Additive Hazards Regression With Covariate Measurement Error

Michal KULICH and D. Y. LIN

The additive hazards model specifies that the hazard function conditional on a set of covariates is the sum of an arbitrary baseline hazard function and a regression function of covariates. This article deals with the analysis of this semiparametric regression model with censored failure time data when covariates are subject to measurement error. We assume that the true covariate is measured on a randomly chosen validation set, whereas a surrogate covariate (i.e., an error-prone version of the true covariate) is measured on all study subjects. The surrogate covariate is modeled as a linear function of the true covariate plus a random error. Only moment conditions are imposed on the measurement error distribution. We develop a class of estimating functions for the regression parameters that involve weighted combinations of the contributions from the validation and nonvalidation sets. The optimal weight can be selected by an adaptive procedure. The resulting estimators are consistent and asymptotically normal with easily estimated variances. Simulation results demonstrate that the asymptotic approximations are adequate for practical use. Illustration with a real medical study is provided.

KEY WORDS: Censoring; Corrected score; Error in variables; Failure time; Surrogate covariate.

1. INTRODUCTION

Many scientific studies are concerned with the effects of covariates on certain outcomes. It is common that a key covariate is very expensive to measure exactly but can be ascertained cheaply with measurement error. In other words, there exists an easily measured *surrogate covariate* that is related to, but not always equal to, the true covariate of interest.

As an example, consider the following problem in Wilms tumor research. Wilms tumor is a rare renal cancer occurring only in children. One of the most important prognostic factors for Wilms tumor patients is the histologic type of the tumor, which can be classified as favorable versus unfavorable. The National Wilms Tumor Study Group (NWTSG) conducted a series of studies to assess the effect of unfavorable histology on the (possibly censored) time to relapse with adjustments for other (discrete and continuous) risk factors (D'Angio et al. 1989; Green et al. 1998). Ideally, the histologic type should be evaluated by the same pathology center for all patients. This was indeed done in the NWTSG studies, but the cost was very high due to the expenses incurred by shipping and storing thousands of tissue samples and maintaining a pathology lab for this purpose. In these studies, the histologic type was initially evaluated by a local pathologist in the institution that treated each particular patient, but this evaluation was less precise and less reliable. It would be much cheaper and logistically simpler if the pathology center evaluated only a fraction of the tissue samples while all of the samples were evaluated by the local pathologists. In fact, such designs are being considered for future NWTSG studies.

It is well known that failure to account for covariate measurement error leads to biased estimators of regression parameters. In particular, the relative risk estimator for unfavorable histology would be biased if one fitted the Cox (1972) proportional hazards model to the NWTSG data by using the histologic type determined by the local rather than the central pathologist. To prevent bias, special estimating techniques must be used. For noncensored data, many such methods are described in the texts by Carroll, Ruppert, and Stefanski (1995) for nonlinear models and Fuller (1987) for linear models. Several methods have been proposed to account for covariate measurement error for the Cox regression with censored data; unfortunately, these methods have substantial limitations. To be specific, the pioneer work of Prentice (1982) assumes that events are rare. This assumption was removed by Zhou and Pepe (1995), but their method requires that all covariates in the nonvalidation set be discrete. The error calibration method of Wang, Hsu, Feng, and Prentice (1996) is only approximately consistent. None of these methods would be satisfactory for the NWTSG studies.

One of the approaches that yields consistent estimators of regression parameters in the presence of covariate measurement error is the corrected score method. The key idea of this approach is to modify the score function so that the bias induced by the measurement error is removed. To this end, a corrected score is constructed such that its expectation over the measurement error distribution approximates the expectation of the standard score function without measurement error. The corrected score method has been used successfully for normal, Poisson, and gamma regression models (Carroll et al. 1995, chap. 6). For the logistic and Cox regression, however, the score functions depend on the covariates through ratios of exponentials. The expectations of such terms are difficult to calculate in general, and in these two cases it can even be shown that no corrected score exists (Nakamura 1992; Stefanski 1989). Under the restrictive as-

Michal Kulich is Assistant Professor, Department of Probability and Statistics, Charles University, 186 75 Prague, Czech Republic (E-mail: *kulich@karlin.mff.cuni.cz*). D. Y. Lin is Professor, Department of Biostatistics, Box 357232, University of Washington, Seattle, WA 98195. This research was supported by the National Institute of Health grants GM47845 and CA42326 and by the Grant Agency of the Czech Republic grant 201/98/P136. The authors are grateful to the National Wilms Tumor Study Group for providing the dataset used in this article and to the editor and referees for helpful comments and suggestions.

Journal of the American Statistical Association
March 2000, Vol. 95, No. 449, Theory and Methods

sumption of normal error with known variance, Nakamura (1992) proposed an approximately corrected score for the Cox model. The asymptotic properties of the resultant estimator are unknown.

We show that the corrected score method can be extended to the additive hazards (AH) model (Breslow and Day 1987, p. 182; Cox and Oakes 1984, p. 74; Lin and Ying 1994). Like the Cox model, the AH model is a semiparametric regression model for censored failure time data. The regression parameters in the AH model pertain to the risk difference, which can be interpreted as the expected numbers of events occurring during a unit time interval due to a unit change in the covariate. This measure is very relevant to public health decisions, because it translates directly into the expected number of disease cases that would be prevented in the population by removing a certain exposure. Thus epidemiologists are keenly interested in estimating this quantity.

In this article we derive consistent estimators for the parameters of the AH model when covariates are subject to measurement error. We achieve this by extending the corrected score approach. Unlike the case of the Cox model, a corrected score for the AH model exists and is easy to calculate. We assume that a surrogate covariate is measured on all subjects and that the exact values of the true covariate are available on a randomly chosen validation set. By introducing an adaptively selected weight, which governs how much information from the nonvalidation set is used in the estimation, we optimize the asymptotic properties of the estimators and ensure that the method works well in moderate-sized samples even if the measurement error is very large. In particular, the asymptotic efficiency of the estimator is guaranteed to be higher than that of the estimator based solely on the complete cases. The proof of the asymptotic results for the corrected score estimator is nontrivial, because the martingale theory is not sufficient. Instead, our proofs rely on the modern empirical process theory. The proposed method yields consistent estimators of regression parameters when the model involves multiple covariates, which can be discrete or continuous. None of the aforementioned methods for censored data shares this feature.

The article is organized as follows. In Section 2 we provide some essential background on AH regression without measurement error. In Section 3 we introduce the corrected score estimator for the univariate case, and in Section 4 we discuss two special error structures and show the results of Monte Carlo studies. We generalize the estimator to the multiple covariates in Section 5 and discuss application of our method to the NWTSG data in Section 6. We finish with some concluding remarks in Section 7 and provide technical details in the Appendix.

2. ADDITIVE HAZARDS REGRESSION

Let T be the failure time, C be the censoring time, and $\mathbf{Z}(\cdot)$ be a p-vector of possibly time-varying covariates. The AH model relates the hazard function of T to $\mathbf{Z}(\cdot)$ through the equation

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) + \boldsymbol{\beta}_0^T \mathbf{Z}(t), \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\beta}_0$ is a p-vector of unknown regression parameters. Denote the cumulative baseline hazard function by $\Lambda_0(t) \equiv \int_0^t \lambda_0(s)\, ds$.

Write $X = \min(T, C)$ and $\Delta = I(T \le C)$, where $I(\cdot)$ is the indicator function. In the absence of measurement error, the observed data consist of independent triplets $(X_i, \Delta_i, \mathbf{Z}_i(t); 0 \le t \le X_i), i = 1, \ldots, n$. Let $N_i(t) \equiv I(X_i \le t, \Delta_i = 1)$ count the number of observed failures for the ith subject up to time t, and let $Y_i(t) \equiv I(X_i \ge t)$ indicate whether or not the ith subject is at risk for failure at time t. Write $\bar{\mathbf{Z}}(t) = \sum_{i=1}^n Y_i(t)\mathbf{Z}_i(t) / \sum_{i=1}^n Y_i(t)$, which is the average covariate vector for the subjects at risk at time t.

Lin and Ying (1994) introduced a pseudoscore function for $\boldsymbol{\beta}_0$ as $\mathbf{U}_A(\boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{\psi}_i^{(A)}(\boldsymbol{\beta})$, where

$$\boldsymbol{\psi}_i^{(A)}(\boldsymbol{\beta}) = \int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}\{dN_i(t) - \mathbf{Z}_i(t)^T \boldsymbol{\beta} Y_i(t)\, dt\}.$$

The parameter vector $\boldsymbol{\beta}_0$ is estimated by solving the estimating equation $\mathbf{U}_A(\boldsymbol{\beta}) = \mathbf{0}$. The resulting estimator takes an explicit form,

$$\hat{\boldsymbol{\beta}}_A = \left[\sum_{i=1}^n \int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2} Y_i(t)\, dt\right]^{-1} \times \left[\sum_{i=1}^n \int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}\, dN_i(t)\right], \tag{2}$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$.

