Corrected Score Estimation in the Proportional Hazards Model with Misclassified Discrete Covariates

David M. Zucker

Department of Statistics, Hebrew University

Mount Scopus, 91905 Jerusalem, Israel

and

Donna Spiegelman

Departments of Epidemiology and Biostatistics

Harvard School of Public Health

677 Huntington Avenue, Boston, MA 02115, USA

Short Title: Cox Model With Misclassification

Corresponding Author:

David M. Zucker

Department of Statistics, Hebrew University

Mt. Scopus, 91905 Jerusalem, Israel

Phone: +972-2-588-1291

Fax: +972-2-588-3549

E-mail: mszucker@mscc.huji.ac.il

Summary

We consider Cox proportional hazards regression when the covariate vector includes error-prone discrete covariates along with error-free covariates which may be discrete or continuous. The misclassification in the discrete error-prone covariates is allowed to be of arbitrary form. Building on work of Nakamura and his colleagues, we present a corrected score method for this setting. The method can handle all three major study designs (internal validation design, external validation design, and replicate measures design), both functional and structural error models, and time-dependent covariates satisfying a certain "localized error" condition. We derive the asymptotic properties of the method, and indicate how to adjust the covariance matrix of the regression coefficient estimates to account for estimation of the misclassification matrix. We present results of a finite-sample simulation study under Weibull survival with a single binary covariate having known misclassification rates. The performance of the method described here was similar to that of related methods we have examined in previous work. Specifically, our new estimator performed as well as or, in a few cases, better than the full Weibull maximum likelihood estimator. We also present simulation results for our method for the case where the misclassification probabilities are estimated from an external replicate measures study. Our method generally performed well in these simulations. The new estimator has a broader range of applicability than many other estimators proposed in the literature, including those described in our own earlier work, in that it can handle time-dependent covariates with an arbitrary misclassification structure. We illustrate the method on data from a study of the relationship between dietary calcium intake and distal colon cancer.

KEY WORDS: Errors in variables, nonlinear models, proportional hazards

1. INTRODUCTION

Many regression analyses involve explanatory variables that are measured with error. It is well known that failing to account for covariate error can lead to biased estimates of the regression coefficients. For linear models, theory for handling covariate error has been developed over the past 50 or more years; Fuller [1] provides an authoritative exposition. For nonlinear models, theory has been developing over the past 20 or so years. Carroll, Ruppert, and Stefanski [2] provide a comprehensive summary of the first 10-15 years of development; currently, the covariate error problem for nonlinear models remains an active research area. In particular, beginning with Prentice [3], a growing literature has developed on the Cox [4] proportional hazards survival regression model when some covariates are measured with error. In this paper, we focus on discrete covariates subject to misclassification, which are of interest in many epidemiological studies.

Three basic design setups are of interest. In all three designs, we have a main survival cohort for which surrogate covariate measurements and survival time data are available on all individuals. The three designs are as follows: (1) the internal validation design, where the true covariate values are available on a subset of the main survival cohort, (2) the external validation design, where the measurement error distribution is estimated from data outside the main survival study, and (3) the replicate measurements design, where replicate surrogate covariate measurements are available, either on a subset of the survival study cohort or on individuals outside the main survival study. Also, two types of models for the measurement error are of interest (see [1, p. 2] and [2, Sec. 1.2]): structural models, where the true covariates are random variables, and functional models, where the true covariates are fixed values. Structural model methods generally involve estimation of some aspect of the distribution of the true covariate values; in functional model methods, this process is avoided.

The Cox survival regression model with covariate errors has been examined in a number of settings. Much of the existing work focuses on the independent additive

error model, which assumes that the observed covariate value is equal to the true value plus a random error whose distribution is independent of the true value. In the case of discrete covariates subject to misclassification, this model practically never holds, and so the methods built upon it do not apply.

The work of Zhou and Pepe [5] and of Zhou and Wang [6] deals with the internal validation design without the independent additive error assumption. Their approach involves empirically estimating the conditional mean of a certain function of the true covariate vector conditional on the observed covariate vector. This process entails stratification or smoothing with respect to the observed covariate vector. When the covariate vector is of moderate to high dimension, the "curse of dimensionality" causes this approach to break down, even if only one of the covariates is error-prone. Chen [7] presents an alternate method for the internal validation design. His method combines the regression coefficient estimate based on the validation sample only with information gleaned from the rest of the main study cohort. Chen's approach assumes that it is possible to form a satisfactory initial estimate of the regression coefficient vector based on the validation sample alone. This is not the case, however, for studies where the event rate is low to moderate, the main study sample size is in the thousands, and the validation study sample size is a few hundred. Thus, in such cases, which often arise in practice, Chen's approach breaks down. In addition, the methods of Zhou and Pepe, Zhou and Wang, and Chen do not cover the external validation or replicate measures setups.

Spiegelman, McDermott, and Rosner [26] and Wang et al. [9] discuss the simple and well-known regression calibration method, which applies to all three design setups. This method, however, is only an approximate method and does not yield a consistent estimator. Zucker and Spiegelman [10] and Zucker [11] present general methods suitable for all three designs, but their methods cover only time-independent covariates and their approaches do not seem generalizable to time-dependent covariates. Hu, Tsiatis, and Davidian [12] present a method that can handle time-dependent covariates under

all three designs in a more general setting, but their approach is complex and its asymptotic properties were not formally examined. Most of the methods cited above apply only to structural models.

In principle, it is possible to apply methods that have been developed in the missing covariate literature, such as that of Herring and Ibrahim [13]. Most of this work deals with the internal validation setup, though the approach could be extendable to the other two design setups. These methods, however, are highly complex, and they cannot effectively handle the functional model setting or time-dependent covariates.

Thus, the currently existing methods are subject to substantial limitations, even for the internal validation design, and all the more so for the external validation and replicate measures designs. There is a need for a new method that overcomes these limitations. In particular, there is is a need for a convenient method for all three study designs that can handle general measurement error structures, both functional and structural models, and time-dependent covariates.

The aim of this paper is to present such a method for the case where the error-prone covariates are discrete. The misclassification is allowed to be arbitrarily complex. The error-free covariates are allowed to be either discrete or continuous. The time-dependent covariates are required to satisfy a certain "localized error" condition which we describe later. We present basic asymptotic properties of the method and examine its finite-sample performance in a simulation study.

In the case where the classification probabilities are estimated from replicate measurements data, it is necessary in this estimation process to regard the true covariate value as a random variable, as in a structural model (see Sec. 5 below). But when *external* replicate data are used, the marginal distribution of the true covariate need not be the same in the replicate sample as in the main study; we need only portability of the conditional distribution of the observed covariate given the true covariate.

Our proposed method follows the corrected score approach. Nakamura [14, 15] described the basic idea behind the approach. He then developed it some detail under the

independent additive error model. However, as noted above, this error model is not appropriate for discrete covariates. Accordingly, we build instead on the work of Akazawa, Kinukawa, and Nakamura [16], which dealt with logistic regression with discrete covariates subject to misclassification. We extend the work of Azakawa, Kinukawa, and Nakamura in a number of directions. First, we extend their approach from logistic regression to the Cox survival regression model; this extension involves a certain amount of technical development. Second, we allow for the case where, in addition to the error-prone discrete covariates, there may be a large number of other covariates measured without error; these other covariates can be either discrete or continuous. Finally, while Akazawa, Kinukawa, and Nakamura assume the classification probabilities to be known, we allow them to be estimated, and we derive corrections to the estimated covariance matrix of the parameter estimates that reflect the error in estimating the classification probabilities. In the absence of misclassification, our method reduces to the classical Cox partial likelihood method.

