*\}, \qquad (11)$$

where $\lambda_0^*(t) = \lambda_0(t)\exp[\boldsymbol{\beta}_0'\{\mathbf{I} - \boldsymbol{\Sigma}_*(\boldsymbol{\Sigma}_* + \mathbf{D}_*)^{-1}\}\boldsymbol{\mu}_* + .5\boldsymbol{\beta}_0'\{\boldsymbol{\Sigma}_* - \boldsymbol{\Sigma}_*(\boldsymbol{\Sigma}_* + \mathbf{D}_*)^{-1}\boldsymbol{\Sigma}_*\}\boldsymbol{\beta}_0]$. Hence comparing models (11) and (4) indicates that the asymptotic limit of the naive estimate is approximated by

$$\boldsymbol{\beta}^* = (\boldsymbol{\Sigma}_* + \mathbf{D}_*)^{-1}\boldsymbol{\Sigma}_*\boldsymbol{\beta}_0. \qquad (12)$$

If either $\boldsymbol{\Sigma}_*$ or $\mathbf{D}_*$ is nondiagonal (i.e., the independence assumptions are not satisfied), we can always construct the scenarios where some individual components of $\boldsymbol{\beta}^*$ have upward biases or zero crossings. For example, consider a bicovariate model (9) with $l_1 = 2$ and $l_2 = 0$, and assume $\mathbf{X} = (X_1, X_2) \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_*)$ and measurement error $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{D}_*)$. Set the true parameter values as $\beta_{1x}^{(0)} = 1, \beta_{2x}^{(0)} = 3.6$. Using (12), we compute the asymptotic limits of the naive estimates by considering different structures of $\boldsymbol{\Sigma}_*$ and $\mathbf{D}_*$:

- Correlated covariates with independent error. Let $\boldsymbol{\Sigma}_* = (\Sigma_{ij})_{2\times 2}$, where $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = .9$, and $\mathbf{D}_* = \text{diag}(.4, .4)$. Then $\beta_{1x}^* = 1.64$ ($> \beta_{1x}^{(0)}$) and $\beta_{2x}^* = 2.16$.
- Independent covariates with correlated error. Let $\boldsymbol{\Sigma}_* = \mathbf{I}_2$ and $\mathbf{D}_* = (D_{ij})_{2\times 2}$, where $D_{11} = D_{22} = 1$ and $D_{12} = D_{21} = -.65$. Then $\beta_{1x}^* = 1.21$ ($> \beta_{1x}^{(0)}$) and $\beta_{2x}^* = 2.19$.
- Zero crossing. Let $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = -.9$, and $\mathbf{D}_* = \text{diag}(.4, .4)$. Then $\beta_{1x}^* = -.61$ (< 0) and $\beta_{2x}^* = 1.53$.

3.2 A First-Order Bias-Correcting Estimator

Immediately, when σ^2 is small, an approximation to the true $\boldsymbol{\beta}_0$ can be obtained by inverting a Taylor expansion. Specifically, expanding $\boldsymbol{\beta}^*(\sigma^2)$ around $\sigma^2 = 0$ gives

$$\boldsymbol{\beta}_0 = \boldsymbol{\beta}^*(\sigma^2) - \sigma^2 \frac{\partial}{\partial\sigma^2}\boldsymbol{\beta}^*(\sigma^2)\bigg|_{\sigma^2=0} + o(\sigma^2), \qquad (13)$$

where

$$\frac{\partial}{\partial\sigma^2}\boldsymbol{\beta}^*(\sigma^2)\bigg|_{\sigma^2=0} = -\left\{\int_0^{T_o} E(gC)\,dt\right\}\mathbf{V}^{-1}(\boldsymbol{\beta}_0, 0)\mathbf{F}\mathbf{D}\boldsymbol{\beta}_x^{(0)}, \qquad (14)$$

with

$$\mathbf{V}(\boldsymbol\beta,\sigma^2)=\int_0^{T_o}\frac{E(gC)}{\{E(GCe^{\boldsymbol\beta'\mathbf{X}_*})\}^2}\times\Biggl(E\Biggl[GCe^{\boldsymbol\beta'\mathbf{X}_*}\Biggl\{\mathbf{X}_*(\mathbf{X}_*)'+p_1\biggl(\frac{1}{2}\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x,\sigma^2\biggr)(\mathbf{FD}\boldsymbol\beta_x)^{\otimes 2}+p\biggl(\frac{1}{2}\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x,\sigma^2\biggr)\mathbf{FDF}\Biggr\}\Biggr]E\bigl(GCe^{\boldsymbol\beta'\mathbf{X}_*}\bigr)-\bigl\{E\bigl(GC\mathbf{X}_*e^{\boldsymbol\beta'\mathbf{X}_*}\bigr)\bigr\}^{\otimes 2}\Biggr)\,dt. \tag{15}$$

Omitting the higher-order terms in σ^2 from (13) and considering (14), we derive in Appendix A.5 a first-order bias-correcting estimator,

$$\tilde{\boldsymbol\beta}=(\mathbf{I}-\sigma^2\bar N\hat{\mathbf{V}}^{-1}\mathbf{FDF}')^{-1}\hat{\boldsymbol\beta}_{\text{naive}}, \tag{16}$$

where $\mathbf{I}$ is an $(l_1+l_2)\times(l_1+l_2)$ identity matrix; and $m\bar N=\sum_{i=1}^m N_i(T_0)$, the number of events observed during study; $\hat{\boldsymbol\beta}_{\text{naive}}$ is the naive estimate; $\hat{\mathbf{V}}$ is an approximation to $\mathbf{V}(\boldsymbol\beta_0,0)$, given by

$$\hat{\mathbf{V}}=\int_0^{T_o}\left[\frac{\mathbf{S}^{(2)}(\hat{\boldsymbol\beta}_{\text{naive}},t)}{S^{(0)}(\hat{\boldsymbol\beta}_{\text{naive}},t)}-\left\{\frac{\mathbf{S}^{(1)}(\hat{\boldsymbol\beta}_{\text{naive}},t)}{S^{(0)}(\hat{\boldsymbol\beta}_{\text{naive}},t)}\right\}^{\otimes 2}\right]\times\frac{1}{m}\sum_{i=1}^m dN_i(t);$$

and $S^{(j)}(\boldsymbol\beta,t)=m^{-1}\sum_i\mathbf{W}_{i*}^{\otimes j}Y_i(t)\exp(\boldsymbol\beta'\mathbf{W}_{i*})$, $j=0,1,2$. Here σ^2 is assumed to be known or can be consistently estimated from validation datasets or reliability samples. Notice that $(m\hat{\mathbf{V}})^{-1}$ is the naive variance–covariance estimate of $\hat{\boldsymbol\beta}_{\text{naive}}$, which is available from commonly used statistical software (e.g., SAS PROC PHREG and S–PLUS COXPH). Hence the corrected estimates can be conveniently obtained from the output of the naive fitting, involving considerably less computation. As indicated in Appendix A.5, the corrected estimator is applicable in much more general situations without conditions (C.0)–(C.5) listed in Appendix A.3, particularly the independent assumptions on the covariates and the measurement error, which are required for the theoretical bias analyses. Note that (16) reduces to the naive estimates when $\sigma^2=0$, and that the form of (16) resembles the corrected estimator in linear regressions (Carroll et al. 1995, chap. 2) and in survival analysis with constant measurement error variance (Kong 1999).

An application of the delta method yields an approximate "sandwich" variance estimator for (16),

$$\widehat{\text{var}}(\tilde{\boldsymbol\beta})=(\mathbf{I}-\sigma^2\bar N\hat{\mathbf{V}}^{-1}\mathbf{FDF}')^{-1}\times\widehat{\text{var}}(\hat{\boldsymbol\beta}_{\text{naive}})(\mathbf{I}-\sigma^2\bar N\hat{\mathbf{V}}^{-1}\mathbf{FDF}')^{-T}. \tag{17}$$

Here $\widehat{\text{var}}(\hat{\boldsymbol\beta}_{\text{naive}})$ is given in (A.1), Appendix A.2, and is available by invoking the option of robust variance in Splus COXPH and SAS PHREG (version 8.1) (or SAS PHLEV macro, downloadable from www.mayo.edu/hsr/biostat.html), which agrees well with the naive variance obtained from the Martingale residuals for a relatively small σ^2.