It is easy to see that

$$\mathbf{U}_A(\boldsymbol{\beta}_0) = \sum_{i=1}^n \int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}\, dM_i(t),$$

where

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\{d\Lambda_0(s) + \mathbf{Z}_i(s)^T \boldsymbol{\beta}_0\, ds\}, \tag{3}$$

which is a martingale. Using this martingale representation, Lin and Ying showed that $n^{-1/2}\mathbf{U}_A(\boldsymbol{\beta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_0)$ converge in distribution to zero-mean normal random vectors with covariance matrices $\boldsymbol{\Sigma}_A$ and $\mathbf{D}_A^{-1}\boldsymbol{\Sigma}_A\mathbf{D}_A^{-1}$, where $\boldsymbol{\Sigma}_A = E\int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2}\, dN_i(t)$ and $\mathbf{D}_A = E\int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2} Y_i(t)\, dt$. In addition, the cumulative baseline hazard $\Lambda_0(t)$ can be consistently estimated by

$$\hat{\Lambda}_A(t) = \int_0^t \sum_{i=1}^n dN_i(s) \Big/ \sum_{i=1}^n Y_i(s) - \int_0^t \bar{\mathbf{Z}}(s)^T \hat{\boldsymbol{\beta}}_A\, ds. \tag{4}$$

3. CORRECTED PSEUDOSCORE ESTIMATOR

Throughout the article, we assume that model (1) holds. In this section we propose a corrected pseudoscore and

study the asymptotic properties of the estimator assuming that Z is a time-independent scalar. We study the multivariate case in Section 5.

3.1 Construction of the Corrected Pseudoscore Estimator

Let $\xi_1, \ldots, \xi_n$ be independent Bernoulli variables with $P(\xi_i = 1) = \alpha$, where α is a known constant. If $\xi_i = 1$, then the ith subject is included in the validation set; otherwise, it is included in the nonvalidation set. Denote the validation set by $\mathcal{V} \equiv \{i: \xi_i = 1\}$. In principle, $P(\xi_i = 1)$ could depend on (Δ_i, X_i) or on any other observed data. For simplicity of description, however, we assume that α is a constant and that ξ_i is independent of (Z_i, Δ_i, X_i). The notation $\bar{\xi}_i$ for $1 - \xi_i$ and $\bar{\alpha}$ for $1 - \alpha$ will be used in the sequel.

Write W_i for the surrogate covariate. The data are given by iid quintuples $(X_i, \Delta_i, Z_i, W_i, \xi_i), i = 1, \ldots, n$; however, for $i \notin \mathcal{V}$, we observe only $(X_i, \Delta_i, W_i, \xi_i)$. Suppose that the data are observed on the time interval $[0, \tau]$, where $0 < \tau < \infty$ is a fixed quantity. The goal is to consistently estimate β_0 using only the observed data.

We impose the following conditions on the surrogate covariate W_i:

S1. $E(W_i|Z_i) = \gamma_0 + \gamma_1 Z_i$, where $\gamma_1 \neq 0$
S2. $\mathrm{var}(W_i|Z_i) \equiv V(Z_i) = v_0 + v_1 Z_i + v_2 Z_i^2 > 0$
S3. W_i and W_j are independent given Z_i, Z_j $(i \neq j)$
S4. Given Z_i, W_i is independent of $N_i(t)$ and $Y_i(t)$.

Conditions S1 and S2 specify the first two conditional moments of W given Z. The variance function $V(\cdot)$ in condition S2 has a specific form, which is motivated later. Condition S3 ensures that errors are independent. Condition S4 states that W is a true surrogate; that is, all of the effects of W on failure and censoring are mediated through Z. The parameters $\gamma_0, \gamma_1, v_0, v_1$, and v_2 involved in conditions S1 and S2 are called the *error parameters*. Note that the model for W specified by S1 and S2 could be made more general by including more covariates, provided that the linear form of both model components is preserved.

Let us first assume that the error parameters are known. Define the bias-adjusted covariate $\hat{Z}_i = \gamma_1^{-1}(W_i - \gamma_0)$. Clearly, $E(\hat{Z}_i|Z_i) = Z_i$ and $\mathrm{var}(\hat{Z}_i|Z_i) = \gamma_1^{-2} V(Z_i)$. A naive method for estimating β_0 is obtained by taking the pseudoscore U_A and substituting the values of $\hat{Z}_i$ for the unknown Z_i in the nonvalidation set. Specifically, with $R_i = \xi_i Z_i + \bar{\xi}_i \hat{Z}_i$ and $\bar{R}(t) = \sum Y_i(t) R_i / \sum Y_i(t)$, the naive estimator $\hat{\beta}_{\mathrm{N}}$ is defined as the solution of $U_{\mathrm{N}}(\beta) \equiv \sum_i \psi_i^{(N)}(\beta) = 0$, where

$$\psi_i^{(N)}(\beta) = \int_0^\tau \{R_i - \bar{R}(t)\}\{dN_i(t) - R_i \beta Y_i(t)\, dt\}. \quad (5)$$

Although $E(R_i|Z_i) = Z_i$, the naive estimator is not consistent for β_0 because (as we show later) $EU_{\mathrm{N}}(\beta_0) \neq 0$. To obtain an unbiased pseudoscore, we calculate $EU_{\mathrm{N}}(\beta)$ conditionally on the σ-algebra $\mathcal{F}_\tau$ generated by the failure, censoring, and (true) covariate information for all the subjects over $[0, \tau]$. Straightforward calculation shows that $E(U_{\mathrm{N}}(\beta)|\mathcal{F}_\tau) = U_A(\beta) - \gamma_1^{-2}\beta \sum_{i=1}^n \bar{\xi}_i V(Z_i) X_i + o_P(n^{1/2})$. Because $EU_A(\beta_0) = 0$, an asymptotically unbiased pseudoscore could be obtained by adding $\gamma_1^{-2}\beta \sum_{i=1}^n \bar{\xi}_i V(Z_i) X_i$ to $U_{\mathrm{N}}(\beta)$. However, $V(Z_i)$ depends on the unobserved Z_i. This is why a restriction on V is needed. Indeed, if condition S2 holds, then $E\{1 + \gamma_1^{-2} v_2)^{-1} V(\hat{Z}_i)|\mathcal{F}_\tau\} = V(Z_i)$. Thus the bias correction is $(\gamma_1^2 + v_2)^{-1}\beta \sum_{i=1}^n \bar{\xi}_i V(\hat{Z}_i) X_i$.

We define the corrected pseudoscore by combining the contributions of the validation set subjects and of the nonvalidation set subjects. To increase efficiency, we introduce a downweighting constant $0 \leq w \leq 1$, which reduces the influence of nonvalidation observations. Thus the average covariate at risk is estimated by

$$\bar{R}(t, w) \equiv \frac{\sum (\xi_i Z_i + w\bar{\xi}_i \hat{Z}_i) Y_i(t)}{\sum (\xi_i + w\bar{\xi}_i) Y_i(t)}. \quad (6)$$

We occasionally drop the argument w from $\bar{R}(t, w)$ and other quantities involving w.

The pseudoscore contribution of a validation set subject is

$$\psi_i^{(\mathrm{V})}(\beta) = \int_0^\tau \{Z_i - \bar{R}(t)\}\{dN_i(t) - Z_i \beta Y_i(t)\, dt\}. \quad (7)$$

Nonvalidation set members contribute

$$\psi_i^{(\mathrm{NV})}(\beta) = \int_0^\tau \{\hat{Z}_i - \bar{R}(t)\}\{dN_i(t) - \hat{Z}_i \beta Y_i(t)\, dt\} + \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2}\, \beta X_i. \quad (8)$$

These two parts are combined into a single estimating function,

$$U_C(\beta, w) = \sum_{i=1}^n \{\xi_i \psi_i^{(\mathrm{V})}(\beta) + w\bar{\xi}_i \psi_i^{(\mathrm{NV})}(\beta)\}. \quad (9)$$

The corrected pseudoscore (CS) estimator $\hat{\beta}_C \equiv \hat{\beta}_C(w)$ is defined as the solution of $U_C(\beta, w) = 0$. The estimator has a closed form similar to (2),

$$\hat{\beta}_C = \left(\sum_{i=1}^n \rho_i \int_0^\tau \left[\{R_i - \bar{R}(t)\}^2 - \bar{\xi}_i \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2} \right] Y_i(t)\, dt \right)^{-1} \times \left[\sum_{i=1}^n \rho_i \int_0^\tau \{R_i - \bar{R}(t)\}\, dN_i(t) \right],$$

where $\rho_i = \xi_i + w\bar{\xi}_i$. Because w spans the interval [0, 1], (9) defines an entire class of estimators. When $w = 0$, no nonvalidation data are used, and the resulting estimator is the complete-case AH estimator based on the validation set $\mathcal{V}$.