The plan of the paper is as follows. Section 2 reviews the corrected score technique in a general likelihood setting. Section 3 describes our extension of the technique to the Cox [4] survival regression model. In Sections 2 and 3 the misclassification probabilities are assumed known. Section 4 discusses the case where the misclassification probabilities are estimated. Section 5 presents a simulation study of the method in the case of a single binary error-prone covariate. In the simulations, we consider both the case where the misclassification probabilities are known and the case where these probabilities are estimated from external replicate measurement data. Section 6 presents an application to data from the Nurses Health Study on the relationship between dietary calcium intake and distal colon cancer [17]. Section 7 provides a summary and discussion.

2. REVIEW OF THE CORRECTED SCORE TECHNIQUE

We review here the corrected score technique of Nakamura [14, 15] and Akazawa, Kinukawa, and Nakamura [16] in a general likelihood setting. The setup is as follows.

We have data on n individuals, which we assume here are independent, though extensions to correlated data are possible. Associated with each individual i is a response variable T_i and a p-vector of covariates $\mathbf{X}_i$. The conditional density or mass function of T_i given $\mathbf{X}_i$ is denoted $f(t|\mathbf{X}_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a q-vector of unknown parameters, which includes regression coefficients and auxiliary parameters such as the variance of the response variable (given $\mathbf{X}$). We have in mind mainly generalized linear models such as linear, logistic, and Poisson regression, but we present the theory in a general way. We denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$. Extending Akazawa et al. [16], we partition the vector $\mathbf{X}_i$ into $\mathbf{W}_i$ and $\mathbf{Z}_i$, where $\mathbf{W}_i$ is a p_1-vector of error-prone covariates and $\mathbf{Z}_i$ is a p_2-vector of error-free covariates. We denote the observed value of $\mathbf{W}_i$ by $\tilde{\mathbf{W}}_i$. The vector $\mathbf{W}_i$ is assumed to be discrete, with its possible values (each one a p_1-vector) denoted by $\mathbf{w}_1, \ldots, \mathbf{w}_K$. The range of values of $\tilde{\mathbf{W}}_i$ is assumed to be the same as that for $\mathbf{W}_i$. We denote by $k(i)$ the value of k such that $\tilde{\mathbf{W}}_i = \mathbf{w}_k$. The vector $\mathbf{Z}_i$ of error-free covariates is allowed to be either discrete or continuous. The case where the model $f(t|\mathbf{X}_i, \boldsymbol{\theta})$ involves interaction terms between the error-prone and error-free covariates can be accommodated with suitable minor notational changes.

We denote $A_{kl}^{(i)} = \Pr(\tilde{\mathbf{W}}_i = \mathbf{w}_l | \mathbf{W}_i = \mathbf{w}_k, \mathbf{Z}_i, T_i)$, which defines a square matrix $\mathbf{A}^{(i)}$ of classification probabilities. The setup here differs from that of Zucker and Spiegelman [10]. There, we worked with the conditional distribution of the true covariate vector given the observed covariate vector, thus following the structural model approach. Moreover, in [10], *all* covariates were assumed to be discrete, and the classification rate matrix involved probabilities relating to the entire covariate vector. Here, we work with the conditional distribution of the observed covariate vector given the true covariate vector–a functional model approach–and only the error-prone covariates are assumed to be discrete. In addition, the classification rate matrix $\mathbf{A}^{(i)}$ here involves only probabilities relating to the error-prone covariates. Also, in [10], we worked only with overall classification probabilities for the entire population, but here we allow the classification probabilities to depend on individual-specific factors, including the error-free covariates. For now, we assume that $\mathbf{A}^{(i)}$ is known. In Section 4, we will consider

the case where $\mathbf{A}^{(i)}$ is estimated.

We denote by $\mathbf{B}^{(i)}$ the matrix inverse of $\mathbf{A}^{(i)}$, which is assumed to exist. This assumption will generally be satisfied provided the misclassification is not too extreme; cf. [10, Appendix A.1].

Define $\mathbf{u}(t, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}) = [\partial/\partial\boldsymbol{\theta}] \log f(t|\mathbf{w}, \mathbf{z}, \boldsymbol{\theta})$ and $\mathbf{u}_i(\boldsymbol{\theta}) = \mathbf{u}(T_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta})$. The classical normalized likelihood score function, for the case of no covariate error, is then given by $\mathbf{U}(\boldsymbol{\theta}) = n^{-1} \sum_i \mathbf{u}_i(\boldsymbol{\theta})$, and the maximum likelihood estimate (MLE) is obtained by solving the equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$. Under classical conditions, $E_{\boldsymbol{\theta}_0}[\mathbf{U}(\boldsymbol{\theta}_0)] = \mathbf{0}$ and the MLE is consistent and asymptotically normal.

The idea of the corrected score approach is to find a function $\mathbf{u}^*(t, \tilde{\mathbf{w}}, \mathbf{z}, \boldsymbol{\theta})$ such that

$$E[\mathbf{u}^*(T_i, \tilde{\mathbf{W}}_i, \mathbf{Z}_i, \boldsymbol{\theta})|\mathbf{W}_i, \mathbf{Z}_i, T_i] = \mathbf{u}(T_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}). \tag{1}$$

Then, with $\mathbf{u}_i^*(\boldsymbol{\theta}) = \mathbf{u}^*(T_i, \tilde{\mathbf{W}}_i, \mathbf{Z}_i, \boldsymbol{\theta})$, we use the modified likelihood score function $\mathbf{U}^*(\boldsymbol{\theta}) = n^{-1} \sum_i \mathbf{u}_i^*(\boldsymbol{\theta})$ in place of $\mathbf{U}(\boldsymbol{\theta})$ as the basis for estimation. The estimation equation thus becomes $\mathbf{U}^*(\boldsymbol{\theta}) = \mathbf{0}$.

In the case of discrete error-prone covariates, as shown by Akazawa et al., a function $\mathbf{u}^*$ satisfying (1) may be constructed by a simple device: we define

$$\mathbf{u}_i^*(\boldsymbol{\theta}) = \sum_{l=1}^{K} B_{k(i)l}^{(i)} \mathbf{u}(T_i, \mathbf{w}_l, \mathbf{Z}_i, \boldsymbol{\theta}), \tag{2}$$

where $k(i)$ is as defined previously and $B_{kl}^{(i)}$ is the (k, l) element of the matrix $\mathbf{B}^{(i)}$. We then have

$$\begin{aligned} E[\mathbf{u}_i^*(\boldsymbol{\theta})|\mathbf{W}_i = \mathbf{w}_m, \mathbf{Z}_i, T_i] &= \sum_{k=1}^{K} A_{mk}^{(i)} \sum_{l=1}^{K} B_{kl}^{(i)} \mathbf{u}(T_i, \mathbf{w}_l, \mathbf{Z}_i, \boldsymbol{\theta}) \\ &= \sum_{l=1}^{K} \left(\sum_{k=1}^{K} A_{mk}^{(i)} B_{kl}^{(i)} \right) \mathbf{u}(T_i, \mathbf{w}_l, \mathbf{Z}_i, \boldsymbol{\theta}) \\ &= \mathbf{u}(T_i, \mathbf{w}_m, \mathbf{Z}_i, \boldsymbol{\theta}), \end{aligned} \tag{3}$$

so that (1) is indeed satisfied. In particular, we have $E_{\boldsymbol{\theta}_0}[\mathbf{U}^*(\boldsymbol{\theta}_0)] = \mathbf{0}$.