However, (17) may lead to underestimation, especially when σ^2 is large, as it fails to account for the variation of $\bar N$ and $\hat{\mathbf{V}}$ in the formulation. A simple but effective alternative is to apply a Bootstrap procedure (Efron 1979). Specifically, we resample m subjects, with replacement, from $(T_i,\delta_i,\mathbf{W}_{i*})|_{i=1}^m$ to obtain a new dataset $\{T_{(i)},\delta_{(i)},\mathbf{W}_{(i*)}\}|_{i=1}^m$. Given this new dataset, we use (16) to compute the corrected estimates. Such a procedure can be repeated K times to obtain a sequence of estimates, $\tilde{\boldsymbol\beta}^{(k)}$, $k=1,\ldots,K$. The bootstrap variance estimates can hence be calculated using the sample variances

$$\text{var}_{\text{boot}}(\tilde{\boldsymbol\beta})=\frac{1}{K-1}\sum_{k=1}^K\bigl\{\tilde{\boldsymbol\beta}^{(k)}-\bar{\boldsymbol\beta}_{\text{boot}}\bigr\}\bigl\{\tilde{\boldsymbol\beta}^{(k)}-\bar{\boldsymbol\beta}_{\text{boot}}\bigr\}',$$

where $\bar{\boldsymbol\beta}_{\text{boot}}=\frac{1}{K}\sum_{l=1}^K\tilde{\boldsymbol\beta}^{(k)}$. In practice, it is adequate to choose a moderate number of resamplings, K, say, in the range 25 to 100 (Lange 1998, p. 301). We chose $K=30$ for simulations.

3.3 Simulations and Two Worked Examples

Simulations were performed to examine the finite sample performance of the first-order corrected estimator (16). With similar configurations as in the foregoing section, we varied the sample size from 80 to 500 and the average measurement error variance from .05 to .90. The parameter in the censoring distribution was taken to be .47, corresponding roughly to a 30% censoring proportion. A total of 1,000 simulations were conducted for each configuration. It appears that with a small sample size, the corrected estimator performs well for small and moderate measurement errors (e.g., $0\le\sigma^2\le .5$), while, with a moderate or a large sample size (e.g., $m=200$ or $m=500$), the corrected estimator performed well even for large measurement errors (e.g., $\sigma^2=.9$). To examine the performance of the estimator with different magnitude of the underlying covariate effect and variance function $\sigma(\cdot)$, we varied $\beta_x^{(0)}$ and $\beta_z^{(0)}$ from .25 to 1.0 and chose $\sigma(t)=1$ or $\exp(t)-1$, and found similar patterns. We calculated the averages of the naive estimates and the corrected estimates, along with their empirical standard errors (SEs) and the average bootstrap standard errors, with $\sigma(t)=t$. It appears that the bootstrap variances are in agreement with the empirical counterparts for small sample sizes with small and moderate measurement errors and for large sample sizes even with relatively large measurement errors. For example, with $m=500$, when $\sigma^2=.5$, the averages of the corrected estimates for $\beta_x^{(0)}$ and $\beta_z^{(0)}$ were .9918 (empirical SE = .154, bootstrap SE = .157) and .9928 (empirical SE = .104, bootstrap SE = .101), and when $\sigma^2=.9$, the corresponding averages were 1.0003 (empirical SE = .231, bootstrap SE = .296) and .9942 (empirical SE = .127, bootstrap SE = .143).

We illustrate the practical use of (16) with two published studies. The first is Framingham study, a subset of which was analyzed by Xie et al. (2001). Of particular interest was the effect of long-term average systolic blood pressure (SBP) on developing coronary heart disease for middle-aged men. A total of 85 failures of 423 subjects were observed in a span of 10 years of follow-up. The true covariate X_i is the long-term average SBP, which was subject to error due to variations of

SBP from time to time. Multiple measurements of SBP were made for each individual, which were assumed to follow

$$W_{ij} = X_i + u_{ij} \tag{18}$$

for $i = 1, \ldots, n_i$. Here $u_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_u^2)$, and are independent of X_i. The within-subject average of measurements (e.g., $\bar{W}_i = \frac{1}{n_i}\sum_i W_{ij}$) was considered as the surrogate for X_i. Hence for each individual, the measurement error variance is σ_u^2/n_i. In this data subset $n_i \equiv 2$, and σ_u^2 was identified to be .039, indicating that for each individual, the average measurement error variance is .019. The reported naive estimate for the Cox model was 1.441, with naive variance .089. Immediately, using (16) and (17) with $l_1 = 1$ and $l_2 = 0$, we found that the first-order–corrected estimate was 1.69 (SE = .350) which matches well with the ordinary regression calibration estimate 1.74 (SE = .366) and the risk set regression calibration estimate 1.76 (SE = .372) reported by Xie et al. (2001).

Another example is the ACTG 116b/117 study (Kahn et al. 1992) on the relationship between the number of CD4 counts and the disease progression-free survival. Of 912 patients accrued, a total of 334 events, including disease progression or death, were observed during study. The number of CD4 replicates varied across patients: 56 patients had 1 replicate measurement, 310 patients had 2 replicate measurements, 541 patients had 3 replicate measurements, and 5 patients had 4 replicate measurements. Under model (18), Hu et al. (1998) calculated $\sigma_u^2 = .1505$ and reported a naive estimate of β_x of $-.6015$ with naive variance .00319. The average measurement error variance were calculated by $\sigma^2 = (56 \times \sigma_u^2 + 310 \times \sigma_u^2/2 + 541 \times \sigma_u^2/3 + 5 \times \sigma_u^2/4)/912 = .0646$. Hence applying formulas (16) and (17) yielded the corrected estimate $-.6497$ (SE = .061), which is in agreement with the regression calibration estimate of $-.6307$ (SE = .059), tha fully parametric estimate of $-.6597$ (SE = .062), the fully nonparametric estimate of $-.6216$ (SE = .057), and the semiparametric estimate of $-.6641$ (SE = .063) reported by Hu et. al (1998).

4. HYPOTHESIS TESTING IN THE PRESENCE OF GENERAL COVARIATE MEASUREMENT ERROR

A relatively unexplored subject in survival analysis is hypothesis testing in the presence of mistiming error or general measurement error. For simplicity of exposition, we postulate that the true hazard adheres to a single covariate proportional hazards model,

$$\lambda(t, X_i) = \lambda_0(t)\exp(\beta X_i).$$

For testing the null hypothesis $\beta = 0$, a naive partial likelihood score test would practically be used by ignoring mistiming error or measurement error and computing

$$m^{-1/2}U_m = m^{-1/2}\sum_{i=1}^{m}\int_0^{T_o}\{W_i - \bar{W}(t)\}\,dN_i(t), \tag{19}$$

where $\bar{W}(t) = \sum_{i=1}^m Y_i(t)W_i / \sum_{i=1}^m Y_i(t)$. Under the null hypothesis, the naive model (4) coincides with the true model. Therefore, test (19) is asymptotically a mean-0, normally distributed random variable. With a properly calculated variance (given in Thm. 4), test (19) is valid and retains the nominal level under the null hypothesis. But, as shown in the following, the induced survival model under the alternative hypothesis does not preserve proportionality, and as a result, the loss of efficiency in test (19) would be expected (Lagakos 1988). Hence, it will be of substantial interest to analytically characterize the efficiency loss due to covariate measurement errors.

4.1 Asymptotic Relative Efficiency and Its Applications

To facilitate the following discussion, suppose that we invert (5) so that

$$X_i = \alpha_i^* + \rho_i^* W_i + \epsilon_i, \tag{20}$$

where $\rho_i^* = \frac{\sigma_x^2}{\sigma_x^2 + \tau_i}$, $\sigma_x^2 = \mathrm{var}(X)$, α_i^* is a constant possibly depending on τ_i, and ϵ_i is a mean-0 random variable that is independent of W_i, conditional on τ_i. Indeed, (20) holds if we assume that the unobserved covariate X_i follows a normal distribution.

Similarly, we may calculate the hazard function for each counting process N_i based on the observed covariates W_i and measurement error variance τ_i,

$$\lambda(t|W_i, \tau_i) = \lambda_0(t)\exp\{\beta\alpha_i^* + \beta\rho_i^* W_i + \psi(\beta, W_i, \tau_i, t)\}, \tag{21}$$

where $\psi(\beta, W_i, \tau_i, t) = \log E(e^{\beta\epsilon_i}|\tilde{T}_i \geq t, W_i, \tau_i)$.

Note that $\psi(\beta, W_i, \tau_i, t) \equiv 0$ when $\tau_i = 0$ or $\beta = 0$, corresponding to the case of no measurement error and the null model. More generally, in the presence of measurement error, the induced hazard function fails to preserve the proportionality. Hence the simple log-rank test (19) may be inefficient relative to that in the absence of measurement error. We explore this more formally by deriving the asymptotic relative efficiency (ARE), a useful device for comparing tests, of the simple log-rank test versus its counterpart in the absence of measurement error. To proceed, we consider a sequence of local alternatives that converges to the null hypothesis at the appropriate rate as sample size increases to infinity (Fleming and Harrington 1991). Under these alternatives, the log-rank statistic has, asymptotically, a finite mean and variance.