3.2 Asymptotic Properties of Corrected Pseudoscore Estimators With Known Error Parameters

To establish the asymptotic properties of $\hat{\beta}_{\mathrm{C}}(w)$, we first show that $U_{\mathrm{C}}(\beta_0, w)$ can be approximated by a sum of independent random variables. Denote $\pi_k(t) = EZ_i^k Y_i(t), k = 0, 1$, and $e(t) = \pi_1(t)/\pi_0(t)$. We work under the following regularity conditions:

R1. $\Lambda_0(\tau) < \infty$
R2. $P(Y_i(\tau) = 1) > 0$
R3. $E|Z_i|^3 < \infty$
R4. $\Sigma_A(\beta_0) = E\int_0^\tau \{Z_i - e(t)\}^2\, dN_i(t) > 0$.

Conditions R3 and R4 are necessary conditions. Conditions R1 and R2 could probably be weakened, but then the proofs would get more complicated. The asymptotic linearity of the CS is stated in the following theorem.

Theorem 1. Under conditions R1–R4,

$$n^{-1/2}U_C(\beta_0, w) = n^{-1/2}\sum_{i=1}^n \{\xi_i\tilde{\psi}_i^{(\mathrm{V})}(\beta_0) + w\bar{\xi}_i\tilde{\psi}_i^{(\mathrm{NV})}(\beta_0)\} + o_P(1),$$

where $\tilde{\psi}_i^{(\mathrm{V})}(\beta_0) = \int_0^\tau \{Z_i - e(t)\}\, dM_i(t)$ and

$$\tilde{\psi}_i^{(\mathrm{NV})}(\beta_0) = \int_0^\tau \{\hat{Z}_i - e(t)\}\{dN_i(t) - Y_i(t)\, d\Lambda_0(t) - \hat{Z}_i\beta_0 Y_i(t)\, dt\} + \beta_0\, \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2}\, X_i. \tag{10}$$

The proof of Theorem 1 is given in Appendix section A.1. It is easy to see that $E\tilde{\psi}_i^{(\mathrm{V})}(\beta_0) = E\tilde{\psi}_i^{(\mathrm{NV})}(\beta_0) = 0$ and that $(\tilde{\psi}_i^{(\mathrm{V})}, \tilde{\psi}_i^{(\mathrm{NV})})$ are iid. We also have var $\tilde{\psi}_i^{(\mathrm{V})}(\beta_0) = \Sigma_A(\beta_0)$, and we denote $\Sigma_B(\beta_0) =$ var $\tilde{\psi}_i^{(\mathrm{NV})}(\beta_0)$. The asymptotic normality of the CS follows immediately from Theorem 1.

Corollary 1. If $EW_i^4 < \infty$, then $n^{-1/2}U_C(\beta_0, w)$ converges in distribution to a zero-mean normal random variable with variance

$$\Sigma_C(\beta_0, w) = \alpha\Sigma_A(\beta_0) + \bar{\alpha}w^2\Sigma_B(\beta_0).$$

The next theorem proclaims the consistency and asymptotic normality of the CS estimator and is proven in Appendix section A.2.

Theorem 2. Suppose that conditions R1–R4 hold and that $EW_i^4 < \infty$. Then $n^{1/2}(\hat{\beta}_C(w) - \beta_0)$ converges in distribution to a zero-mean normal random variable with variance $\Sigma_C(\beta_0, w)/D_C^2(w)$, where $D_C(w) = (\alpha + \bar{\alpha}w)D_A$.

We now turn to estimation of the limiting variance of $\hat{\beta}_C$. Because D_C is the limit of the negative derivative of $n^{-1}U_C(\beta)$, its empirical counterpart is

$$\hat{D}_C \equiv n^{-1}\sum_{i=1}^n \left(\xi_i \int_0^\tau \{Z_i - \bar{R}(t)\}^2 Y_i(t)\, dt + \bar{\xi}_i w \int_0^\tau \left[\{\hat{Z}_i - \bar{R}(t)\}^2 - \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2}\right] Y_i(t)\, dt\right).$$

To find a consistent estimator for $\Sigma_B(\beta_0)$, we need to estimate the cumulative baseline hazard Λ_0. We propose using the estimator $d\hat{\Lambda}_C(t) \equiv \sum dN_i(t)/\sum Y_i(t) - \bar{R}(t)\hat{\beta}_C\, dt$, which is an obvious extension of (4). The estimator $\hat{\Lambda}_C(t)$ may not be monotone, but can be modified to be so along the lines of work of Lin and Ying (1994). The consistency of $\hat{D}_C$ and $\hat{\Lambda}_C(t)$ is stated and proven in Lemma A.2 of Appendix section A.3. We estimate Σ_A and Σ_B by $\hat{\Sigma}_A \equiv (\alpha n)^{-1}\sum \xi_i \int_0^\tau \{Z_i - \bar{R}(t)\}^2\, dN_i(t)$ and $\hat{\Sigma}_B \equiv (\bar{\alpha}n)^{-1}\sum_{i=1}^n \bar{\xi}_i(\hat{\psi}_i^{(\mathrm{NV})})^2$, where

$$\hat{\psi}_i^{(\mathrm{NV})} = \int_0^\tau \{\hat{Z}_i - \bar{R}(t)\}\left\{dN_i(t) - \frac{Y_i(t)}{\sum Y_j(t)}\sum dN_j(t)\right\} - \hat{\beta}_C \int_0^\tau \left[\{\hat{Z}_i - \bar{R}(t)\}^2 - \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2}\right] Y_i(t)\, dt.$$

We then estimate $\Sigma_C(\beta_0, w)$ by $\hat{\Sigma}_C \equiv \alpha\hat{\Sigma}_A + w^2\bar{\alpha}\hat{\Sigma}_B$. The consistency of $\hat{\Sigma}_C$ is proven in Appendix section A.3.

The CS estimator $\hat{\beta}_C(w)$ is consistent for any weight $w \in [0, 1]$. However, its limiting variance depends on the choice of w. By differentiating the variance function with respect to w, we see that the limiting variance achieves its minimum at $w_{\mathrm{opt}} \equiv \Sigma_A(\beta_0)/\Sigma_B(\beta_0)$. Clearly, $w_{\mathrm{opt}} \le 1$. The optimum weight w_{opt} defines an estimator that is the most efficient among all $\hat{\beta}_C(w)$ $(0 \le w \le 1)$. In particular, $\hat{\beta}_C(w_{\mathrm{opt}})$ is guaranteed to be more efficient than the complete-case estimator $\hat{\beta}_C(0)$. The optimal weight w_{opt} can be estimated by $\hat{w}_{\mathrm{opt}} = \hat{\Sigma}_A/\hat{\Sigma}_B$. It follows from Lemma A.3 that $\hat{w}_{\mathrm{opt}} \to_p w_{\mathrm{opt}}$. Because $\hat{w}_{\mathrm{opt}}$ depends on $\hat{\beta}_C$, these two estimators can be obtained simultaneously by a simple iterative procedure, or, perhaps preferably, a preliminary consistent estimator of β_0 can be calculated with $w = 0$ and plugged into the expression for $\hat{w}_{\mathrm{opt}}$, which in turn is used to obtain the final $\hat{\beta}_C(\hat{w}_{\mathrm{opt}})$.

3.3 Corrected Pseudoscore Estimators With Unknown Error Parameters

We now relax the assumption that the error parameters describing the conditional distribution of W given Z are known. From now on, let $\boldsymbol{\theta}$ denote all the error parameters taken as a column vector and let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$. The unknown error parameters can be estimated from the validation set and be replaced by their estimators in the CS. The resulting CS estimator remains consistent and asymptotically normal, although its variance needs to be adjusted.

Suppose that a consistent estimator $\hat{\theta}$ of $\boldsymbol{\theta}_0$ is obtained by solving $\sum \xi_i\boldsymbol{\phi}_i(\boldsymbol{\theta}) = \mathbf{0}$. Quasi-likelihood estimating equations provide a consistent estimator in any case, although they can be replaced by another set of unbiased estimating equations. We substitute $\hat{\theta}$ for the unknown $\boldsymbol{\theta}_0$ involved in (9) and then solve $U_C(\beta, \hat{\theta}) = 0$. The resulting estimator is still denoted by $\hat{\beta}_C$.