Next, define the matrix $\mathbf{J}_i(\boldsymbol{\theta})$ by $J_{i,rs}(\boldsymbol{\theta}) = (\partial/\partial\theta_s)u_{i,r}(\theta)$ and let $\mathbf{J}_i^*(\boldsymbol{\theta})$ be defined correspondingly with $\mathbf{u}_i^*$ in place of $\mathbf{u}_i$. Then, analogously to (3), we have $E[J_{i,rs}^*(\boldsymbol{\theta})|\mathbf{W}_i, \mathbf{Z}_i, T_i] = J_{i,rs}(\boldsymbol{\theta})$.

Under the typical conditions assumed in generalized estimation equations (GEE) theory [18], the estimator will be consistent and asymptotically normal. The limiting covariance matrix $\mathbf{V}$ of $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be estimated using the sandwich estimator

$$\hat{\mathbf{V}} = \mathbf{D}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{H}(\hat{\boldsymbol{\theta}})\mathbf{D}(\hat{\boldsymbol{\theta}})^{-1}, \tag{4}$$

where

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{u}_i^*(\boldsymbol{\theta})\mathbf{u}_i^*(\boldsymbol{\theta})^T \tag{5}$$

and $\mathbf{D}(\boldsymbol{\theta}) = -n^{-1}\sum_i \mathbf{J}_i^*(\boldsymbol{\theta})$.

In the internal validation design, for each individual i in the internal validation sample we can carry out the estimation of $\boldsymbol{\theta}$ with $\tilde{\mathbf{W}}_i$ replaced by $\mathbf{W}_i$ and $\mathbf{A}^{(i)}$ replaced by the identity matrix. Alternatively, we can employ the hybrid scheme of Zucker and Spiegelman [10, Sec. 5], where a separate estimator of $\boldsymbol{\theta}$ is computed for the validation sample and for the main study excluding the validation sample, and the two estimators are then combined. The hybrid scheme is likely to be more efficient when the validation sample is sizable.

The case where there are replicate measurements $\tilde{\mathbf{W}}_{ij}$ of $\tilde{\mathbf{W}}$ on the individuals in the main study can be handled in various ways. A simple approach is to redefine the quantity $\mathbf{u}_i^*(\boldsymbol{\theta})$ given in (2) by replacing $B_{k(i)l}^{(i)}$ with the mean of $B_{k(i,j)l}^{(i)}$ over the replicates for individual i, with $k(i,j)$ defined as the value of k such that $\tilde{\mathbf{W}}_{ij} = \mathbf{w}_k$. The development then proceeds as before.

3. APPLICATION TO THE COX SURVIVAL MODEL

3.1. *The Setup*

We now show how to apply the foregoing corrected score approach to the Cox [4] proportional hazards regression model. Nakamura [15] described a related corrected score technique for the case of a normally-distributed covariate with independent additive measurement error. We assume a standard survival analysis setup. The survival time is denoted by T_i°. As usual, observation is subject to right censoring with censoring time C_i, and the observed survival data consist of the observed follow-up time $T_i = \min(T_i^\circ, C_i)$ and the event indicator $\delta_i = I(T_i^\circ \leq C_i)$. We let $Y_i(t) = I(T_i \geq t)$ denote the at-risk indicator. We assume the failure process and the censoring process are conditionally independent given the covariate process in the sense described by Kalbfleisch and Prentice [19, Sec. 5.3.2]. The covariate processes are assumed to be left continuous with right limits. Left truncation can be handled by setting $Y_i(t)$ to zero until the time at which individual i comes under observation.

The covariate structure is as described in the preceding section, except that the covariates are allowed to be time-dependent, so that we write $k(i,t)$ and $\mathbf{Z}_i(t)$. We assume that the measurement error process is "localized" in the sense that it depends only on the current true covariate value. More precisely, the assumption is that, conditional on the value of $\mathbf{X}_i(t)$, the value of $\tilde{\mathbf{W}}_i(t)$ is independent of the survival and censoring processes and of the values of $\mathbf{X}_i(s)$ for $s \neq t$. This assumption is plausible in many circumstances, such as situations in which the main source of error is technical or laboratory error, or reading/coding error, as with diagnostic X-rays and dietary intake assessments. The assumption will not be directly satisfied for covariates that represent cumulative exposure, though it may be possible to adapt our approach to cumulative exposure variables by working with the successive increments in observed exposure. For time-independent covariates, the assumption reduces to an assumption that the measurement error process is independent of the survival and censoring processes.

We note that with no change in the theory, the classification probabilities $A_{kl}^{(i)}$ can be allowed to depend upon t, but we shall not indicate this dependence explicitly in the notation. With this extension, we can account for improvements in measurement

techniques over time. In addition, if internal validation data are available, this extension allows us to dispense with the localized error assumption. The assumption can be avoided by using classification rate estimates based on the internal validation sample units that are still at risk at each given point in time. Also, as before, interaction terms between the error-prone and error-free covariates can be accommodated with suitable minor notational changes.

In the proportional hazards model, the hazard function is taken to be of the form

$$\lambda(t|\mathbf{X}(t)) = \lambda_0(t)\psi(\mathbf{X}(t);\boldsymbol{\beta}), \tag{6}$$

with $\lambda_0(t)$ being a baseline hazard function of unspecified form. The function $\psi(\mathbf{x};\boldsymbol{\beta})$, which involves a p-vector $\boldsymbol{\beta}$ of unknown regression parameters which are to be estimated, represents the relative risk for an individual with covariate vector $\mathbf{x}$. The classical Cox [4] model assumes $\psi(\mathbf{x};\boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T\mathbf{x}}$. In line with Thomas [20] and Breslow and Day [21, Sec. 5.1(c)], we allow a general relative risk function $\psi(\mathbf{x};\boldsymbol{\beta})$ which is assumed to be positive in a neighborhood of the true $\boldsymbol{\beta}$ for all $\mathbf{x}$ and to be twice differentiable with respect to the components of $\boldsymbol{\beta}$. We assume further that $\psi(\mathbf{x};\mathbf{0}) = 1$, which simply means that $\boldsymbol{\beta} = \mathbf{0}$ corresponds to no covariate effect. In many applications, it will be desirable to take $\psi(\mathbf{x};\boldsymbol{\beta})$ to be a function that is monotone in each component of $\mathbf{x}$ for all $\boldsymbol{\beta}$. We let $\boldsymbol{\beta}_0$ denote the true value of $\boldsymbol{\beta}$.