For notational convenience, denote

$$\phi(\beta, W_i, \tau_i, t) = \beta\alpha_i^* + \beta\rho_i^* W_i + \psi(\beta, W_i, \tau_i, t), \tag{22}$$

$$s_\beta^{(j)}(t) = E\{Y_i(t)W_i^j\phi_\beta(\beta, W_i, \tau_i, t)\lambda_0(t)\}, \tag{23}$$

and

$$s^{(j)}(t) = E\{Y_i(t)W_i^j\lambda_0(t)\}, \tag{24}$$

where the expectation is taken with respect to N, Y, W, and τ for $j = 0, 1, 2$, and $\phi_\beta(\cdot)$ is the partial derivative of $\phi(\cdot)$ with respect to β. Assuming regularity conditions A–D of Andersen and Gill (1982, sec. 3), and a sequence of local alternatives, $m^{-1/2}\eta$ for some fixed η $(-\infty < \eta < \infty)$, we have the following theorem (the proof of which is given in App. A.6) about the asymptotic behavior of the log-rank test (19).

Theorem 3. With conditions A–D given by Andersen and Gill (1982, sec. 3) and under the hypothesis $\beta_m = m^{-1/2}\eta$, $m^{-1/2}U_m$ converges in distribution to a normal distribution with mean

$$\int_0^{T_o}\eta\frac{s^{(1)}(t)s_\beta^{(0)}(t) - s^{(0)}(t)s_\beta^{(1)}(t)}{s^{(0)}(t)}\,dt$$

and variance

$$\int_0^{T_o}\left[\frac{s^{(2)}(t)}{s^{(0)}(t)}-\left\{\frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^2\right]s^{(0)}(t)\,dt,$$

where $s_\beta^{(j)}(t)$ and $s^{(j)}(t)$ are defined in (23) and (24), and all of the expectations are taken under the null hypothesis $\beta=0$, and the derivatives are evaluated at $\beta=0$.

Therefore, under the null hypothesis (i.e., $\eta=0$), the naive score test is unbiased, and under the local alternatives (i.e., $\eta\neq 0$), the asymptotic distribution of the naive score test is still normal with the same variance as in the null case but with the asymptotic mean shifted. Thus we can compute the asymptotic efficacy of (19), which is defined by its noncentrality,

$$\eta^2\left(\int_0^{T_o}\frac{s^{(1)}(t)s_\beta^{(0)}-s^{(0)}(t)s_\beta^{(1)}}{s^{(0)}(t)}\,dt\right)^2 \Bigg/ \int_0^{T_o}\left[\frac{s^{(2)}(t)}{s^{(0)}(t)}-\left\{\frac{s^{(1)}(t)}{s^{(0)}(t)}\right\}^2\right]s^{(0)}(t)\,dt.$$

It is reasonable in practice to assume that the censoring time is independent of X_i and W_i, which results in a more simplified efficacy. Under the null hypothesis $\beta=0$, because the observed survival time T is independent of W, for $j=0,1,2$,

$$s_\beta^{(j)}(t)=\pi(t)\lambda_0(t)E\{W_i^j\phi_\beta(\beta,W_i,\tau_i,t)\}$$

and

$$s^{(j)}(t)=\pi(t)\lambda_0(t)E(W_i^j),$$

where $\pi(t)=E\{Y_i(t)\}=P(T_i\geq t)$. Direct calculations yield

$$\frac{\partial}{\partial\beta}\psi(\beta,W,\tau,t)\bigg|_{\beta=0}=0, \tag{25}$$

and hence $\phi_\beta(\beta,W,\tau,t)|_{\beta=0}=\alpha^*+\rho^*W_i$, where $\rho^*=\sigma_x^2/(\sigma_x^2+\tau)$. It follows that the simplified efficacy is

$$\mathit{eff}_{\text{naive}}=\frac{\eta^2[E(W)E(\rho^*W)-E(\rho^*W^2)]^2}{\operatorname{var}(W)}\times\int_0^{T_o}\pi(t)\lambda_0(t)\,dt. \tag{26}$$

Write $P_{T_o}=\int_0^{T_o}\pi(t)\lambda_0(t)\,dt$. With basic martingale theory, it follows that $P_{T_o}=E\{N(T_o)\}$ under the null hypothesis, indicating that P_{T_o} is the probability of observing a failure by time T_0. Moreover, using double expectation gives that $\operatorname{var}(W)=\sigma_x^2+\sigma^2$, $E(\rho^*W)=E\{E(\rho^*W|\tau)\}=E(\rho^*)E(X)$, and $E(\rho^*W^2)=E\{E(\rho^*W^2|\tau)\}=E\{\rho^*(\sigma_x^2+\tau)\}+E(\rho^*)\{E(X)\}^2=\sigma_x^2+E(\rho^*)\{E(X)\}^2$. Hence (26) can be rewritten as

$$\mathit{eff}_{\text{naive}}=\frac{\eta^2\sigma_x^4}{\sigma_x^2+\sigma^2}P_{T_o}. \tag{27}$$

Therefore, in the absence of measurement error ($\sigma^2=0$), the maximum efficacy of test (19) is achieved at $\eta^2\sigma_x^2P_{T_o}$.

By calculating the ratio of efficacies, it follows that the asymptotic relative efficiency (ARE) comparing the log-rank test in the presence of measurement error to that in the absence of measurement error is $\sigma_x^2/(\sigma_x^2+\sigma^2)$, which is a monotonically decreasing function of σ^2/σ_x^2, the relative variability of measurement error with respect to the true underlying covariate. A relative efficiency less than 1 indicates that more observations are needed for the test in the presence of measurement error. Noticeably in the context of mistimed covariate problem, this result depends not on the specific forms of the variance function $\sigma(t)$ in (3) and the distribution of the measurement time $\mathcal{T}_i$, but only on σ^2, the average error variance.

The foregoing calculations have an immediate application in the study design. Specifically, for a fixed alternative $\beta=\beta_a$ which is in the range of $O(m^{-1/2})$, by Theorem 4, the distribution of the simple log-rank test under the true model and under the null hypothesis is approximately given by

$$\mathrm{N}\{m^{1/2}\beta_a\sigma_x^2P_{T_o},(\sigma_x^2+\sigma^2)P_{T_o}\}.$$

It then follows that the sample size needed to detect $H_a:\beta=\beta_a\neq 0$ versus $H_0:\beta=0$ for the naive test with power δ and one-sided type I error level of ϵ is

$$\frac{(\mathcal{Z}_{1-\epsilon}+\mathcal{Z}_\delta)^2}{\beta_a^2\mathit{eff}_{\text{naive}}},$$

or, equivalently, the number of events required is

$$(\mathcal{Z}_{1-\epsilon}+\mathcal{Z}_\delta)^2\frac{\sigma_x^2+\sigma^2}{\beta_a^2\sigma_x^4}, \tag{28}$$

where $\mathcal{Z}_q$ is the $100\times q$ percentile of a standard normal distribution. Note that (28) is essentially an extension of Schoenfeld's sample size formula (Schoenfeld 1983), which applies in the absence of measurement error. Hence if a conventional partial likelihood score test is opted for, then the sample size should be inflated by σ^2/σ_x^2 to compensate for the efficiency loss due to measurement error.

4.2 Optimality of Log-Rank Tests With General Covariate Measurement Error

A natural question would be whether the efficiency of test (19) can be improved within a more general class of weighted log-rank tests,

$$m^{-1/2}\tilde{U}_m=m^{-1/2}\sum_{i=1}^m\int_0^{T_o}r_m(t)\{W_i-\bar{W}(t)\}\,dN_i(t), \tag{29}$$

by choosing a proper weight function $r_m(t)$. Here $r_m(t)$ is a bounded predictable process, converging uniformly in probability to a bounded nonrandom function $r(t)$ over $[0,T_o]$. For example, $r_m(t)=\{\hat{S}(t-)\}^\rho\{1-\hat{S}(t-)\}^\gamma$, where $\hat{S}(\cdot)$ is the Kaplan–Meier survival estimate, corresponds to the $G^{\rho,\gamma}$ class test (see, e.g., Fleming and Harrington 1991, chap. 7). Yet, as revealed by the following theorem (proved in App. A.7), with a noninformative censoring mechanism, test (19) is in fact the optimal test within the general class (29).

Theorem 4. Assume that for each i, the censoring time $\tilde{C}_i$ is independent of the observed covariate W_i. Test (19) achieves the maximum efficacy within the general log-rank test class (29).