By the Taylor expansion of $U_C(\beta_0, \hat{\theta})$ at $\boldsymbol{\theta}_0$,

$$U_C(\beta_0, \hat{\theta}) = \sum_{i=1}^n \{\xi_i\tilde{\psi}_i^{(\mathrm{V})}(\beta_0) + w\bar{\xi}_i\tilde{\psi}_i^{(\mathrm{NV})}(\beta_0, \boldsymbol{\theta}_0) - \alpha^{-1}\bar{\alpha}w\xi_i\boldsymbol{\Gamma}(\beta_0, \boldsymbol{\theta}_0)^T\phi_i(\boldsymbol{\theta}_0)\} + o_P(n^{1/2}),$$

where $\boldsymbol{\Gamma}(\beta_0, \boldsymbol{\theta}_0)^T = \mathbf{D}_2(\beta_0, \boldsymbol{\theta}_0)\mathbf{D}_1^{-1}(\boldsymbol{\theta}_0), \mathbf{D}_1(\boldsymbol{\theta}) = -E\{\partial\phi(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\}$, and $\mathbf{D}_2(\beta, \boldsymbol{\theta}) = -E\{\partial\tilde{\psi}_i^{(\mathrm{NV})}(\beta, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\}$. Write

$\boldsymbol{\Sigma}_\phi = \text{var } \boldsymbol{\phi}_i(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_E(\beta_0) = \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_\phi \boldsymbol{\Gamma}$. It then follows from the central limit theorem and Appendix A.3.6 of Carroll et al. (1995) that $n^{-1/2} U_C(\beta_0, \hat{\theta})$ is asymptotically zero-mean normal with variance

$$\Sigma_{CN}(\beta_0, w) = \alpha \Sigma_A(\beta_0) + \bar{\alpha} w^2 \Sigma_B(\beta_0) + \frac{\bar{\alpha}^2}{\alpha} w^2 \Sigma_E(\beta_0),$$

and $n^{1/2}(\hat{\beta}_C - \beta_0)$ is asymptotically zero-mean normal with variance Σ_{CN}/D_C^2. One may estimate the limiting variance of $\hat{\beta}_C$ consistently by replacing Σ_{CN} and D_C with their respective sample estimators. The resulting variance estimator adjusts for the fact that the vector of error parameters $\boldsymbol{\theta}_0$ is estimated from the validation dataset. The explicit form of the correction vector $\boldsymbol{\Gamma}$ varies from application to application but is not difficult to derive for each special case.

Again, by differentiating the variance function $\Sigma_{CN}(\beta_0, w)$ with respect to w, we see that the optimal weight w_{opt} with unknown error parameters can be estimated by $\hat{\Sigma}_A/(\hat{\Sigma}_B + \bar{\alpha}/\alpha \hat{\Sigma}_E)$, where $\hat{\Sigma}_E$ is a consistent estimator of Σ_E. Such an estimator can be easily obtained.

4. SPECIAL CASES

The class of CS estimators described in Section 3 is quite general. The actual estimating procedure depends on the type of covariate and on the error structure. In this section we consider two important special cases: continuous covariate measured with error of constant variance and misclassified binary covariate.

4.1 Continuous Covariate Measured With Error of Constant Variance

This is the simplest practical application of the CS method. Let $W_i = Z_i + \varepsilon_i$, where $\varepsilon_1, \ldots, \varepsilon_n$ are independent zero-mean random variables that are independent of (Z_i, N_i, Y_i). Let var $\varepsilon_i = \sigma^2$, where σ^2 is unknown. In this case $\gamma_1 = 1, \gamma_0 = 0, \hat{Z}_i = W_i$, and $V(Z_i) = V(\hat{Z}_i) = \sigma^2$. We estimate σ^2 by $\hat{\sigma}^2 = \sum \xi_i (W_i - Z_i)^2 / \sum \xi_i$, which is the solution of $\sum \xi_i \phi_i = 0$ with $\phi_i = (W_i - Z_i)^2 - \sigma^2$. The variance correction factor Γ is just $-\beta_0 E X_i$.

A simulation study was conducted to investigate the performance of the CS estimator in this setting. The population distribution of the covariate Z was standard normal truncated at ± 1.96. The random error ε was zero-mean normal with standard error σ. The baseline hazard was constant, and censoring was uniform. The total number of observations was 1,000, about 55% of which were censored; 1,000 datasets were generated and analyzed for each setting. Four estimates of β_0 were calculated: the full-data estimate given by (2), the naive estimate defined by (5), the complete-case estimate with $w = 0$, and the CS estimate $\hat{\beta}_C(\hat{w}_{\text{opt}})$, which substitutes the complete-case estimate into $\hat{\Sigma}_A$ and $\hat{\Sigma}_B$ to yield $\hat{w}_{\text{opt}}$ and estimates the error variance σ^2 from the validation data. Simulation results are summarized in Table 1.

The table shows the behavior of the four estimators under different validation set sizes and magnitudes of measurement error. The measurement error variance ranges from one-quarter of to four times the covariate variance. The validation set consists of either one-fifth or one-half of the observations. As expected, the naive estimator is severely biased. The CS estimator removes the bias quite well. The sample standard error of the CS estimates is always smaller than that of the complete-case estimates, even when the error variance is large. Moreover, the estimated standard error of the CS estimator is on average close to the sample standard error of the simulated CS estimates. The coverage probabilities of the 95% confidence intervals based on the CS estimator and its estimated standard error range from .945 to .961, which is quite satisfactory.

The average estimated optimal weight was around .8, .5, and .2 for σ equal to .5, 1, and 2. To see how much efficiency was gained by using the optimal weight, we also evaluated the more intuitive estimator $\hat{\beta}_C(1)$. For small and moderate errors, it was only slightly worse than $\hat{\beta}_C(\hat{w}_{\text{opt}})$; but its performance was very poor when the measurement error was large. For example, with $\sigma = 2$ and $\alpha = .5$, the standard error of $\hat{\beta}_C(1)$ was about .364, compared to .214 for $\hat{\beta}_C(\hat{w}_{\text{opt}})$ and .230 for the complete-case estimator. When α was .2, the estimator with $w = 1$ even broke down in some simulations. Thus using the optimal weight not only minimizes the variance, but also improves the numerical behavior.

4.2 Misclassified Binary Covariate

Suppose now that Z is a binary covariate with $P(Z = 1) = p_Z$ and W is a misclassified surrogate for Z. Denote the sensitivity of W for Z by $\eta \equiv P\ (W = 1 | Z = 1)$ and the specificity of W for Z by $\nu \equiv P\ (W = 0 | Z = 0)$. Then the expectation of W given Z is $E(W|Z) = 1 - \nu + (\eta + \nu - 1)Z$. Thus $E(W|Z)$ has the required linear form with $\gamma_0 = 1 - \nu$ and $\gamma_1 = \eta + \nu - 1$. We assume that $\gamma_1 \neq 0$. The conditional variance of W given Z is $V(Z) = \text{var}(W|Z) = \nu(1-\nu) + \{\eta(1-\eta) - \nu(1-\nu)\}Z$, also a linear function of Z. Because all of the error parameters are functions of sensitivity and specificity, we parameterize the problem in terms of η and ν. It is, however, convenient to write γ_1 for $\eta + \nu - 1$.

The bias-corrected version of W is $\hat{Z} = \gamma_1^{-1}(W - 1 + \nu)$. An unbiased estimator for $V(Z)$ is given by $V(\hat{Z})$. Let us write $\bar{Z}_i$ for $1 - Z_i, \bar{W}_i$ for $1 - W_i$ and $\bar{p}_Z$ for $1 - p_Z$. The error parameters $\boldsymbol{\theta}_0 = (\eta_0, \nu_0)^T$ can be estimated by $\hat{\eta} = n_{11}/(n_{10} + n_{11})$ and $\hat{\nu} = n_{00}/(n_{00} + n_{01})$, where $n_{ij} = \sum_k \xi_k I\ (Z_k = i, W_k = j)$. That is, $\hat{\theta}$ is the solution of $\sum \xi_i \boldsymbol{\phi}_i(\boldsymbol{\theta}) = \mathbf{0}$ with $\boldsymbol{\phi}_i(\boldsymbol{\theta}) = (Z_i W_i - \eta Z_i, \bar{Z}_i \bar{W}_i - \nu \bar{Z}_i)^T$.

Simple algebra shows that the error parameter variance adjustment is given by

$$\boldsymbol{\Gamma}^T \boldsymbol{\phi}_i(\boldsymbol{\theta}) = \frac{\beta_0}{\gamma_1} \{p_Z^{-1}(K - E Z_i X_i) Z_i (\eta - W_i) + (\bar{p}_Z)^{-1}(K - E \bar{Z}_i X_i) \bar{Z}_i (\nu - \bar{W}_i)\},$$

where $K = \int_0^\tau e(t)\{1 - e(t)\}\pi_0(t)\ dt$. This expression is quite easy to estimate.