3.2. *The Method*

We now describe the method. Let $\psi_r'(\mathbf{x};\boldsymbol{\beta})$ denote the partial derivative of $\psi(\mathbf{x};\boldsymbol{\beta})$ with respect to β_r and define $\xi_r(\mathbf{x};\boldsymbol{\beta}) = \psi_r'(\mathbf{x};\boldsymbol{\beta})/\psi(\mathbf{x};\boldsymbol{\beta})$. Then the classical Cox [4, 22] partial likelihood score function in the case with no measurement error is given by

$$U_r(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\xi_r(\mathbf{X}_i(T_i);\boldsymbol{\beta}) - \frac{e_{1r}(T_i)}{e_0(T_i)}\right), \tag{7}$$

where

$$\begin{aligned} e_0(t) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\psi(\mathbf{X}_i(t);\boldsymbol{\beta}), \\ e_{1r}(t) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\psi_r'(\mathbf{X}_i(t);\boldsymbol{\beta}). \end{aligned}$$

Now define

$$\begin{aligned}
\psi_i^*(t,\boldsymbol{\beta}) &= \sum_{l=1}^{K} B_{k(i,t)l}^{(i)} \psi(\mathbf{w}_l, \mathbf{Z}_i(t); \boldsymbol{\beta}), \\
\eta_{ir}(t,\boldsymbol{\beta}) &= \sum_{l=1}^{K} B_{k(i,t)l}^{(i)} \psi_r'(\mathbf{w}_l, \mathbf{Z}_i(t); \boldsymbol{\beta}), \\
\xi_{ir}^*(t,\boldsymbol{\beta}) &= \sum_{l=1}^{K} B_{k(i,t)l}^{(i)} \xi_r(\mathbf{w}_l, \mathbf{Z}_i(t); \boldsymbol{\beta}), \\
e_0^*(t) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\psi_j^*(t,\boldsymbol{\beta}), \\
e_{1r}^*(t) &= \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\eta_{jr}(t,\boldsymbol{\beta}).
\end{aligned}$$

Then our proposed corrected score function is the following obvious analogue of (7):

$$U_r^*(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \left(\xi_{ir}^*(T_i,\boldsymbol{\beta}) - \frac{e_{1r}^*(T_i)}{e_0^*(T_i)} \right). \tag{8}$$

As before, the proposed corrected score estimator is the solution to $\mathbf{U}^*(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{U}^*$ denotes the vector whose components are U_r^*.

We have

$$\begin{aligned}
E[Y_i(t)\psi_i^*(t,\boldsymbol{\beta})|\mathbf{X}_i(t)] &= E[Y_i(t)E[\psi_i^*(t,\boldsymbol{\beta})|\mathbf{X}_i(t), Y_i(t)]|\mathbf{X}_i(t)] \\
&= E[Y_i(t)E[\psi_i^*(t,\boldsymbol{\beta})|\mathbf{X}_i(t)]|\mathbf{X}_i(t)] \\
&= E[Y_i(t)\psi(\mathbf{X}_i(t);\boldsymbol{\beta})|\mathbf{X}_i(t)], \qquad (9)
\end{aligned}$$

where the second equality follows from the localized error assumption and the third follows from the argument used with (3). Similarly,

$$E[Y_i(t)\eta_{ir}^*(t,\boldsymbol{\beta})|\mathbf{X}_i(t)] = E[Y_i(t)\psi_r'(\mathbf{X}_i(t),\boldsymbol{\beta})|\mathbf{X}_i(t)], \tag{10}$$

$$E[Y_i(t)\xi_{ir}^*(t,\boldsymbol{\beta})|\mathbf{X}_i(t)] = E[Y_i(t)\xi_r(\mathbf{X}_i(t),\boldsymbol{\beta})|\mathbf{X}_i(t)]. \tag{11}$$

Thus, referring to the quantity in parentheses in (8), the first term and the numerator and denominator of the second term all have the correct expectation. It follows that $\mathbf{U}^*(\boldsymbol{\beta})$ is an asymptotically unbiased score function.

Accordingly, under standard conditions like those of Andersen and Gill [23] and of Prentice and Self [24], our corrected score estimator will be consistent and asymptotically normal. The Appendix presents an outline of the asymptotic arguments. See Huang and Wang [25] for a related discussion in a similar context. The asymptotic covariance matrix of $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ may be estimated by the sandwich formula

$$\hat{\mathbf{V}} = \mathbf{D}(\hat{\boldsymbol{\beta}})^{-1}\mathbf{H}(\hat{\boldsymbol{\beta}})\mathbf{D}(\hat{\boldsymbol{\beta}})^{-1}. \tag{12}$$

Here $\mathbf{D}(\boldsymbol{\beta})$ is -1 times the matrix of derivatives of $\mathbf{U}^*(\boldsymbol{\beta})$ with respect to the components of $\boldsymbol{\beta}$ and $\mathbf{H}(\boldsymbol{\beta})$ is an empirical estimate of the covariance matrix of $n^{\frac{1}{2}}\,\mathbf{U}^*(\boldsymbol{\beta})$. To define these matrices, some additional notation is needed. We define

$$\hat{\Upsilon}_{ir}(\boldsymbol{\beta}) = \delta_i \left[\xi^*_{ir}(T_i, \boldsymbol{\beta}) - \frac{e^*_{1r}(T_i)}{e^*_0(T_i)}\right] - \frac{1}{n}\sum_{j:T_j \le T_i} \delta_j \left[\frac{\eta_{ir}(T_j, \boldsymbol{\beta})}{e^*_0(T_j)} - \frac{e^*_{1r}(T_j)}{e^*_0(T_j)}\frac{\psi^*_i(T_j, \boldsymbol{\beta})}{e^*_0(T_j)}\right]. \tag{13}$$

In this definition, the first term tends to be the dominant term, especially if the event is rare. We further define $\psi''_{rs}(\mathbf{x}; \boldsymbol{\beta}) = (\partial^2/\partial\beta_r\, \partial\beta_s)\psi(\mathbf{x}; \boldsymbol{\beta})$, and

$$\phi_{rs}(\mathbf{x}; \boldsymbol{\beta}) = \frac{\psi''_{rs}(\mathbf{x}; \boldsymbol{\beta})}{\psi(\mathbf{x}; \boldsymbol{\beta})} - \frac{\psi'_r(\mathbf{x}; \boldsymbol{\beta})}{\psi(\mathbf{x}; \boldsymbol{\beta})}\frac{\psi'_s(\mathbf{x}; \boldsymbol{\beta})}{\psi(\mathbf{x}; \boldsymbol{\beta})}.$$

Finally, we define

$$\begin{aligned}
\phi^*_{irs}(t, \boldsymbol{\beta}) &= \sum_{l=1}^{K} B^{(i)}_{k(i,t)l}\phi_{rs}(\mathbf{w}_l, \mathbf{Z}_i(t); \boldsymbol{\beta}), \\
e^*_{2rs}(t) &= \sum_{j=1}^{n} Y_j(t) \sum_{l=1}^{K} B^{(i)}_{k(j,t)l}\psi''_{rs}(\mathbf{w}_l, \mathbf{Z}_j(t); \boldsymbol{\beta}).
\end{aligned}$$

With these definitions we have

$$H_{rs}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \hat{\Upsilon}_{ir}(\boldsymbol{\beta})\hat{\Upsilon}_{is}(\boldsymbol{\beta}), \tag{14}$$

$$D_{rs}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \left[-\phi^*_{irs}(T_i, \boldsymbol{\beta}) + \frac{e^*_{2rs}(T_i)}{e^*_0(T_i)} - \left(\frac{e^*_{1r}(T_i)}{e^*_0(T_i)}\right)\left(\frac{e^*_{1s}(T_i)}{e^*_0(T_i)}\right)\right]. \tag{15}$$

The expression for $D_{rs}(\boldsymbol{\beta})$ is derived by straightforward differentiation. The derivation of the expression for $H_{rs}(\boldsymbol{\beta})$ is given in the Appendix.