Hence tests beyond the classical log-rank test family (29) need to be considered for improving efficiency. Alternative choices, include generalized log-rank tests in which a subject-specific weight function is assigned to each individual,

$$m^{-1/2}U_m = m^{-1/2}\sum_{i=1}^{m}\int_0^{T_o} r_{m,i}(t)\{W_i - \tilde{W}_m(t)\}\,dN_i(t),$$

where $\tilde{W}_m(t) = \sum_{i=1}^{m} Y_i(t)r_{m,i}(t)W_i/(\sum_{i=1}^{m} Y_i(t)r_{m,i}(t))$. A heuristic strategy would be to put more weight on subjects whose covariates are measured with more precision, although this warrants a detailed investigation. In a different context, Kong and Slud (1997) and DiRienzo and Lagakos (2001) considered improving test (29) for two treatment comparisons with a misspecified proportional hazard model.

5. DISCUSSION

In this article we have discussed survival analysis with mistimed covariates within a more general heterogenuous measurement error framework. To understand the consequence of mistiming error on parameter estimation, we have focused on the asymptotic behavior of the naive maximum partial likelihood estimates and have shown that under the heterogeneous measurement error structure and the assumption that all components of the covariate vector and the measurement error vector combined are mutually independent, these naive estimates will shrink toward 0, and that the degree of attenuation increases as the measurement error increases. We also gave several counterexamples for reverse attenuation when these independence conditions are violated. To our knowledge, this result is also new in the context of survival analysis and argues against the common conception that measurement error will always lead to attenuation.

Moreover, as a byproduct of our asymptotic bias-analyses, we have obtained a simple bias-correcting estimator that performs well for small sample sizes with small and moderate measurement errors. Because the proposed estimator capitalizes on the existing standard statistical software, the computational burden is considerably less than that for the other bias-correcting approaches, for example, calibration regression or corrected partial likelihood score. However, because our estimator was derived under the assumption of small and (conditionally) normal measurement error, more sophisticated methods, such as likelihood-based methods, may be needed in cases with a large amount of measurement error or nonnormal measurement error.

We have also considered the asymptotic behavior of the commonly used partial likelihood score test for assessing the association between a failure outcome and an exposure in the presence of measurement error or mistiming error. In particular, have derived the asymptotic distribution of the naive partial likelihood score test under a series of local alternatives and calculated the asymptotic relative efficiency. As a result, we obtained a sample size formula, easily implementable by practitioners when designing an observational study to compensate for the efficiency loss due to covariate measurement errors.

APPENDIX: TECHNICAL DETAILS

A.1 Regularity Conditions and a Proof of Theorem 1

We assume the following regularity conditions:

(R.1) There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}^*$ such that

$$\sup_{t\in[0,T_o],\boldsymbol{\beta}\in\mathcal{B}} \left|\mathbf{S}^{(j)}(\boldsymbol{\beta},t) - \mathbf{s}^{(j)}(\boldsymbol{\beta},t)\right| \xrightarrow{p} 0.$$

Assume that $\mathbf{s}^{(j)}(\boldsymbol{\beta},t)$, $j=0,1$, are bounded on $\mathcal{B}\times[0,T_o]$, and that $s^{(0)}(\boldsymbol{\beta},t)$ is bounded away from 0 on $\mathcal{B}\times[0,T_o]$.

(R.2) The negative second-order derivative of $\mathbf{h}(\boldsymbol{\beta})$,

$$\mathbf{A}(\boldsymbol{\beta}) = \int_0^{T_o}\left[\frac{\mathbf{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)} - \left\{\frac{\mathbf{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}\right]s^{(0)}(t)\,dt$$

is positive definite at $\boldsymbol{\beta}^*$.

Using the techniques of lemma 3.1 of Andersen and Gill (1982), $\ell(\boldsymbol{\beta})$ in (7) can be shown to be asymptotically equivalent to $H(\boldsymbol{\beta}) = \int_0^{T_o}\boldsymbol{\beta}'\mathbf{s}^{(1)}(t)\,dt - \int_0^{T_o}\log\{s^{(0)}(\boldsymbol{\beta},t)\}s^{(0)}(t)\,dt$. Hence, by theorem 2.1 of Struthers and Kalbfleisch (1986), we immediately have this consistency result.

A.2 Asymptotic Normality of Naive Estimates

Theorem A.1.

$$m^{1/2}(\hat{\boldsymbol{\beta}}_{\text{naive}} - \boldsymbol{\beta}^*) \xrightarrow{d} \mathrm{N}\{\mathbf{0}, \mathbf{A}^{-1}(\boldsymbol{\beta}^*)\mathbf{B}(\boldsymbol{\beta}^*)\mathbf{A}^{-1}(\boldsymbol{\beta}^*)\},$$

where

$$\mathbf{B}(\boldsymbol{\beta}^*) = E\Bigg[\int_0^{T_o}\left\{\mathbf{W}_{i*} - \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right\}dN_i(t) - \int_0^{T_o}\frac{Y_i(t)\exp(\boldsymbol{\beta}^{*\prime}\mathbf{W}_{i*})}{s^{(0)}(\boldsymbol{\beta}^*,t)}\left\{\mathbf{W}_{i*} - \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}^*,t)}{s^{(0)}(\boldsymbol{\beta}^*,t)}\right\}s^{(0)}(t)\,dt\Bigg]^2.$$

Moreover, $\mathbf{A}(\boldsymbol{\beta}^*)$ and $\mathbf{B}(\boldsymbol{\beta}^*)$ can be consistently estimated by $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\text{naive}})$ and $\hat{\mathbf{B}}(\hat{\boldsymbol{\beta}}_{\text{naive}})$. Here

$$\hat{\mathbf{V}}(\boldsymbol{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\int_0^{T_o}\left[\frac{\mathbf{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} - \left\{\frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}\right]dN_i(t),$$

$$\mathbf{G}_i(\boldsymbol{\beta}) = \int_0^{T_o}\left\{\mathbf{W}_{i*} - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right\}dN_i(t) - \sum_{j=1}^{m}\int_0^{T_o}\frac{Y_i(t)\exp(\boldsymbol{\beta}'\mathbf{W}_{i*})}{mS^{(0)}(\boldsymbol{\beta},t)}\left\{\mathbf{W}_{i*} - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right\}dN_j(t),$$

and

$$\hat{\mathbf{B}}(\boldsymbol{\beta}) = m^{-1}\sum_{i=1}^{m}\mathbf{G}_i(\boldsymbol{\beta})^{\otimes 2}.$$

Thus a consistent variance estimator of $\hat{\boldsymbol{\beta}}_{\text{naive}}$ is

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{\text{naive}}) = \frac{1}{m}\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{naive}})\hat{\mathbf{B}}(\hat{\boldsymbol{\beta}}_{\text{naive}})\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{naive}}). \quad \text{(A.1)}$$

Algebraically, this type of variance estimator is equivalent to the approximate robust jackknife variance, and related statistical software is available (see Therneau and Grambsch 2000). Under the null hypothesis $H_0: \boldsymbol{\beta} = 0$, it follows from (8) that $\boldsymbol{\beta}^* = 0$. Hence this theorem suggests that even in the presence of covariate measurement errors, it is possible to perform hypothesis testing based on the naive estimates using asymptotic normality and estimation of the asymptotic variance.

A.3 Additional Regularity Conditions for Bias Analysis (Thm. 3) and Lemmas

We postulate similar regularity conditions as given by in Bretagnolle and Huber-Carol (1988) for Theorem 3 [except for (C.0)]:

(C.0) Assume that the measurement error variance–covariance matrix $\mathbf{D}$ is a diagonal matrix and that the scale measurement error variance τ has a moment-generating function $M_\tau(v,\sigma^2)$ such that $\frac{\partial^2}{\partial v\,\partial\sigma^2}\log M_\tau(v,\sigma^2) > 0$ when $v,\sigma^2 \ge 0$.

(C.1) The additional covariate vector $\mathbf{Z} = (Z_1,\ldots,Z_{l_2})$ is "pertinent" in model (9), in that the true effect of each component, say, $\beta_{jz}^{(0)}$, $j = 1,\ldots,l_2$, is nonzero.

(C.2) The true covariate vector $(\mathbf{X},\mathbf{Z})$ has time-independent components with a finite moment-generating function, and there is no proper linear subspace including $\mathbf{X}$ and $\mathbf{Z}$ almost surely.