The finite-sample performance of the CS estimator for misclassified binary covariates was addressed by a simulation study. Datasets of 1,000 subjects were generated. The

Table 1. Simulation Study of a Continuous Covariate With Random Error Under $\lambda(t|Z) = 3.4 + \beta_0 Z$ Where $\beta_0 = .3$

Method[a]	Validation set proportion	Error SE σ	Mean Estimate	Sample SE	Average estimated SE	Coverage probability[b]	Power[c]
F	[d]	[d]	.300	.157	.162	.960	.468
CC	.2	[d]	.300	.361	.370	.956	.124
	.5	[d]	.296	.232	.230	.943	.256
N	.2	.5	.245	.142	.148	.939	.378
		1	.167	.115	.120	.809	.269
		2	.071	.075	.078	.166	.143
	.5	.5	.262	.152	.153	.953	.416
		1	.190	.131	.132	.864	.293
		2	.104	.094	.093	.442	.194
CS	.2	.5	.299	.173	.178	.961	.398
		1	.312	.212	.217	.955	.300
		2	.315	.291	.314	.961	.148
	.5	.5	.298	.171	.172	.953	.424
		1	.290	.194	.189	.946	.348
		2	.312	.213	.214	.945	.296

[a] F = full data, CC = complete case, N = naive, CS = corrected score.
[b] Proportion of 95% confidence intervals based on estimated parameter and estimated SE that cover the true value of β_0.
[c] Power for testing H_0: $\beta_0 = 0$ against $\beta_0 \neq 0$.
[d] Estimator does not depend on this parameter.

baseline hazard was constant, and censoring was uniform such that failures were observed for about 420 subjects. The probability of being exposed was $p_Z = .5$. The validation set consisted of either 20% or 50% of the subjects. For each of the 1,000 datasets generated, four estimates were calculated: the full-data estimate, the naive estimate, the complete-case estimate, and the CS estimate with estimated optimal weight and estimated error parameters. The results are summarized in Table 2.

Table 2 includes the results for different misclassification rates in the binary covariate: the sensitivity and specificity combinations of (.7, .7), (.9, .7) and (.9, .9) yield misclassification rates of 30%, 20%, and 10%. The table indicates that the CS method successfully removes the bias produced by the naive estimator. The standard error of the CS estimator is always lower than the standard error of the complete-case estimator, even when the misclassification rate is 30% and the validation set is small. The 95% coverage probabilities for the CS estimator are in the range .934–.957, suggesting that both the CS estimator and its variance estimator work reasonably well with binary covariates.

5. MULTIPLE COVARIATES

When one covariate is measured with error, a surrogate is available, and the other covariates are always observed, the pseudoscore correction is applied just to the component of the pseudoscore corresponding to the covariate measured with error. Suppose that the p-vector of covariates $\mathbf{Z}_i$ is partitioned as $\mathbf{Z}_i(t)^T = (Z_{1i}, \mathbf{Z}_{2i}(t)^T)$, where the second part is a possibly time-dependent $(p-1)$-vector. Let Z_{1i} be observed only if $\xi_i = 1$, and let W_i be the surrogate for Z_{1i}. Suppose that conditions S1–S4 are fulfilled for Z_{1i} and W_i. Let $\hat{Z}_{1i}$ be the bias-adjusted surrogate and define the observed covariate vector $\mathbf{R}_i(t)^T = (R_{1i}, \mathbf{Z}_{2i}(t)^T)$, where $R_{1i} = \xi_i Z_{1i} + \bar{\xi}_i \hat{Z}_{1i}$. Other p-vectors, such as $\boldsymbol{\beta}_0$ and $\tilde{\boldsymbol{\psi}}_i^{(\mathrm{V})}$,

Table 2. Simulation Study of a Misclassified Binary Covariate Under $\lambda(t|Z) = 2.6 + \beta_0 Z$ where $\beta_0 = 0.9$

Method[a]	Validation set proportion	Sensitivity η	Specificity ν	Mean Estimate	Sample SE	Average estimated SE	Coverage probability[b]	Power[c]
F	[d]	[d]	[d]	.897	.314	.299	.937	.842
CC	.2	[d]	[d]	.911	.689	.679	.950	.272
	.5	[d]	[d]	.911	.444	.425	.944	.576
N	.2	.9	.9	.616	.253	.247	.780	.707
		.9	.7	.390	.195	.190	.241	.539
		.7	.7	.178	.131	.130	.000	.288
	.5	.9	.9	.708	.268	.263	.868	.764
		.9	.7	.481	.231	.216	.512	.598
		.7	.7	.245	.161	.156	.015	.366
CS	.2	.9	.9	.900	.371	.360	.946	.711
		.9	.7	.932	.442	.435	.957	.583
		.7	.7	.940	.554	.537	.944	.411
	.5	.9	.9	.908	.343	.331	.946	.775
		.9	.7	.902	.387	.362	.934	.693
		.7	.7	.912	.408	.395	.938	.646

[a] F = full data, CC = complete case, N = naive, CS = corrected score.
[b] Proportion of 95% confidence intervals based on estimated parameter and estimated SE that cover the true value of β_0.
[c] Power for testing H_0: $\beta_0 = 0$ against $\beta_0 \neq 0$.
[d] Estimator does not depend on this parameter.

are similarly partitioned into the first component and the remaining $(p-1)$ components.

To define the corrected pseudo-score, we must generalize the downweighting constant w to the multivariate case. We propose weighting the pseudoscore contributions of the nonvalidation subjects by a general $p \times p$ invertible matrix $\mathbf{\Omega}$. Thus the contribution of the ith subject is weighted by the matrix $\mathbf{A}_i = \xi_i \mathbf{I}_p + \bar{\xi}_i \mathbf{\Omega}$, with $E\mathbf{A}_i = \alpha \mathbf{I}_p + \bar{\alpha}\mathbf{\Omega}$. The weighted at-risk covariate average is defined as

$$\bar{\mathbf{R}}(t) = \left[\sum_{i=1}^{n} \mathbf{A}_i Y_i(t)\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{A}_i \mathbf{R}_i(t) Y_i(t)\right].$$

The multivariate corrected pseudo-score is given by

$$\begin{aligned}\mathbf{U}_C(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Bigg[&\mathbf{A}_i \int_0^{\tau} \{\mathbf{R}_i - \bar{\mathbf{R}}(t)\} \\ &\times \{dN_i(t) - \mathbf{R}_i(t)^T \boldsymbol{\beta} Y_i(t)\, dt\} \\ &+ \bar{\xi}_i \mathbf{A}_i \boldsymbol{\beta}_{(1)} \frac{V(\hat{Z}_i)}{\gamma_1^2 + v_2} X_i \Bigg], \end{aligned} \tag{11}$$

where $\boldsymbol{\beta}_{(1)} = (\beta_1, 0, \ldots, 0)^T$. The estimator based on (11) has a closed-form solution. By analogy with the univariate case, the normalized pseudoscore is asymptotically linear; that is,

$$\begin{aligned} &n^{-1/2}\mathbf{U}_C(\boldsymbol{\beta}_0) \\ &= n^{-1/2} \sum_{i=1}^{n} \mathbf{A}_i \begin{bmatrix} \xi_i \tilde{\psi}_{1i}^{(\mathrm{V})}(\boldsymbol{\beta}_0) + \bar{\xi}_i \tilde{\psi}_{1i}^{(\mathrm{NV})}(\boldsymbol{\beta}_0) \\ \xi_i \tilde{\boldsymbol{\psi}}_{2i}^{(\mathrm{V})}(\boldsymbol{\beta}_0) + \bar{\xi}_i \tilde{\boldsymbol{\psi}}_{2i}^{(\mathrm{NV})}(\boldsymbol{\beta}_0) \end{bmatrix} + o_P(1) \\ &= n^{-1/2} \sum_{i=1}^{n} \{\xi_i \tilde{\boldsymbol{\psi}}_i^{(\mathrm{V})}(\boldsymbol{\beta}_0) + \bar{\xi}_i \mathbf{\Omega} \tilde{\boldsymbol{\psi}}_i^{(\mathrm{NV})}(\boldsymbol{\beta}_0)\} + o_P(1), \end{aligned}$$

where

$$\tilde{\psi}_{1i}^{(\mathrm{V})}(\boldsymbol{\beta}_0) = \int_0^{\tau} \{Z_{1i} - e_1(t)\}\, dM_i(t),$$

$$\tilde{\psi}_{1i}^{(\mathrm{NV})}(\boldsymbol{\beta}_0) = \int_0^{\tau} \{\hat{Z}_{1i} - e_1(t)\}\, d\tilde{M}_i(t) + \frac{V(\hat{Z}_{1i})}{\gamma_1^2 + v_2} \beta_{01} X_i,$$

$$\tilde{\boldsymbol{\psi}}_{2i}^{(\mathrm{V})}(\boldsymbol{\beta}_0) = \int_0^{\tau} \{\mathbf{Z}_{2i}(t) - \mathbf{e}_2(t)\}\, dM_i(t),$$

$$\tilde{\boldsymbol{\psi}}_{2i}^{(\mathrm{NV})}(\boldsymbol{\beta}_0) = \int_0^{\tau} \{\mathbf{Z}_{2i}(t) - \mathbf{e}_2(t)\}\, d\tilde{M}_i(t),$$

and $d\tilde{M}_i(t) = dN_i(t) - Y_i(t) d\Lambda_0(t) - \mathbf{R}_i(t)^T \boldsymbol{\beta}_0 Y_i(t)\, dt$. Note that $\tilde{\boldsymbol{\psi}}_i^{(\mathrm{V})}(\boldsymbol{\beta}_0)$ and $\tilde{\boldsymbol{\psi}}_i^{(\mathrm{NV})}(\boldsymbol{\beta}_0)$ are mean-zero iid terms. Thus $n^{-1/2}\mathbf{U}_C(\boldsymbol{\beta}_0)$ converges in distribution to a p-variate zero-mean normal random vector with covariance matrix $\mathbf{\Sigma}_C(\boldsymbol{\beta}_0) \equiv \alpha \mathbf{\Sigma}_A(\boldsymbol{\beta}_0) + \bar{\alpha}\mathbf{\Omega}\mathbf{\Sigma}_B(\boldsymbol{\beta}_0)\mathbf{\Omega}^T$, where $\mathbf{\Sigma}_A$ is the covariance matrix of the full-data pseudoscore and $\mathbf{\Sigma}_B$ is the covariance matrix of $\tilde{\boldsymbol{\psi}}_i^{(\mathrm{NV})}$.