In the case of the classical Cox model with $\psi(\mathbf{x}; \boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T\mathbf{x}}$, we have $\psi'_r(\mathbf{x}; \boldsymbol{\beta}) = x_r e^{\boldsymbol{\beta}^T\mathbf{x}}$, $\xi_r(\mathbf{x}; \boldsymbol{\beta}) = x_r$, $\psi''_{rs}(\mathbf{x}; \boldsymbol{\beta}) = x_r x_s e^{\boldsymbol{\beta}^T x}$, and $\phi_{rs}(\mathbf{x}; \boldsymbol{\beta}) = 0$. We note again that,

for the internal validation design, the available true $\mathbf{W}$ values can be used in the estimation of $\boldsymbol{\beta}$ by replacing $\tilde{\mathbf{W}}_i$ with $\mathbf{W}_i$ and $\mathbf{A}^{(i)}$ by the identity matrix when individual i is in the internal validation sample. Alternatively, the hybrid scheme of Zucker and Spiegelman [10, Sec. 5] can be used. Also, the case where there are replicate measurements $\tilde{\mathbf{W}}_{ij}$ of $\tilde{\mathbf{W}}$ on the individuals in the main study can be handled as described at the end of the preceding section.

3.3. *Estimation of the Cumulative Hazard*

The cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ can be estimated using the Breslow-type estimator

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \frac{\delta_i I(T_i \le t)}{e_0^*(T_i, \hat{\boldsymbol{\beta}})} \, . \tag{16}$$

From the argument given at the end of the Appendix, we find that $n^{\frac{1}{2}}(\hat{\Lambda}_0(t) - \Lambda_0(t))$, for a given t, is asymptotically mean-zero normal with variance that can be estimated by

$$\widehat{\mathrm{Var}}(\hat{\Lambda}_0(t)) = \frac{1}{n}\sum_{i=1}^{n} \hat{\Upsilon}_i^*(\hat{\boldsymbol{\beta}})^2 \, , \tag{17}$$

where

$$\hat{\Upsilon}_i^*(\boldsymbol{\beta}) = \sum_{r=1}^{p} \hat{c}_r(\boldsymbol{\beta})\hat{\Upsilon}_{ir}(\boldsymbol{\beta}) + \frac{\delta_i I(T_i \le t)}{e_0^*(T_i, \boldsymbol{\beta})} - \frac{1}{n}\sum_{j=1}^{n} \delta_j I(T_j \le \min\{T_i, t\}) \frac{\psi_i^*(T_j, \boldsymbol{\beta})}{e_0^*(T_j, \boldsymbol{\beta})} \, .$$

In the above expression, $\hat{\mathbf{c}}(\boldsymbol{\beta}) = -\mathbf{D}(\boldsymbol{\beta})^{-1}\hat{\mathbf{a}}(\boldsymbol{\beta})$, where

$$\hat{a}_r(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i I(T_i \le t) \frac{e_{1r}^*(T_i, \boldsymbol{\beta})}{e_0^*(T_i, \boldsymbol{\beta})^2} \, .$$

4. ESTIMATED CLASSIFICATION PROBABILITIES

We now indicate the changes needed to handle the case where the $A_{kl}^{(i)}$ are estimated. The relevant estimates may be obtained in several ways. In some cases, estimates are obtained from an external validation study, that is, a separate study

with measurements of both $\mathbf{W}_i$ and $\tilde{\mathbf{W}}_i$. Alternately, an internal validation design is used, with some individuals in the main survival study having measurements on both $\mathbf{W}_i$ and $\tilde{\mathbf{W}}_i$. Another possibility is to base the estimates on internal or external replicate measures data; we discuss this in more detail at the end of the section. The theory developed in this section represents a step beyond the work of Akazawa et al. [16], who considered only the case where the classification probabilities are known. This theory is applied in the example presented in Section 7.

The main issue is how to adjust the covariance matrix of the estimates to account for the estimation error in the $A_{kl}^{(i)}$. Following Zucker and Spiegelman [10], we express $\mathbf{A}^{(i)}$ as $\mathbf{A}^{(i)}(\boldsymbol{\omega})$ for some q'-vector of parameters $\boldsymbol{\omega}$. The nature of the function $\mathbf{A}^{(i)}(\boldsymbol{\omega})$ is dictated by the measurement error model employed. As an illustration, consider the simplest case: a single binary covariate with a common classification matrix $\mathbf{A}(\boldsymbol{\omega})$ for all individuals. In this case, with the false positive and false negative rates allowed to be different, $\mathbf{A}(\boldsymbol{\omega})$ takes the following form (where we assume the sum of the off-diagonal elements is less than 1):

$$\mathbf{A}(\boldsymbol{\omega}) = \begin{bmatrix} \omega_1 & (1-\omega_1) \\ (1-\omega_2) & \omega_2 \end{bmatrix}. \tag{18}$$

We presume that the parameter vector $\boldsymbol{\omega}$ is estimated from a study of one of the types described above, with m independent units. We presume further that the study yields an estimator $\hat{\boldsymbol{\omega}}$ having an approximate normal distribution with mean $\boldsymbol{\omega}$ and covariance matrix $m^{-1}\boldsymbol{\Gamma}$, along with an estimator $\hat{\boldsymbol{\Gamma}}$ of the matrix $\boldsymbol{\Gamma}$. This setup is a typical one in practical applications. For example, for the case of a single 0–1 binary covariate with internal or external validation data, the estimates of $\omega_k = \Pr(\tilde{W} = k-1|W = k-1), k = 1, 2$, are given by the obvious sample proportions, and $\boldsymbol{\Gamma}$ is a 2×2 diagonal matrix with $\Gamma_{kk} = \omega_k(1-\omega_k)/\vartheta_k$, where ϑ_k is the fraction of individuals with $W = k-1$ in the validation study. The procedure for a replicate measures study is discussed at the end of this section. For the asymptotics we assume that m and n are of the same order of magnitude, i.e., $m/n \to \zeta$ for some constant ζ as $n \to \infty$.

Otherwise, the error in $\mathbf{A}^{(i)}(\boldsymbol{\omega})$ will either be dominated by or will dominate the error in $\hat{\boldsymbol{\beta}}$ due to the variation in the survival data. Typically ζ will be between 0 and 1.