(C.3) The joint density function of true covariates and measurement error, $(\mathbf{X},\mathbf{Z},\mathbf{u})$, has a decomposition

$$f_{\mathbf{X},\mathbf{Z},\mathbf{u}}\big(x_1,\ldots,x_{l_1},z_1,\ldots,z_{l_2},v_1,\ldots,v_{l_1}\big) = \prod_{i=1}^{l_1} f_{X_i}(x_i)\prod_{j=1}^{l_2} f_{Z_j}(z_j)\prod_{k=1}^{l_1} f_{u_k}(v_k),$$

where $f_{(\cdot)}(\cdot)$ are density functions.

(C.4) $P(\tilde{C}\wedge T_o < \tilde{T}|\mathbf{X},\mathbf{Z}) < 1$ almost surely, to guarantee observability of covariates.

(C.5) The survival function of the censoring does not depend on $\mathbf{X}$ and $\mathbf{Z}$.

As indicated by the following lemma, (C.0) is not restrictive and is satisfied by the nonnegative random variables whose distributions are in the exponential distribution family, including the following examples: Bernoulli, $M_\tau(v,\sigma^2) = (1-\sigma^2) + \sigma^2 e^v$; Poisson, $e^{\sigma^2\{\exp(v)-1\}}$; chi-squared, $(1-2v)^{-1/\sigma^2/2}$, $v < 1/2$; and exponential, $(1-\sigma^2 v)^{-1}$, $v < 1/\sigma^2$. The assumptions is also satisfied for the uniform, for example, $\tau \sim \mathrm{U}[0, 2\sigma^2]$, whose moment-generating function is $M_\tau(v,\sigma^2) = (e^{2v\sigma^2}-1)/2v\sigma^2$.

Lemma A.1. Let τ be any arbitrary nonnegative random variable whose distribution is in the exponential family and let $M_\tau(v,\mu) = E(e^{v\tau})$ be its moment-generating function, where $\mu = E(\tau)$. Then $\frac{\partial^2}{\partial v\,\partial\mu}\log M_\tau(v,\mu) > 0$.

Proof. The probability density function (if τ is continuous) or the probability mass function (if τ is discrete) in the exponential family can be parameterized by $f(\tau;\theta,\phi) = \exp\{\frac{\tau\theta - b(\theta)}{a(\phi)} + c(\tau,\phi)\}$, where $\mu = E(\tau) = b'(\theta)$ and $\mathrm{var}(\tau) = a(\phi)b''(\theta)$ (McCullagh and Nelder 1989).

Direct calculations yield

$$\frac{\partial^2}{\partial v\,\partial\mu}M_\tau(v,\mu) = \frac{E(\tau e^{v\tau}S_\mu)E(e^{v\tau}) - E(\tau e^{v\tau})E(e^{v\tau}S_\mu)}{\{E(e^{v\tau})\}^2}, \quad \text{(A.2)}$$

where $S_\mu = \frac{\partial}{\partial\mu}\log f(\tau;\theta,\phi)$ and the integrals involved in the expectations are taken with respect to the Lebesque measure if τ is continuous or with respect to the counting measure if τ is discrete. By the chain rule, $S_\mu = \frac{\partial}{\partial\theta}\log f(\tau;\theta,\phi)\frac{d\theta}{d\mu} = \frac{\tau-\mu}{\mathrm{var}(\tau)}$. Hence substituting in S_μ yields

$$\frac{E(\tau^i e^{v\tau}S_\mu)}{E(\tau^i e^{v\tau})} = \left\{\frac{E(\tau^{i+1}e^{v\tau})}{E(\tau^i e^{v\tau})} - \mu\right\}\frac{1}{\mathrm{var}(\tau)} \quad \text{(A.3)}$$

for $i = 0, 1$. In addition, an application of the Schwarz inequality yields

$$E(e^{v\tau})E(\tau^2 e^{v\tau}) > \{E(\tau e^{v\tau})\}^2. \quad \text{(A.4)}$$

Combining (A.3) and (A.4), we have that $\frac{E(\tau e^{v\tau}S_\mu)}{E(\tau e^{v\tau})} > \frac{E(e^{v\tau}S_\mu)}{E(e^{v\tau})}$, which implies that the numerator in (A.2) is positive and finishes the proof.

Next state a technical lemma that is useful in the proof of Theorem 3. The proof was given by Bretagnolle and Huber-Carol (1988).

Lemma A.2. A symmetric square matrix $\mathbf{M}$ is called DP+ type if it is positive definite and its off-diagonal elements are all negative. Then each element of the inverse of a DP+ type matrix is positive.

A.4 Proof of Theorem 3

If $\boldsymbol{\beta}_x^{(0)} = \mathbf{0}$, then the naive model (4) correctly specifies the hazard function. Hence for any $\sigma^2 \ge 0$, the limits of the naive estimates are $\boldsymbol{\beta}_x^*(\sigma^2) = \mathbf{0}$ and $\beta_{jz}^*(\sigma^2) = \beta_{jz}^{(0)}$ for $1 \le j \le l_2$.

Without loss of generality, we consider the situation when $\beta_{1x}^{(0)} \ne 0$. We may assume that the other components of $\boldsymbol{\beta}_0$ are nonnegative. Otherwise, if there exists, for instance, a j such that $\beta_{jz}^{(0)} < 0$, then one can always reverse the sign of the corresponding covariate Z_j.

Denote the right side of (10) by $\mathbf{S}(\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_x', \boldsymbol{\beta}_z')' = (\beta_{1x},\ldots,\beta_{l_1x},\beta_{1z},\ldots,\beta_{l_2z})'$. Denote $\mathbf{X}_{-1} = (X_2,\ldots,X_{l_1})'$ and $\boldsymbol{\beta}_{-1,x} = (\beta_{2x},\ldots,\beta_{l_1x})'$ and consider $S_{1x}(0,\boldsymbol{\beta}_{-1,x},\boldsymbol{\beta}_z)$, the first component of $\mathbf{S}(\boldsymbol{\beta})$ when $\beta_{1x} = 0$,

$$S_x(0,\boldsymbol{\beta}_{-1,x},\boldsymbol{\beta}_z) = \int_0^{T_o} E\left[\left\{X_1 - \frac{E(X_1\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG)}{E(\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG)}\right\}gC\right]dt. \quad \text{(A.5)}$$

We adopt a symmetrization technique (see, e.g., Bretagnolle and Huber-Carol 1988) to show that $S_x(0,\boldsymbol{\beta}_{-1,x},\boldsymbol{\beta}_z) > 0$. We first show for each t,

$$E(X\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG) - E(\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG)E(X) \le 0.$$

Let $X_1'', \mathbf{X}_{-1}'', \mathbf{Z}''$ be independent copy of $X_1, \mathbf{X}_{-1}$ and $\mathbf{Z}$, and let C'' and G'' be the same functions as C and G but taken at $\mathbf{X}''$ and $\mathbf{Z}''$. By condition (C.5), C does not depend on $\mathbf{X}$ and $\mathbf{Z}$, which implies $C = C''$. Hence

$$\begin{aligned} &E(X_1\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG) \\ &\quad - E(\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}CG)E(X_1) \\ &= \frac{1}{2}E\big\{(X_1 - X_1'') \\ &\quad \times \big(\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}G \\ &\quad - \exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1}'' + \boldsymbol{\beta}_z'\mathbf{Z}''\}G''\big)C\big\}. \end{aligned} \quad \text{(A.6)}$$

We only prove the situation when $\beta_{1x}^{(0)} > 0$. The proof will apply exactly for the situation when $\beta_{1x}^{(0)} < 0$.

When $\beta_{1x}^{(0)} > 0$, $\exp\{\boldsymbol{\beta}_{-1,x}'\mathbf{X}_{-1} + \boldsymbol{\beta}_z'\mathbf{Z}\}G$ is decreasing in X_1, so is its expectation conditional on X_1. Hence the conditional integrand on the right side of (A.6) is nonpositive, and therefore, (A.6) is nonpositive.

Now we want to prove

$$\psi = \int_0^{T_o} E\big[\{X_1 - E(X_1)\}gC\big]dt = -\int_0^{T_o} E\big[\{X_1 - E(X_1)\}C\,dG(t)\big] > 0. \quad \text{(A.7)}$$

In fact, integrating by parts and noting that $C(0)G(0) \equiv 1$, we have $\psi = \int_0^{T_o} E[\{X_1 - E(X_1)\}G\,dC(t)]$. Again using the symmetrization technique, for each t we have

$$2E\big[\{X_1 - E(X_1)\}G\,dC(t)\big] = E\big[(X_1 - X_1'')(G'' - G)\,dC(t)\big] \ge 0,$$

which follows because G is decreasing in X_1 and so is its conditional expectation on X_1. Moreover, $C(t)$ is decreasing in t. Hence the integrand in (A.7) is nonnegative. By continuity, we need only show that the integrand in (A.7) is strictly positive at some point t. In fact, using the same symmetrization technique and conditions (C.3) and (C.4), we can show the integrand in (A.7) is strictly positive at $t = 0$. Hence we have that $S_x(0, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z) > 0$ for any $\boldsymbol\beta_{-1,x}$ and $\boldsymbol\beta_z$.