It follows that $n^{1/2}(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0)$ is asymptotically zero-mean normal with covariance matrix $\mathbf{D}_A^{-1} \mathbf{\Phi}(\mathbf{\Omega}) \mathbf{D}_A^{-T}$, where $\mathbf{\Phi}(\mathbf{\Omega}) = (\alpha \mathbf{I} + \bar{\alpha}\mathbf{\Omega})^{-1}(\alpha \mathbf{\Sigma}_A + \bar{\alpha}\mathbf{\Omega}\mathbf{\Sigma}_B\mathbf{\Omega}^T)(\alpha \mathbf{I} + \bar{\alpha}\mathbf{\Omega}^T)^{-1}$. The estimation of the limiting covariance matrix of $n^{1/2}(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0)$ and the pseudoscore variance adjustment for unknown error parameters are straightforward generalizations of the univariate results developed in Section 3.

To maximize the efficiency of the estimator, we need to find $\mathbf{\Omega}_0$ such that $\mathbf{\Phi}(\mathbf{\Omega}) - \mathbf{\Phi}(\mathbf{\Omega}_0) \geq \mathbf{0}$ for any invertible weight matrix $\mathbf{\Omega}$. It can be shown through the Gauss–Markov theorem that the optimal weight matrix is given by $\mathbf{\Omega}_0 = \mathbf{\Sigma}_A \mathbf{\Sigma}_B^{-1}$. Clearly, $\mathbf{\Phi}(\mathbf{\Omega}_0) = (\alpha \mathbf{\Sigma}_A^{-1} + \bar{\alpha}\mathbf{\Sigma}_B^{-1})^{-1}$.

If there are more covariates measured with error but all true covariates are available on the validation set $\mathcal{V}$ and have surrogates that are always observed, then the bias correction must be applied jointly to the pseudoscore components corresponding to the mismeasured covariates. The correction then involves the conditional covariance matrix of the vector of surrogates given the vector of true covariates. If this matrix can be estimated unbiasedly from the observed data, then the CS method yields consistent estimators as well.

6. A REAL EXAMPLE

In this section we present an analysis of relapse rate for patients enrolled in two of the studies conducted by the National Wilms Tumor Study Group (NWTSG-3 and NWTSG-4). We are interested in the excess risk of relapse associated with unfavorable histology. The central and institutional histology assessments can be regarded as the true covariate and the surrogate. In the NWTSG studies, central histology was evaluated for nearly all patients. However, it would be interesting to see what the results would be if the pathology center had reviewed only a random subsample of the entire cohort. In the analysis, we would like to adjust for other important risk factors—namely, tumor stage, age at diagnosis, and study number. The stage was coded by I–IV, indicating spread of the tumor from localized to metastatic. Age was measured in years. The study number could be important, because the patients in NWTSG-4 received an improved treatment.

The dataset comprised 4,157 subjects (1,920 in NWTSG-3 and 2,237 in NWTSG-4) with known relapse status, follow-up time, stage, age, and both histology evaluations. Unfavorable central histology was recorded for 11.5% of the patients (11.6% in NWTSG-3 and 11.4% in NWTSG-4) and 14.8% of the patients relapsed (15.6% in NWTSG-3 and 14.1% in NWTSG-4). The overall sensitivity and specificity of institutional histology for central histology were .72 and .98, but these parameters differed by stage, age, and study number. Preliminary analyses of relapse rates confirmed that, given central histology, institutional histology had no effect on relapse.

We first fitted model (1) to the whole data, with central histology, age, and three dummy variables for stage and one dummy variable for study as the covariates. Then we drew validation sets at random and calculated the CS, naive, and complete-case estimators assuming that central histology was available only for the validation set subjects.

Table 3. Estimates of Excess Risk Parameters for Relapse Calculated From the NWTSG Data

Method[a]	Unfavorable histology	Stage 2	Stage 3	Stage 4	Age	NWTSG-4
F	.0579	.0090	.0120	.0216	.0052	−.0029
	(.0054)	(.0018)	(.0021)	(.0034)	(.0018)	(.0019)
N[b]	.0474	.0087	.0108	.0188	.0062	−.0019
	(.0050)	(.0018)	(.0020)	(.0034)	(.0018)	(.0019)
CC[b]	.0586	.0089	.0121	.0218	.0054	−.0031
	(.0100)	(.0034)	(.0038)	(.0063)	(.0033)	(.0036)
CS[b]	.0581	.0087	.0118	.0215	.0055	−.0030
	(.0072)	(.0020)	(.0023)	(.0038)	(.0018)	(.0020)

NOTE: Estimated standard errors are in parentheses.
[a] F = full data, N = naive, CC = complete case, CS = corrected score.
[b] The results are averaged over 1,000 validation sets.

The validation set proportion α was .3. Because η and ν depend on stage, age, and study number, we divided age into two categories, under 1 year and over 1 year, and estimated η and ν separately in each of the 16 combinations of stage, age, and study number. The regression results are shown in Table 3. The full-data parameter estimate for histology was .0579. Because time was measured in years, this number pertains to how many extra relapses per 1 person-year of follow-up are expected in the unfavorable histology group compared to the favorable histology group. To evaluate the naive, complete-case, and CS estimators, we drew 1,000 different validation sets at random and averaged the estimates and the estimated standard errors. We found that both the CS and complete-case estimates were on average close to the full-data estimates and that the naive estimates were biased in nearly all components. The standard errors of the complete-case estimates were nearly twice as large as those of the full-data estimates. The standard error of the CS estimate for histology was on average 33% larger than that of the full-data estimate, whereas other parameters had only slightly elevated standard errors. The estimated optimal weight matrix for the CS estimator had diagonal elements (.34, .76, .77, .71, .92, .85). All off-diagonal elements were much closer to 0.

Because it requires central histology assessment for only about 1,250 subjects, the proposed design would have saved the pathology center about 2,900 tissue sample analyses. This 70% savings would have increased the standard errors of the regression parameter estimators by less than 33%.

7. DISCUSSION

A key assumption for the CS estimator to work is the independence of measurement error and failure/censoring. We formulated it as the conditional independence of the surrogate and failure/censoring given the true covariate. If this condition is not met, then the CS estimator is inconsistent. When measurement error depends on other observed covariates, we can model the surrogate covariate as a linear function of the true covariate and the other covariates, as was done in the NWTSG example.

The CS estimator is well defined even when there is no validation set. Indeed, we can just set $\alpha = 0$ and $\xi_i = 0$ for all i. Then the error parameters must be known, say from extraneous sources, and the results given in Sections 3.1 and 3.2 apply with no change. However, the weight w drops out and, as indicated in the last paragraph of Section 4.1, the CS estimator may break down if the measurement error is large.

As discussed by Breslow and Day (1987) and Lin and Ying (1994), the choice between the proportional and additive hazards models depends on the biological knowledge, the empirical evidence, and the association parameter of interest. Although both models can provide adequate fit to any given dataset if appropriate time-dependent covariates are introduced, the more parsimonious model is normally preferable. In the presence of covariate measurement error, it is much easier to fit the additive hazards model than the proportional hazards model. If the covariates are all discrete, or if one is willing to discretize continuous covariates, then the method of Zhou and Pepe (1995) can be used to fit the Cox model; otherwise, one would have to resort to the approximate solutions of Prentice (1982), and Wang et al. (1996) or just use the additive hazards model. Simple goodness-of-fit methods, such as plots of cumulative hazard estimates, can be used to check the adequacy of the proportional hazards and additive hazards models, although no method is available to formally test which of the two models fits the data better even in the absence of measurement error.