Let us now write the corrected score function as $\mathbf{U}^*(\boldsymbol{\theta}, \boldsymbol{\omega})$ to indicate explicitly the dependence on $\boldsymbol{\omega}$, where we mean here to cover the setups of both Sections 2 and 3, with the $\boldsymbol{\theta}$ of Section 3 being $\boldsymbol{\beta}$. Also, let us denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$ (as before) and the true value of $\boldsymbol{\omega}$ by $\boldsymbol{\omega}_0$. Since we are now estimating $\boldsymbol{\omega}_0$ by $\hat{\boldsymbol{\omega}}$, our estimating equation for $\boldsymbol{\theta}$ is now $\mathbf{U}^*(\boldsymbol{\theta}, \hat{\boldsymbol{\omega}}) = \mathbf{0}$. Using Taylor's theorem, we can write

$$\mathbf{0} = \mathbf{U}^*(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}}) \doteq \mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0) - \mathbf{D}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \dot{\mathbf{U}}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0),$$

where $-D_{rs}$ is the partial derivative of $U_r^*(\boldsymbol{\theta}, \boldsymbol{\omega})$ with respect to θ_s evaluated at $\boldsymbol{\omega}_0$ (as presented in Section 2 or 3, as appropriate), and $\dot{\mathbf{U}}(\boldsymbol{\theta}, \boldsymbol{\omega})$ is a matrix whose (r, ν) element is the partial derivative of $U_r(\boldsymbol{\theta}, \boldsymbol{\omega})$ with respect to ω_ν. Hence we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \doteq \mathbf{D}(\boldsymbol{\theta}_0)^{-1} \left[\mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0) + \dot{\mathbf{U}}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) \right]. \tag{19}$$

If $\boldsymbol{\omega}$ is estimated from an external data, then $\mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)$ and $\hat{\boldsymbol{\omega}}$ are obviously independent. When $\boldsymbol{\omega}$ is estimated from an internal validation sample, the following argument can be given to show that these two quantities are asymptotically independent. The contribution to $\mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)$ made by each individual in the study has exactly (Section 2) or asymptotically (Section 3) expectation zero conditional on the true covariate values. This is true in particular of the individuals in the internal validation sample. It hence follows from an iterated expectation argument that $\mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)$ and $\hat{\boldsymbol{\omega}}$ are asymptotically uncorrelated (cf. [8, Appendix A]). Since we have asymptotic normality as well, this implies asymptotic independence. As a result, by Slutsky's theorem, the two terms in brackets in (19) are asymptotically independent, since $\dot{\mathbf{U}}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)$ converges to a deterministic limit.

It therefore follows that, when external data or an internal validation sample is used to estimate $\boldsymbol{\omega}$, the necessary covariance adjustment may be accomplished by replacing the matrix $\mathbf{H}(\boldsymbol{\theta})$ with the following corrected version:

$$\mathbf{H}^{(\mathrm{corr})}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \mathbf{H}(\boldsymbol{\theta}, \boldsymbol{\omega}) + \zeta^{-1} \dot{\mathbf{U}}^*(\boldsymbol{\theta}, \boldsymbol{\omega}) \hat{\boldsymbol{\Gamma}} \dot{\mathbf{U}}^*(\boldsymbol{\theta}, \boldsymbol{\omega})^T. \tag{20}$$

To present the formulas for $\dot{U}^*_{r\nu}(\boldsymbol{\theta})$, we define $\dot{\mathbf{A}}^{(i)}_\nu$ and $\dot{\mathbf{B}}^{(i)}_\nu$ to be, respectively, the partial derivative of the matrices $\mathbf{A}^{(i)}$ and $\mathbf{B}^{(i)}$ with respect to ω_ν. By the rule for differentiating an inverse matrix, we have $\dot{\mathbf{B}}^{(i)}_\nu = -\mathbf{B}^{(i)}\dot{\mathbf{A}}^{(i)}_\nu\mathbf{B}^{(i)}$. For the setup of Section 2, we then have simply

$$\dot{U}_{r\nu}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\dot{B}^{(i)}_{k(i)l:\nu}u_r(T_i, \mathbf{w}_l, \mathbf{Z}_i, \boldsymbol{\theta}). \tag{21}$$

For the setup in Section 3, we have

$$\dot{U}_{r\nu}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\dot{\xi}_{ir:\nu}(T_i, \boldsymbol{\beta}) - \frac{\dot{e}^*_{1r:\nu}(T_i)}{e^*_0(T_i)} + \left(\frac{e^*_{1r}(T_i)}{e^*_0(T_i)}\right)\left(\frac{\dot{e}^*_{0:\nu}(T_i)}{e^*_0(T_i)}\right)\right], \tag{22}$$

where $\dot{\xi}^*_{ir:\nu}, \dot{e}^*_{0:\nu}$, and $\dot{e}^*_{1r:\nu}$ are defined analogously to ξ^*_{ir}, e^*_0, and e^*_{1r} with $B^{(i)}_{k(i)l}$ replaced by $\dot{B}^{(i)}_{k(i)l:\nu}$.

When $\boldsymbol{\omega}$ is estimated from an internal replicate measures sample, $\mathbf{U}^*(\boldsymbol{\theta}_0, \boldsymbol{\omega}_0)$ and $\hat{\boldsymbol{\omega}}$ are no longer asymptotically independent, and so we must work out the covariance between them. We consider, for concreteness, an i.i.d. setup where $\boldsymbol{\omega}$ is estimated by maximum likelihood. Let R_i denote the number of replicates on individual i, and let $g_i(\boldsymbol{\omega})$ denote the log likelihood function for $(\tilde{\mathbf{W}}_{i1}, \ldots, \tilde{\mathbf{W}}_{iR_i})$. The overall normalized log likelihood is then $g(\boldsymbol{\omega}) = m^{-1}\sum_{i\in\mathcal{R}} g_i(\boldsymbol{\omega})$, where $\mathcal{R}$ denotes the set of individuals in the internal replicate measures sample. Let $\mathbf{g}'(\boldsymbol{\omega})$ and $\mathbf{g}''(\boldsymbol{\omega})$ denote the gradient vector and Hessian matrix, respectively, of $g(\boldsymbol{\omega})$, and let $\mathbf{g}'_i(\boldsymbol{\omega})$ denote the gradient of $\mathbf{g}_i(\boldsymbol{\omega})$. We can then express $\hat{\boldsymbol{\omega}}$ in terms of the classic asymptotic approximation $\hat{\boldsymbol{\omega}} \doteq -\mathbf{g}''(\boldsymbol{\omega}_0)^{-1}\mathbf{g}'(\boldsymbol{\omega}_0)$. For the setting of Section 2, the limiting value of

$$\boldsymbol{\Phi} = \mathrm{Cov}\left(\left[\frac{1}{\sqrt{m}}\sum_{i\in\mathcal{R}}\mathbf{u}^*_i(\boldsymbol{\omega})\right], \sqrt{m}\,\mathbf{g}'(\boldsymbol{\omega}_0)\right)$$

can be estimated empirically by

$$\hat{\boldsymbol{\Phi}} = \frac{1}{m}\sum_{i\in\mathcal{R}}\mathbf{u}^*_i(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}})\mathbf{g}'_i(\hat{\boldsymbol{\omega}})^T. \tag{23}$$

For the setting of Section 3, the relevant expression for $\hat{\boldsymbol{\Phi}}$ is given by replacing $\mathbf{u}^*_i(\hat{\boldsymbol{\theta}})$ with the vector comprising the quantities $\hat{\Upsilon}_{ir}(\hat{\boldsymbol{\beta}})$. The appropriate corrected version of $\hat{\mathbf{H}}$ is then

$$\hat{\mathbf{H}}^{(\mathrm{corr})} = \hat{\mathbf{H}} + \zeta^{-1}\dot{\mathbf{U}}^*(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}})\hat{\boldsymbol{\Gamma}}\dot{\mathbf{U}}^*(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}})^T - \hat{\boldsymbol{\Phi}}\mathbf{g}''(\hat{\boldsymbol{\omega}})^{-1}\dot{\mathbf{U}}^*(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}})^T, \tag{24}$$

where for the present setup we have $\hat{\mathbf{\Gamma}} = \mathbf{g}''(\hat{\boldsymbol{\omega}})^{-1}$.