Now write

$$Q(a,t) = \frac{E[X_1 \exp\{aX_1 + \boldsymbol\beta'_{-1,x}\mathbf{X}_{-1} + \boldsymbol\beta'_z\mathbf{Z}\}CG]}{E[\exp\{aX_1 + \boldsymbol\beta'_{-1,x}\mathbf{X}_{-1} + \boldsymbol\beta'_z\mathbf{Z}\}CG]}.$$

For each $t \ge 0$, we can show, using the Cauchy–Schwarz inequality, that $\frac{\partial}{\partial a}Q(a,t) > 0$.

A direct calculation yields

$$\partial S_{1x}(\beta_{1x}, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z)/\partial\beta_{1x} = -\int_0^{T_o} E(gC)\left\{\frac{\partial}{\partial\beta_{1x}}Q(\beta_{1x}, t) + q(\beta_{1x}, \sigma^2)\right\} dt,$$

where $q(\beta_{1x}, \sigma^2) = p_1(\frac12\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x, \sigma^2)D_{11}^2\beta_{1x}^2 + p(\frac12\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x, \sigma^2)$ and $p_1(v,\sigma^2) = \frac{\partial}{\partial v}p(v,\sigma^2) = \frac{\partial^2}{\partial v^2}\log M_\tau(v,\sigma^2)$. Because τ is a nonnegative random variable, $p(v,\sigma^2) \ge 0$, and by the convexity of $\log M_\tau(v,\sigma^2)$ with respect to v, $p_1(v,\sigma^2) \ge 0$. Therefore, $q(\beta_{1x}, \sigma^2) > 0$. Hence $\partial S_{1x}(\beta_{1x}, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z)/\partial\beta_{1x} < 0$, indicating that $S_{1x}(\beta_{1x}, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z)$ is a decreasing function of β_{1x} for any fixed $\boldsymbol\beta_{-1,x}, \boldsymbol\beta_z$. Because $S(0, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z) > 0$, the solution to $S_x(\beta, \boldsymbol\beta_{-1,x}, \boldsymbol\beta_z) = 0$ for a fixed $\boldsymbol\beta_z$ will be strictly positive. In particular, $\beta^*_{1x}(\sigma^2) > 0$. With exactly the same argument, we can show that $\beta^*_{jz}(\sigma^2) > 0$ for all $1 \le j \le l_2$.

We now prove the monotonicity with respect to σ^2. Differentiating (10) with respect to σ^2 on both sides and collecting terms yields

$$\frac{\partial}{\partial\sigma^2}\boldsymbol\beta^*(\sigma^2) = -\left[p_2\left\{\frac12(\boldsymbol\beta_x^{*\prime}\mathbf{D}\boldsymbol\beta_x^*, \sigma^2)\right\} \times \int_0^{T_o} E(gC)\,dt\right]\mathbf{V}^{-1}(\boldsymbol\beta^*, \sigma^2)\mathbf{F}\mathbf{D}\boldsymbol\beta_x^*, \quad \text{(A.8)}$$

where we abbreviate $\boldsymbol\beta^*(\sigma^2)$ by $\boldsymbol\beta^* = (\boldsymbol\beta_x^{*\prime}, \boldsymbol\beta_z^{*\prime})'$ and

$$p_2\left\{\frac12\boldsymbol\beta_x^{*\prime}\mathbf{D}\boldsymbol\beta_x^*, \sigma^2\right\} \stackrel{\text{def}}{=} \frac{\partial}{\partial\sigma^2}p\left\{\frac12\boldsymbol\beta_x^{*\prime}\mathbf{D}\boldsymbol\beta_x^*, \sigma^2\right\} = \frac{\partial^2}{\partial v\,\partial\sigma^2}\log M_\tau\left(\frac12\boldsymbol\beta_x^{*\prime}\mathbf{D}\boldsymbol\beta_x^*, \sigma^2\right) > 0,$$

by the assumption, and $\mathbf{V}(\boldsymbol\beta^*, \sigma^2)$ is defined in (15). Note that only conditions (R.1) and (R.2) are needed for (A.8) to hold; the other conditions (C.0)–(C.5) are not necessary.

Hence, for any $\mathbf{x} = (x_1, \ldots, x_{l_1+l_2})' \in R^{l_1+l_2}$ such that $\mathbf{x} \ne \mathbf{0}$,

$$\mathbf{x}'\mathbf{V}(\boldsymbol\beta^*, \sigma^2)\mathbf{x} > E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}(\mathbf{x}'\mathbf{X}_*)^2\big)E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}\big) - \big\{E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}\mathbf{x}'\mathbf{X}_*\big)\big\}^2 \ge 0,$$

where the last inequality is by the Cauchy–Schwarz inequality. Here for a column vector $\mathbf{a}$, $\|\mathbf{a}\|_2^2 = \mathbf{a}'\mathbf{a}$. Hence, $\mathbf{V}(\boldsymbol\beta^*, \sigma^2)$ is positive. Note that the off-diagonal element of $\mathbf{V}(\boldsymbol\beta^*, \sigma^2)$ is

$$V_{s_1,s_2} = E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}X^*_{s_1}X^*_{s_2}\big)E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}\big) - E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}X^*_{s_1}\big)E\big(GCe^{\boldsymbol\beta^{*\prime}\mathbf{X}_*}X^*_{s_2}\big),$$

where $1 \le s_1 \ne s_2 \le l_1 + l_2$ and $X^*_{s_1}$ and $X^*_{s_2}$ are the s_1th and s_2th elements of $\mathbf{X}_*$. Again, using the symmetrization technique and the assumption that $\beta_{1x}^{(0)} > 0$ and $\beta_{jz}^{(0)} > 0$, $1 \le j \le l_2$, one can show that $V_{s_1,s_2} < 0$ for $1 \le s_1 \ne s_2 \le l_1 + l_2$. Hence, $\mathbf{V}(\boldsymbol\beta^*, \sigma^2)$ is of DP+ type. By Lemma A.2, $\mathbf{V}^{-1}(\boldsymbol\beta^*, \sigma^2)$ is a matrix with all of its elements positive, and the theorem follows immediately.

A.5 Bias-Correcting Estimator

When $\sigma^2 = 0$ is small, expanding $\boldsymbol\beta^*(\sigma^2)$ around $\sigma^2 = 0$ yields (13), where $\frac{\partial}{\partial\sigma^2}\boldsymbol\beta^*(\sigma^2)|_{\sigma^2=0}$ is given by (A.8) evaluated at $\sigma = 0$. Direct calculation gives $p_2(v, 0) \equiv 1$ for $v \ge 0$. Hence (14), and also (16), hold.

Using the fact that $p_1(v,0) \equiv 0$ and $p(v,0) \equiv 0$ for $v \ge 0$, and hence $q\{\frac12\{\boldsymbol\beta_x^{(0)}\}'\mathbf{D}\boldsymbol\beta_x^{(0)}, 0\} = 0$, we can rewrite $\mathbf{V}(\boldsymbol\beta_0, 0)$, defined by (15), as

$$\mathbf{V}_0 = \int_0^{T_o}\left[\frac{E(GCe^{\boldsymbol\beta_0'\mathbf{X}^*}\mathbf{X}^*(\mathbf{X}_*)')}{E(GCe^{\boldsymbol\beta_0'\mathbf{X}_*})} - \left\{\frac{E(GC\mathbf{X}_*e^{\boldsymbol\beta_0'\mathbf{X}_*})}{E(GCe^{\boldsymbol\beta_0'\mathbf{X}_*})}\right\}^{\otimes 2}\right]E(gC)\,dt.$$

Additionally, denote by

$$S^{(j)}(\boldsymbol\beta, t) = m^{-1}\sum_i \mathbf{W}_{i*}^{\otimes j}Y_i(t)\exp(\boldsymbol\beta'\mathbf{W}_{i*}), \qquad j = 0, 1, 2,$$

and

$$M_\tau = M_\tau\left\{\frac12\{\boldsymbol\beta_x^{(0)}\}'\mathbf{D}\boldsymbol\beta_x^{(0)}, \sigma^2\right\},$$

$$M_\tau^{(1)} = \frac{\partial}{\partial\boldsymbol\beta_x}M_\tau\left(\frac12\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x, \sigma^2\right)\Big|_{\boldsymbol\beta_x = \boldsymbol\beta_x^{(0)}},$$

and

$$M_\tau^{(2)} = \frac{\partial^2}{\partial\boldsymbol\beta_x\,\partial\boldsymbol\beta_x'}M_\tau\left(\frac12\boldsymbol\beta_x'\mathbf{D}\boldsymbol\beta_x, \sigma^2\right)\Big|_{\boldsymbol\beta_x = \boldsymbol\beta_x^{(0)}}.$$