The theoretical development presented in this article extends naturally to the case of general counting processes $N_i(t)$ that can take multiple jumps. With some refinements of the proofs, it also works for arbitrary at-risk processes $Y_i(t)$. Furthermore, the error model determined by assumptions S1 and S2 could be generalized by allowing the error parameters to vary with other covariates, as in the NWTSG example, or by incorporating other covariates directly into the error model.

For simplicity, we have assumed that the validation set is chosen by independent Bernoulli sampling with a common probability. As discussed in Section 3.1, this assumption is not essential. In fact, we may let the selection process depend on the data. Specifically, let $p_i = P\,(\xi_i = 1|\mathcal{F}_\tau)$. Then a corrected pseudoscore can be formed by weighting each subject's contribution to the pseudoscore and to the numerator and the denominator of $\bar{R}(t, w)$ by the respective inverse sampling probability. The resulting weighted CS estimator is consistent and asymptotically normal, although the vari-

ance has a slightly more complicated form. The details will be presented elsewhere.

APPENDIX: PROOFS OF ASYMPTOTIC PROPERTIES

A.1 Asymptotic Linearity of the Corrected Pseudoscore

To prove Theorem 1, we start with a technical lemma.

Lemma A.1. Let $A_n(t), A_n^*(t)$, and $B_n(t)$ be three sequences of bounded processes on $[0,\tau]$. Suppose that (a) $B_n(t)$ converges weakly to a tight limit $B(t)$ with almost surely continuous sample paths; (b) $A_n(t)$ and $A_n^*(t)$ are monotone in t; and (c) there exist processes $A(t)$ and $A^*(t)$, both right continuous at 0 and left continuous at τ, such that $\sup_{0\le t\le\tau}|A_n(t)-A(t)|\to_p 0$ and $\sup_{0\le t\le\tau}|A_n^*(t)-A^*(t)|\to_p 0$. Then

$$\sup_{0\le t\le\tau}\left|\int_0^t\{A_n(s)A_n^*(s)-A(s)A^*(s)\}\,dB_n(s)\right|\to_p 0.$$

Proof. By the Skorokhod strong embedding theorem (Shorack and Wellner 1986, p. 47), there exists another probability space on which A_n, A_n^*, and B_n can be defined so that $\{A_n(t), A_n^*(t), B_n(t)\}$ converges to $\{A(t), A^*(t), B(t)\}$ almost surely. We work on this probability space. Through integration by parts,

$$\int_0^t A_n(s)\,dB_n(s)=A_n(t)B_n(t)-\int_0^t B_n(s-)\,dA_n(s).$$

By Helly's second theorem (Serfling 1980, p. 352) and the assumptions of this lemma, it can be shown that

$$\sup_{0\le t\le\tau}\left|\int_0^t B_n(s-)\,dA_n(s)-\int_0^t B(s)\,dA(s)\right|\xrightarrow{\text{a.s.}}0$$

(Kulich 1997, lemma 4.2). Thus

$$\sup_{0\le t\le\tau}\left|\int_0^t A_n(s)\,dB_n(s)-\int_0^t A(s)\,dB(s)\right|\overset{\text{a.s.}}{\to}0.$$

By the same argument,

$$\sup_{0\le t\le\tau}\left|\int_0^t A_n(s)A_n^*(s)\,dB_n(s)-\int_0^t A(s)A^*(s)\,dB(s)\right|\to 0.$$

Likewise,

$$\sup_{0\le t\le\tau}\left|\int_0^t A(s)A^*(s)\,dB_n(s)-\int_0^t A(s)A^*(s)\,dB(s)\right|\to 0.$$

Both convergences hold almost surely on the new probability space and hence weakly on the original space. Hence, the statement of the lemma follows.

Proof of Theorem 1. By the law of large numbers,

$$n^{-1}\sum_{i=1}^n\rho_iR_i^kY_i(t)\to_p(\alpha+w\bar\alpha)\pi_k(t)\tag{A.1}$$

for $k=0,1$. It follows from corollary III.2 of Andersen and Gill (1982) that the convergence is uniform in t. Hence

$$\sup_{0\le t\le\tau}|\bar R(t)-e(t)|\to_p 0\quad\text{for any}\quad w\in[0,1].\tag{A.2}$$

We can write $U_C(\beta_0)=U_1+wU_2$, where $U_1=\sum_{i=1}^n\xi_i\int_0^\tau\{Z_i-\bar R(t)\}\,dM_i(t)$ and

$$\begin{aligned}U_2 &= \sum_{i=1}^n\bar\xi_i\int_0^\tau\{\hat Z_i-\bar R(t)\}\\ &\quad\times\{dN_i(t)-Y_i(t)d\Lambda_0(t)-\hat Z_i\beta_0Y_i(t)\,dt\}\\ &\quad+\sum_{i=1}^n\bar\xi_i\,\frac{V(\hat Z_i)}{\gamma_1^2+v_2}\,\beta_0X_i.\end{aligned}$$

We make the decomposition $n^{-1/2}U_2=\sum_{k=1}^4Q_k$, where

$$\begin{aligned}Q_1&=n^{-1/2}\sum_{i=1}^n\bar\xi_i\int_0^\tau\{\hat Z_i-e(t)\}\\ &\qquad\times\{dN_i(t)-Y_i(t)d\Lambda_0(t)-\hat Z_i\beta_0Y_i(t)\,dt\},\\ Q_2&=n^{-1/2}\sum_{i=1}^n\bar\xi_i\int_0^\tau\{e(t)-\bar R(t)\}\,dM_i(t),\\ Q_3&=n^{-1/2}\sum_{i=1}^n\bar\xi_i\int_0^\tau\{e(t)-\bar R(t)\}(Z_i-\hat Z_i)\beta_0Y_i(t)\,dt,\end{aligned}$$

and

$$Q_4=n^{-1/2}\sum_{i=1}^n\bar\xi_i\,\frac{V(\hat Z_i)}{\gamma_1^2+v_2}\,\beta_0X_i.$$

Note that Q_2 is a martingale integral whose variance converges to 0 by (A.2) and condition R3. Thus $Q_2\to_p 0$. We can write Q_3 as $\beta_0\int_0^\tau\{e(t)-\bar R(t)\}\,dB_n(t)$, where

$$B_n(t)=n^{-1/2}\sum_{i=1}^n\bar\xi_i(Z_i-\hat Z_i)\int_0^tY_i(s)\,ds.$$

Because $B_n(t)$ is the difference of two nondecreasing processes, it follows from example 2.11.16 of van der Vaart and Wellner (1996, p. 215) that $B_n(t)$ converges weakly to a process $B(t)$ with continuous sample paths. Set $A_n^{-1}(t)=n^{-1}\sum\rho_iY_i(t)$ and $A_n^*(t)=n^{-1}\sum\rho_iR_iY_i(t)$. Because $R_i=\max(R_i,0)-\max(-R_i,0)$, we can assume without loss of generality that $R_i\ge 0$. Then $A_n(t)$ is nondecreasing and $A_n^*(t)$ is nonincreasing. It follows from (A.1) that $A_n(t)\to_p A(t)$ and $A_n^*(t)\to_p A^*(t)$ uniformly in t. It then follows from lemma A.1 that $Q_3\to_p 0$. Hence

$$\begin{aligned}n^{-1/2}U_2 &= n^{-1/2}\sum_{i=1}^n\bar\xi_i\int_0^\tau\{\hat Z_i-e(t)\}\\ &\quad\times\{dN_i(t)-Y_i(t)d\Lambda_0(t)-\hat Z_i\beta_0Y_i(t)\,dt\}\\ &\quad+\frac{V(\hat Z_i)}{\gamma_1^2+v_2}\,\beta_0X_i+o_P(1).\end{aligned}$$

By similar arguments, $n^{-1/2}U_1$ can be written as

$$n^{-1/2}U_1=n^{-1/2}\sum_{i=1}^n\xi_i\int_0^\tau\{Z_i-e(t)\}\,dM_i(t)+o_P(1).$$

This completes the proof of Theorem 1.

A.2 Asymptotic Normality of Corrected Pseudoscore Estimators

Proof of Theorem 2. By Taylor expansion, $U_C(\hat\beta_C)-U_C(\beta_0)=(\hat\beta_C-\beta_0)\partial U_C(\beta)/\partial\beta$. The partial derivative is a constant, because U_C is linear in β. Thus $n^{-1/2}U_C(\beta_0)=n^{1/2}(\hat\beta_C-$

$\beta_0)A_n$, where $A_n \equiv -n^{-1}\partial U_C(\beta)/\partial\beta$ converges in probability to $D_C = (\alpha + w\bar{\alpha})D_A$ (see lemma A.2 in Sec. A.3). By Corollary 1, $n^{-1/2}U_C(\beta_0)$ converges in distribution to zero-mean normal with variance $\Sigma_C(\beta_0)$. The consistency of $\hat{\beta}_C$ follows from the asymptotic normality. Thus Theorem 2 is proven.