The estimate (17) for $\mathrm{Var}(\hat{\Lambda}_0(t))$ can be corrected as follows. Define $\hat{\mathbf{h}} = \hat{\mathbf{h}}(\boldsymbol{\beta}, \boldsymbol{\omega})$ by

$$\hat{h}_\nu = [\dot{\mathbf{U}}^*(\boldsymbol{\beta}, \boldsymbol{\omega})^T \hat{\mathbf{c}}(\boldsymbol{\beta})]_\nu$$
$$-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_j I(T_j \le \min\{T_i, t\}) s_0(T_j, \boldsymbol{\beta})^{-1} \sum_{l=1}^{K}[\dot{B}_\nu^{(i)}]_{k(i,T_j)l}\psi(\mathbf{w}_l, \mathbf{Z}_i(T_j); \boldsymbol{\beta}).$$

With external data or an internal validation study, it is necessary merely to add to (17) the term $\zeta^{-1}\hat{\mathbf{h}}^T\hat{\mathbf{\Gamma}}\hat{\mathbf{h}}$. For the case of internal replicate data, the additional term is

$$\zeta^{-1}\hat{\mathbf{h}}^T\hat{\mathbf{\Gamma}}\hat{\mathbf{h}} + \left[\frac{1}{m}\sum_{i\in\mathcal{R}}\hat{\Upsilon}_i^*(\hat{\boldsymbol{\beta}})\mathbf{g}_i'(\hat{\boldsymbol{\omega}})\right]^T \mathbf{g}''(\hat{\boldsymbol{\omega}})^{-1}\hat{\mathbf{h}}.$$

We now discuss in more detail the estimation of classification probabilities from a replicate measures study. The simplest case is that of a single binary (0–1) error-prone covariate W, with subject i in the replicate measures study having R_i replicate measurements $\tilde{W}_{ij}$ on the surrogate measure $\tilde{W}$, where the replicates are conditionally i.i.d. given W. This case is relevant to many risk factors of interest in epidemiological studies, such as high blood pressure and high estradiol level, where the risk factor is assessed using direct physiological measurements. In this setting, the subject totals $\tilde{W}_i^{(\mathrm{tot})} = \sum_j \tilde{W}_{ij}$ are sufficient statistics. Denote $\alpha_1 = \Pr(\tilde{W} = 1|W = 0)$ and $\alpha_2 = \Pr(\tilde{W} = 1|W = 1)$. Then, conditional on $W = k - 1$ $(k = 1, 2)$, $\tilde{W}_i^{(\mathrm{tot})}$ has a $\mathrm{Bin}(R_i, \alpha_k)$ distribution. Defining π to be $\Pr(W = 1)$ within the replicate measures sample, the marginal distribution of $\tilde{W}_i^{(\mathrm{tot})}$ is a mixture of the $\mathrm{Bin}(R_i, \alpha_1)$ and $\mathrm{Bin}(R_i, \alpha_2)$ distributions, with respective mixture probabilities $1 - \pi$ and π:

$$\Pr(\tilde{W}_i^{(\mathrm{tot})} = j) = (1-\pi)\begin{pmatrix} R_i \\ j \end{pmatrix}\alpha_1^j(1-\alpha_1)^{R_i-j} + \pi\begin{pmatrix} R_i \\ j \end{pmatrix}\alpha_2^j(1-\alpha_2)^{R_i-j}. \quad (25)$$

The model is identifiable provided that some positive proportion of subjects have $R_i \ge 3$ and the correlation between W and $\tilde{W}$ is positive. The latter condition is equivalent to the condition $\Pr(\tilde{W} = 1|\ W = 0) + \Pr(\tilde{W} = 0|W = 1) < 1$. The likelihood function

can be written down directly from (25), and the parameters π, α_1, and α_2 can then be estimated by maximum likelihood.

In more general settings, the basic ideas are similar, but of course the details are more complex. Several papers have discussed methods for using replicate data to estimate classification probabilities for more general settings, including parsimonious models for polychotomous W and models that allow conditional distribution of the replicates given W to involve some dependence [27]-[30]. The case of correlated replicates is of particular relevance to risk factors that are measured through self-report, such as dietary or physical activity variables.

5. SIMULATION STUDY

To investigate how our method performs, we carried out a simulation study in the setting of a single 0–1 binary covariate. The design of our simulation study was patterned after Zucker and Spiegelman [10, Sec. 6] and Zucker [11, Sec. 3.1]. The assumed study duration was 5 years. The baseline survival distribution was taken to be Weibull, with baseline hazard function $\lambda_0(t) = \alpha\mu(\mu t)^{\alpha-1}$. The power parameter α was taken equal to 5, which is typical of many types of cancer [31], [21, Sec. 6.3]. The scale parameter μ was chosen so as to yield a 25% 5-year cumulative incidence rate for the unexposed population. Censoring was taken to be exponential with a rate of 1% per year. For brevity of presentation, the false positive rate $\Pr(\tilde{W} = 1|W = 0)$ and the false negative rate $\Pr(\tilde{W} = 0|W = 1)$ were taken to be equal to a common classification error rate. A range of values was explored for the prevalence of the risk factor (5%, 25%, 40%), the classification error rate (1%, 5%, 10%, 20%), and the true relative risk (1.5, 2.0). The number of simulation replications was 5,000.

In our first simulation scenario, we assumed that the classification probabilities are known. We took the sample size to be 2,000, leading to approximately 500 events

total. Table 1 shows the results. For comparison, we also present the simulation results given in [10] for the naive Cox partial likelihood estimator ignoring the measurement error and for the parametric log relative risk estimator obtained by maximizing the full Weibull log likelihood under the relevant measurement error model.

The naive Cox estimator was typically badly biased except under 1% misclassification with exposure prevalence of 25% or 40%. By contrast, our method exhibited excellent performance, comparable to that of the fully parametric Weibull estimator. Under an exposure prevalence of 25% or 40%, our method yielded nearly zero bias in the estimated log relative risk, nearly unbiased standard deviation estimates, and accurate confidence interval coverage. With an exposure prevalence of 5%, the performance of all three estimators under consideration was degraded. This finding is not surprising, because the 5% exposure situation presents two difficulties: (1) the expected number of events in the exposed group is only on the order of 25–50, (2) with a misclassification rate of 5% or more, the predictive value of an observed positive exposure is low. The naive Cox estimator was drastically biased. Our estimator and the Weibull estimator were dramatically less biased, but still exhibited some bias. This bias was due in part to outlying values; for both our estimator and the Weibull estimator, the deviation between the median value of the estimates and the true log relative risk was noticeably lower than the deviation between the mean estimated value and the true value. Overall, in terms of mean square error, the performance of our estimator was found to be nearly identical to that of Weibull estimator. In a few cases, our estimator was better; this reflects the fact that, for a given finite sample size, the asymptotically optimal parametric MLE can be outperformed by an alternate estimator. The performance of the method proposed here essentially matches that of the methods of Zucker and Spiegelman [10] and Zucker [11], except that the method here was better for the problematic cases with 5% exposure prevalence. The same pattern is seen under a sample size of 10,000 with a cumulative incidence rate of 5% for the unexposed (results not shown).