With regularity condition (R.1), we can obtain that on a finite interval $[0, T_0]$,

$$\begin{aligned}\tilde S^{(0)}(\boldsymbol\beta_0, t) &\stackrel{\text{def}}{=} M_\tau^{-1}S^{(0)}(\boldsymbol\beta_0, t)\\ &\to M_\tau^{-1}s^{(0)}(\boldsymbol\beta_0, t)\\ &= E\big(GCe^{\boldsymbol\beta_0'\mathbf{X}_*}\big),\end{aligned}$$

$$\begin{aligned}\tilde{\mathbf{S}}^{(1)}(\boldsymbol\beta_0, t) &\stackrel{\text{def}}{=} M_\tau^{-1}\mathbf{S}^{(1)}(\boldsymbol\beta_0, t) - M_\tau^{-2}S^{(0)}(\boldsymbol\beta_0, t)\mathbf{F}M_\tau^{(1)}\\ &\to M_\tau^{-1}\mathbf{s}^{(1)}(\boldsymbol\beta_0, t) - M_\tau^{-2}s^{(0)}(\boldsymbol\beta_0, t)\mathbf{F}M_\tau^{(1)}\\ &= E\big(GC\mathbf{X}_*e^{\boldsymbol\beta_0'\mathbf{X}_*}\big),\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathbf{S}}^{(2)}(\boldsymbol\beta_0, t) \stackrel{\text{def}}{=}\ & M_\tau^{-1}\mathbf{S}^{(2)}(\boldsymbol\beta_0, t) - M_\tau^{-2}M_\tau^{(1)}\big\{\mathbf{S}^{(1)}(\boldsymbol\beta_0, t)\big\}' \otimes \mathbf{F}\\ &+ S^{(0)}(\boldsymbol\beta_0, t)\mathbf{F}\big[2M_\tau^{-3}M_\tau^{(1)}\big\{M_\tau^{(1)}\big\}' - M_\tau^{-2}M_\tau^{(2)}\big]\mathbf{F}'\\ \to\ & M_\tau^{-1}\mathbf{s}^{(2)}(\boldsymbol\beta_0, t) - M_\tau^{-2}M_\tau^{(1)}\big(\mathbf{s}^{(1)}(\boldsymbol\beta_0, t)\big)' \otimes \mathbf{F}\\ &+ s^{(0)}(\boldsymbol\beta_0, t)\mathbf{F}\big[2M_\tau^{-3}M_\tau^{(1)}\big\{M_\tau^{(1)}\big\}' - M_\tau^{-2}M_\tau^{(2)}\big]\mathbf{F}'\\ =\ & E\big\{GCe^{\boldsymbol\beta_0'\mathbf{X}_*}\mathbf{X}_*\mathbf{X}_*'\big\},\end{aligned}$$

in probability uniformly in a finite interval $[0, T_0]$, where for two square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B} = \mathbf{A}'\mathbf{B}' + \mathbf{BA}$. Hence, with similar arguments as in the proof of lemma 3.1 of Andersen and Gill (1982), we can show that

$$\hat{\mathbf{V}}(\boldsymbol{\beta}_0) \overset{\text{def}}{=} \int_0^{T_o} \left[\frac{\tilde{\mathbf{S}}^{(2)}(\boldsymbol{\beta}_0, t)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0, t)} - \left\{ \frac{\tilde{\mathbf{S}}^{(2)}(\boldsymbol{\beta}_0, t)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0, t)} \right\}^{\otimes 2} \right] \frac{1}{m} \sum_{i=1}^m dN_i(t) \to \mathbf{V}_0$$

in probability. Because $\hat{\mathbf{V}}(\beta)$ is continuous in β, when σ^2 is relatively small, $\hat{\mathbf{V}}(\boldsymbol{\beta}_0)$ can be well approximated by $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\text{naive}})$, where $\hat{\boldsymbol{\beta}}_{\text{naive}}$ is the naive partial likelihood estimate. Also, by martingale theory and the strong law of large numbers, we have $\bar{N} \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^m \int_0^{T_o} dN_i(t) \to \int_0^{T_o} E(gC)\,dt$ almost surely. Thus, a first-order approximation to (13) is

$$\tilde{\boldsymbol{\beta}} = \left\{ \mathbf{I} - \sigma^2 \bar{N} \hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{naive}}) \mathbf{FDF}' \right\}^{-1} \hat{\boldsymbol{\beta}}_{\text{naive}},$$

where $\mathbf{I}$ is an $(l_1 + l_2) \times (l_1 + l_2)$ identity matrix. Note that this formula is indeed applicable in much more general situations, because the derivation of (A.8) holds without conditions (C.0)–(C.5) required for the theoretical bias analyses.

When σ^2 is small, $M_\tau = 1 + O(\sigma^2)$, $M_\tau^{(1)} = O(\sigma^2)$, and $M_\tau^{(2)} = O(\sigma^2)$. Therefore, $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\text{naive}})$ can be further approximated by

$$\hat{\mathbf{V}} = \int_0^{T_o} \left[\frac{\mathbf{S}^{(2)}(\hat{\boldsymbol{\beta}}_{\text{naive}}, t)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\text{naive}}, t)} - \left\{ \frac{\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}_{\text{naive}}, t)}{S^{(0)}(\hat{\boldsymbol{\beta}}_{\text{naive}}, t)} \right\}^{\otimes 2} \right] \frac{1}{m} \sum_{i=1}^m dN_i(t),$$

which results in (16).

A.6 Proof to Theorem 4

Define

$$A(t) = \frac{\sum_{i=1}^m Y_i(t) W_i}{\sum_{i=1}^m Y_i(t)} \quad \text{and}$$

$$\tilde{A}(t) = \frac{\sum_{i=1}^m Y_i(t) W_i e^{\phi(\beta, W_i, \tau_i, t)}}{\sum_{i=1}^m Y_i(t) e^{\phi(\beta, W_i, \tau_i, t)}}.$$

Then

$$m^{-1/2} U = m^{-1/2} \sum_i \int_0^{T_o} \{W_i - \tilde{A}(t)\}\, dN_i(t) + m^{-1/2} \sum_i \int_0^{T_o} \{\tilde{A}(t) - A(t)\}\, dN_i(t). \quad \text{(A.9)}$$

Following Schoenfeld (1983), we can show that the first term on the right side of (A.9) is equal to $m^{-1/2} \sum_i \int_0^{T_o} \{W_i - \tilde{A}(t)\}\, d\tilde{M}_i(t)$, where $\tilde{M}_i(t) = N_i(t) - \int_0^t Y_i(u) \lambda_0(u) \exp\{\phi(\beta, W_i, \tau_i, u)\}\, du$ is the martingale with respect to the filtration $\mathcal{F}_t = \sigma\{N_i(u), Y_i(u+), W_i, \tau_i, i = 1, \ldots, m, 0 \le u < t\}$. Hence, by theorem 4.2 of Andersen and Gill (1982), under the local alternative $\beta_m = m^{-1/2}\eta \to 0$, it converges weakly to a normal distribution with mean 0 and variance

$$\int_0^{T_o} \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t)\, dt,$$

where the expectations involved are taken under $\beta = 0$.

Similarly, by a Taylor expansion about $\beta = 0$ and applying empirical processes theory (e.g., van de Geer 2000, chap. 3), we may show that

$$m^{1/2}\{\tilde{A}(t) - A(t)\} \to \eta \frac{s^{(1)}(t) s_\beta^{(0)}(t) - s^{(0)}(t) s_\beta^{(1)}(t)}{\{s^{(0)}(t)\}^2}$$

in probability uniformly on $[0, T_o]$, where the expectations involved are taken under $\beta = 0$ and the derivatives are evaluated at $\beta = 0$. Hence it follows with the argument of lemma 3.1 of Andersen and Gill (1982), the second term on the right side of (A.9) converges in probability to

$$\int_0^{T_o} \eta \frac{s^{(1)}(t) s_\beta^{(0)}(t) - s^{(0)}(t) s_\beta^{(1)}(t)}{s^{(0)}(t)}\, dt.$$

Then, applying Slutsky's theorem yields the desired result immediately.