A.3 Consistency of Variance Estimators

Lemma A.2. Let conditions R1–R4 hold and let $EW_i^4 < \infty$. Then

(a) $\hat{D}_C \to_p D_C$, and
(b) $\sup_{0\le t\le\tau} |\hat{\Lambda}_C(t) - \Lambda_0(t)| \to_p 0$.

Proof. For part (a), the validation part of $\hat{D}_C$ can be written as

$$\begin{aligned} & n^{-1}\sum \xi_i \int_0^\tau \{Z_i - \bar{R}(t)\}^2 Y_i(t)\,dt \\ &= n^{-1}\sum \xi_i \int_0^\tau \{Z_i - e(t)\}^2 Y_i(t)\,dt \\ &\quad + 2n^{-1}\sum \xi_i \int_0^\tau \{Z_i - e(t)\}\{e(t) - \bar{R}(t)\}Y_i(t)\,dt \\ &\quad + n^{-1}\sum \xi_i \int_0^\tau \{e(t) - \bar{R}(t)\}^2 Y_i(t)\,dt. \end{aligned}$$

The first term on the right side converges in probability to αD_A. The absolute value of the second term is bounded by $\sup_t |e(t) - \bar{R}(t)|$ multiplied by a term that converges to a constant. Similarly, the third term tends to 0. By the same arguments, $\bar{R}(t)$ in the nonvalidation part of $\hat{D}_C$ may be replaced by $e(t)$ with an error of $o_P(1)$. It follows that the nonvalidation part tends to $w\bar{\alpha}D_A$. Hence $\hat{D}_C \to_p (\alpha + w\bar{\alpha})D_A = D_C$.

For part (b), we express $\hat{\Lambda}_C(t) - \Lambda_0(t)$ as

$$\begin{aligned} & \int_0^t \frac{\sum_i(\xi_i + w\bar{\xi}_i)\,dM_i(s)}{\sum_j(\xi_j + w\bar{\xi}_j)Y_j(s)} + \int_0^t \frac{\beta_0 - \hat{\beta}_C}{\sum_j(\xi_j + w\bar{\xi}_j)Y_j(s)} \\ &\quad \times \sum_i (\xi_i + w\bar{\xi}_i)(\xi_i Z_i + \bar{\xi}_i\hat{Z}_i)Y_i(s)\,ds \\ &\quad + \int_0^t \frac{w\beta_0}{\sum_j(\xi_j + w\bar{\xi}_j)Y_j(s)} \sum_i \bar{\xi}_i(Z_i - \hat{Z}_i)Y_i(s)\,ds \\ &\quad + \sum_i \int_0^t \left\{ \frac{1}{\sum_j Y_j(s)} - \frac{\xi_i + w\bar{\xi}_i}{\sum_j(\xi_j + w\bar{\xi}_j)Y_j(s)} \right\} dN_i(s). \end{aligned}$$

The first term is a martingale integral with a variance function converging to 0 in probability uniformly in t. The second term converges to 0 uniformly in t by consistency of $\hat{\beta}_C$. The third term tends to 0 uniformly in t as well, because its integrand converges uniformly to 0 by (A.1). The fourth term is asymptotically equivalent to

$$n^{-1}\sum_i \int_0^t \left\{ \frac{1}{\pi_0(s)} - \frac{\xi_i + w\bar{\xi}_i}{(\alpha + w\bar{\alpha})\pi_0(s)} \right\} dN_i(s),$$

which converges to 0 by condition R2 and the law of large numbers. Thus $\hat{\Lambda}_C(t)$ is uniformly consistent.

Lemma A.3. Let the conditions of Lemma A.2 hold, and let Z_i be bounded. Then $\hat{\Sigma}_C \to_p \Sigma_C(\beta_0)$.

Proof. Without loss of generality, assume that $\gamma_0 = 0, \gamma_1 = 1$, and $v_2 = 0$. It suffices to show that $\hat{\Sigma}_B \to_p \Sigma_B(\beta_0)$. Denote $\hat{G}_i(t) = \{W_i - \bar{R}(t)\}Y_i(t)$ and $G_i(t) = \{W_i - e(t)\}Y_i(t)$. The nonvalidation part of $\hat{\Sigma}_C$ can be written as $w^2\sum_{k=1}^6\sum_{i=1}^n \bar{\xi}_i\hat{Q}_{ik}$, where $\hat{Q}_{i1} = \{W_i - \bar{R}(X_i)\}^2$,

$$\hat{Q}_{i2} = \left\{ \int_0^\tau \hat{G}_i(t)\,d\hat{\Lambda}_C(t) \right\}^2,$$

$$\hat{Q}_{i3} = \left(\int_0^\tau [\{W_i - \bar{R}(t)\}W_i - V(W_i)]Y_i(t)\hat{\beta}_C\,dt \right)^2,$$

and $\hat{Q}_{i4}, \hat{Q}_{i5}$, and $\hat{Q}_{i6}$ are the remaining cross-product terms. We need to prove that $n^{-1}\sum_{i=1}^n \bar{\xi}_i\hat{Q}_{ik} \to_p \bar{\alpha}Q_k, k = 1, \ldots, 6$, where Q_k is the expectation of $\hat{Q}_{ik}$ with $\hat{\beta}_C$ replaced by $\beta_0, \hat{\Lambda}_C$ replaced by Λ_0, and $\bar{R}(t)$ replaced by $e(t)$.

For $k = 1$, this is trivial. For $k = 2$, we have to show that $\int_0^\tau\int_0^\tau n^{-1}\sum \bar{\xi}_i\hat{G}_i(s)\hat{G}_i(t)\,d\hat{\Lambda}_C(s)\,d\hat{\Lambda}_C(t)$ converges in probability to $\int_0^\tau\int_0^\tau \bar{\alpha}EG_i(s)G_i(t)\,d\Lambda_0(s)\,d\Lambda_0(t)$. The first step is to verify

$$\sup_{0\le s,t\le\tau} \left| n^{-1}\sum_{i=1}^n \bar{\xi}_i\hat{G}_i(s)\hat{G}_i(t) - \bar{\alpha}EG_i(s)G_i(t) \right| \to_p 0.$$

This can be done by approximating $\bar{\alpha}EG_i(s)G_i(t)$ by the empirical average of $\bar{\xi}_iG_i(s)G_i(t)$ and by using (A.2) to show that the average of $\hat{G}_i(s)\hat{G}_i(t) - G_i(s)G_i(t)$ converges to 0 in probability. Now it suffices to show that

$$\left| \int_0^\tau \int_0^\tau \bar{\alpha}EG_i(s)G_i(t)\{d\hat{\Lambda}_C(s) - d\Lambda_0(s)\}\{d\hat{\Lambda}_C(t) - d\Lambda_0(t)\} \right|$$

converges to 0 in probability. But this convergence follows from integration by parts and from Lemma 2, part (b). The remaining cases $(k = 3, \ldots, 6)$ can be treated similarly.

[Received April 1998. Revised June 1999.]

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.

Breslow, N. E., and Day, N. E. (1987), *Statistical Models in Cancer Research, 2. The Design and Analysis of Cohort Studies*, Lyon: IARC.

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

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

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.

D'Angio, G. J., Breslow, N., Beckwith, J. B., Evans, A., Baum, H., de Lorimier, A., Fernbach, D., Hrabovsky, E., Jones, B., and Kelalis, P. (1989), "Treatment of Wilms Tumor: Results of the Third National Wilms Tumor Study," *Cancer*, 64, 349–360.

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

Green, D. M., Breslow, N. E., Beckwith, J. B., Finklestein, J. Z., Grundy, P. G., Thomas, P. R. M., Kim, T., Shochat, S., Haase, G. M., Ritchey, M. L., Kelalis, P. P., and D'Angio, G. J. (1998), "Comparison Between Single-Dose and Divided-Dose Administration of Dactinomycin and Doxorubicin for Patients With Wilms Tumor. A Report From the National Wilms Tumor Study Group," *Journal of Clinical Oncology*, 16, 237–245.

Kulich, M. (1997), "Additive Hazards Regression With Incomplete Covariate Data," Ph.D. dissertation, Seattle: University of Washington.

Lin, D. Y., and Ying, Z. (1994), "Semiparametric Analysis of the Additive Risk Model," *Biometrika*, 81, 61–71.

Nakamura, T. (1992), "Proportional Hazards Model With Covariates Subject to Measurement Error," *Biometrics*, 48, 829–838.

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

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.

Stefanski, L. A. (1989), "Unbiased Estimation of a Nonlinear Function of a Normal Mean With Applications to Measurement Error Models," *Communications in Statistics, Part A*, 18, 4335–4358.

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

Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997), "Regression Calibration in Failure Time Regression," *Biometrics*, 53, 131–145.

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