A.7 Proof to Theorem 5

Similar to the proof of Theorem 4, under the local alternative $\beta_m = m^{-1/2}\eta \to 0$, the general test statistic $m^{-1/2}\tilde{U}_m$ converges weakly to a normal distribution with mean

$$\int_0^{T_o} \eta r(t) \frac{s^{(1)}(t) s_\beta^{(0)}(t) - s^{(0)}(t) s_\beta^{(1)}(t)}{s^{(0)}(t)}\, dt$$

and variance

$$\int_0^{T_o} r^2(t) \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^2 \right] s^{(0)}(t)\, dt,$$

where the expectations involved are taken under $\beta = 0$ and the derivatives involved are evaluated at $\beta = 0$. Evaluating $s^{(1)}(t)$, $s^{(2)}(t)$, $s_\beta^{(0)}(t)$, and $s_\beta^{(1)}(t)$ under $\beta = 0$ and under the assumption that the censoring time $\tilde{C}$ is independent of W, we can derive the efficacy of test (29),

$$\frac{\eta^2 \sigma_x^4}{(\sigma_x^2 + \sigma^2)^2} \frac{\{\int_0^{T_o} r(t) s^{(0)}(t)\, dt\}^2}{\int_0^{T_o} r^2(t) s^{(0)}(t)\, dt}. \quad \text{(A.10)}$$

For fixed $\eta \neq 0$, $\sigma_x^2 > 0$ and σ^2, by the Schwarz inequality, efficacy (A.10) achieves maximum when $r(t)$ is a nonzero constant function. That is, test (19) achieves maximal efficacy within the class of (29).

[Received January 2003. Revised January 2004.]

REFERENCES

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large-Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Bretagnolle, J., and Huber-Carol, C. (1988), "Effects of Omitting Covariates in Cox's Model for Survival Data," *Scandinavian Journal of Statistics*, 15, 125–138.

Carroll, R., Gail, M. H., and Lubin, J. H. (1993), "Case-Control Studies With Errors in Predictors," *Journal of the American Statistical Association*, 88, 177–191.

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

Carroll, R., and Stefanski, L. (1990), "Approximate Quasi-Likelihood Estimation in Model With Surrogate Predictors," *Journal of the American Statistical Association*, 85, 652–663.

Chen, H. (2002), "Double-Semiparametric Method for Missing Covariates in Cox Regression Models," *Journal of the American Statistical Association*, 97, 565–576.

Chen, H., and Little, R. J. (1999), "Proportional Hazards Regression With Missing Covariates," *Journal of the American Statistical Association*, 94, 896–908.

Clayton, D. G., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Cox, D. R. (1972), "Regression Models and Life-Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

DiRienzo, A. G., and Lagakos, S. W. (2001), "Bias Correction for Score Tests Arising From Mis-Specified Proportional Hazards Regression Models," *Biometrika*, 88, 421–434.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

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

Gold, D. R., Burge, H. A., Carey, V., Milton, D. K., Platts-Mills, T., and Weiss, S. (1999), "Predictors of Repeated Wheeze in the First Year of Life. The Relative Roles of Cockroach, Birth Weight, Acute Lower Respiratory Illness, and Maternal Smoking," *American Journal of Respiratory Critical Care. Medicine*, 160, 227–236.

Herring, A. H., and Ibrahim, J. G. (2001), "Likelihood-Based Methods for Missing Covariates in the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, 96, 292–302.

Higgins, K. M., Davidian, M., and Giltinan, D. M. (1997), "A Two-Step Approach to Measurement Error in Time-Dependent Covariates in Nonlinear Mixed-Effects Models, With Application to IGF-I Pharmacokinetics," *Journal of the American Statistical Association*, 92, 436–448.

Hu, C., and Lin, D. Y. (2002), "Cox Regression With Covariate Measurement Error," *Scandinavian Journal of Statistics*, 29, 637–655.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998), "Estimating the Parameters in the Cox Model When Covariate Variables Are Measured With Error," *Biometrics*, 54, 1407–1419.

Huang, Y., and Wang, C. (2000), "Cox Regression With Accurate Covariates Unascertainable: A Nonparametric Correction Approach," *Journal of the American Statistical Association*, 95, 1209–1219.

Hughes, M. D. (1993), "Regression Dilution in the Proportional Hazards Model," *Biometrics*, 49, 1056–1066.

Kahn, J. O., Lagakos, S. W. et al. (1992), "A Controlled Trial Comparing Continued Zidovudine With Didanosine in Human Immunodeficiency Virus Infection," *New England Journal of Medicine*, 327, 581–587.

Kakihara, Y. (1997), *Multidimensional Second-Order Stochastic Processes*, London: World Scientific.

Keiding, N. (1992), "Independent Delayed Entry," *Survival Analysis: State of the Art*, 211, 309–326.

Kong, F. (1999), "Adjusting Regression Attenuation in the Cox Proportional Hazards Model," *Journal of Statistical Planning and Inference*, 79, 31–44.

Kong, F., and Slud, E. (1997), "Robust Covariate-Adjusted Logrank Tests," *Biometrika*, 84, 847–862.

Lagakos, S. W. (1988), "The Loss in Efficiency From Misspecifying Covariates in Proportional Hazards Regression Models," *Biometrika*, 75, 156–160.

Lange, K. (1998), *Numerical Analysis for Statisticians*, New York: Springer-Verlag.

Leong, T., Lipsitz, S. R., and Ibrahim, J. G. (2001), "Incomplete Covariates in the Cox Model With Applications to Biological Marker Data," *Journal of the Royal Statistical Society*, Ser. C, 50, 467–484.

Lin, D. Y., and Wei, L. J. (1989), "The Robust Inference for the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, 89, 1074–1078.

Lin, D., and Ying, Z. (1993), "Cox Regression With Incomplete Covariate Measurements," *Journal of the American Statistical Association*, 88, 1341–1349.

Lipsitz, S. R., and Ibrahim, J. G. (1998), "Estimating Equations With Incomplete Categorical Covariates in the Cox Model," *Biometrics*, 54, 1002–1013.

Martinussen, T. (1999), "Cox Regression With Incomplete Covariate Measurements Using the EM Algorithm," *Scandinavian Journal of Statistics*, 26, 479–491.

McCullagh, P., and Nelder, J. (1989), *Generalized Linear Models*, London: Chapman and Hall.

Paik, M. C., and Tsai, W. (1997), "On Using the Cox Proportional Hazards Model With Missing Covariates," *Biometrika*, 84, 579–593.

Pons, O. (2002), "Estimation in the Cox Model With Missing Covariate Data," *Journal of Nonparametric Statistics*, 14, 223–247.

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

Schmoor, C., and Schumacher, M. (1997), "Effects of Covariate Omission and Categorization When Analysing Randomized Trials With the Cox Model," *Statistics in Medicine*, 16, 225–237.

Schoenfeld, D. A. (1983), "Sample-Size Formula for the Proportional-Hazards Regression Model," *Biometrics*, 39, 499–503.

Struthers, C. A., and Kalbfleisch, J. D. (1986), "Misspecified Proportional Hazard Models," *Biometrika*, 73, 363–369.

Therneau, T. M., and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, New York: Springer-Verlag.

Tsiatis, A. A., and Davidian, M. (2001), "A Semiparametric Estimator for the Proportional Hazards Model With Longitudinal Covariates Measured With Error," *Biometrika*, 88, 447–458.

van de Geer, S. (2000), *Applications of Empirical Process Theory*, New York: Cambridge University Press.

Wang, C. Y., Xie, S. X., and Prentice, R. L. (2001), "Recalibration Based on an Approximate Relative Risk Estimator in Cox Regression With Missing Covariates," *Statistica Sinica*, 11, 1081–1104.

Wang, N., Lin, X., Gutierrez, R., and Carroll, R. J. (1998), "Bias Analysis and SIMEX Approach in Generalized Linear Mixed Error Models," *Journal of the American Statistical Association*, 93, 249–261.

Xie, S. X., Wang, C. Y., and Prentice, R. L. (2001), "A Risk Set Calibration Method for Failure Time Regression by Using a Covariate Reliability Sample," *Journal of the Royal Statistical Society*, Ser. B, 63, 855–870.

Zhong, M., Sen, P. K., and Cai, J. (1996), "Cox Regression Model With Mismeasured Covariates or Missing Covariate Data," in *Proceedings of the Biometrics Section*, *American Statistical Association*, pp. 323–328.

Zhou, H., and Pepe, M. S. (1995), "Auxiliary Covariate Data in Failure Time Regression," *Biometrika*, 82, 139–149.

Zhou, H., and Wang, C. Y. (2000), "Failure Time Regression With Continuous Covariates Measured With Error," *Journal of the Royal Statistical Society*, Ser. B, 62, 657–665.