In our second simulation scenario, we assumed that the classification probabilities are estimated from an external replicate measures study with 250 subjects and 3 replicate measurements per subject. The procedure for estimating the classification probabilities is described at the end of the preceding section. To get around low cell counts when the misclassification rate is very small (viz. 0.01), we added $\frac{1}{2}$ to all the cell counts. We ran two sets of simulations, one for a sample size of 2,000 (about 500 events total) and the other for a sample size of 1,000 (about 250 events total). Table 2 presents the results.

Our method performed very well when the exposure prevalence was 25% or 40%. Across the board, for both $n = 2,000$ and $n = 1,000$, the bias in the estimated log relative risk was minimal, the standard deviation estimate was on target, and the confidence interval coverage was accurate. With $n = 2,000$, in most cases there was minimal change in the standard deviation of the log relative risk estimate due to the estimation of the misclassification rates, as compared with the standard deviation under known misclassification rates (shown in Table 1). The one exception to this was the case of 20% misclassification, where there was a 5–20% increase in the standard deviation due to the estimation of the misclassification rates. The standard deviations for $n = 1,000$ were greater than those for $n = 2,000$ by about the expected factor of $\sqrt{2}$.

The method performed somewhat less well when the exposure prevalence was 5%. The bias was higher in this situation, in some cases reaching the 10-20% level. Still, this is much better than the bias of the naive Cox estimate (shown for $n = 2,000$ in Table 1). The standard deviation estimates and confidence interval coverage was noticeably inaccurate in some cases. Also, under true misclassification rates of 20%, the misclassification rates could not be successfully estimated in around 10–15% of the simulation replications.

In summary, our method performed generally very well. Good performance was maintained even with estimated misclassification probabilities, except for some prob-

lems when the exposure prevalence rate was very low.

6. EXAMPLE

We illustrate our method on data from the Nurses Health Study concerning the relationship between dietary calcium (Ca) intake and incidence of distal colon cancer [17, Table 4]. The data consist of observations on female nurses whose calcium intake was assessed through a food frequency questionnaire (FFQ) in 1984 and were followed up to May 31, 1996 for distal colon cancer occurrence. Our analysis includes data on 60,575 nurses who reported in 1984 that they had never taken calcium supplements, and focuses on the effect of baseline calcium intake after adjustment for baseline body mass index (BMI) and baseline aspirin use. In line with Wu et al.'s analysis, we use the classical Cox relative risk function $\psi(\boldsymbol{\beta}; \mathbf{x}) = e^{\boldsymbol{\beta}^T \mathbf{x}}$, and, as in Wu et al.'s Table 4, we work with a binary "high Ca" risk factor defined as 1 if the calcium intake was greater than 700 mg/day and 0 otherwise. Note that one glass of milk contains approximately 300 mg of calcium. BMI is expressed in terms of the following categories: <22 kg/m^2, 22 to <25 kg/m^2, 25 to <30 kg/m^2, and 30 kg/m^2 or greater, Aspirin use is coded as yes (1) or no (0). Thus, our model has five explanatory variables, one for the binary risk factor (W), three dummy variables for BMI (Z_1, Z_2, Z_3), and one for aspirin use (Z_4). BMI and aspirin use status are assumed to be measured without error.

It is well known that the FFQ measures dietary intake with some degree of error and more reliable information can be obtained from a diet record (DR) [32, Chapter 6]. We thus take W to be the Ca risk factor indicator based on the DR and $\tilde{W}$ to be the Ca risk factor indicator based on the FFQ. The classification probabilities are estimated using data from the Nurse's Health Study validation study [32, pp. 122–126]. The estimates obtained were $\Pr(\tilde{W} = 0|W = 0) = 0.78$ and $\Pr(\tilde{W} = 1|W = 1) = 0.72$, with corresponding estimated standard errors of 0.042 and 0.046.

Table 3 presents the results of the following analyses: (1) a naive classical Cox regression analysis ignoring measurement error, corresponding to an assumption that

there is no measurement error, (2) our method with **A** assumed known and set according to the foregoing estimated classification probabilities, ignoring the estimation error in these probabilities, and (3) our method with **A** estimated as above with the estimation error in the probabilities taken into account (main study / external validation study design). The last of these analyses makes use of the theory developed in Sec. 4.

The results followed the expected pattern. Adjusting for the misclassification in calcium intake had a marked effect on the estimated relative risk for high calcium intake. Accounting for the error in estimating the classification probabilities increased (modestly) the standard error of the log relative risk estimate. The relative risk estimates for high calcium intake and corresponding 95% confidence intervals obtained in the three analyses were as follows:

Method	Estimate	95% CI
Naive Cox	0.71	[0.51,0.99]
A Known	0.49	[0.24,1.01]
A Estimated	0.49	[0.23,1.04]

The misclassification adjustment had a small effect on the estimated regression coefficients for the BMI dummy variables and essentially no effect on the estimated regression coefficient for aspirin use.

7. SUMMARY AND DISCUSSION

We have considered the Cox [4] proportional hazards model with a set of covariates that includes error-prone discrete covariates along with error-free covariates which may be discrete or continuous. The misclassification in the discrete error-prone covariates is allowed to be of arbitrary form. Building on work of Nakamura [14, 15] and Akazawa, Kinukawa, and Nakamura [16], we have developed an easily-implemented corrected score method for this setting. The method can handle all three major study designs (internal validation design, external validation design, and replicate measures design),

both functional and structural error models, and time-dependent covariates satisfying the "localized error" condition described in Sec. 3. Also, for the internal validation design, the "localized error" condition can be eliminated by using time-dependent classification rate estimates. The method thus represents a significant advance relative to other methods in the literature for this problem. The method performed well in a simulation study, both with misclassification probabilities known and with misclassification probabilities estimated from an external replicate measures study.

In most applications, the new method developed in this paper will be easier to apply than and preferable to our previous method based on weighted transformed Kaplan-Meier curves [10]. Our previous method requires defining strata for every possible configuration of the entire covariate vector (both the error-prone and error-free part). Except when the number of configurations is small, this leads to cumbersome implementation and loss of data for strata having no events. Our current method avoids this problem. Also, our previous method cannot handle continuous error-free covariates, while the current method can. Additionally, our previous method cannot handle time-dependent covariates (nor can the method of Zucker [11]), whereas our current method can handle such covariates if the "localized error" condition applies or internal validation data are available. In some applications, our previous method might be preferred on account of reduced computational burden. Also, our previous method may be more convenient when it is desired to apply a measurement error correction based on published Kaplan-Meier curves for various risk groups as presented in medical and other subject-matter journals. This point is of particular relevance to meta-analysis applications.

This work focuses on the case where the error-prone covariates are discrete. This is admittedly a limitation. However, much of the existing work on the Cox model with covariate error focuses on continuous covariates with independent additive error, and as such does not apply or generalize easily to discrete covariates with misclassification. In many epidemiological studies, the error-prone covariates of interest are in fact discrete.

Thus, the method presented here fills a definite need.

Still, there are cases where it is of interest to investigate continuous error-prone risk factors. In the case of a single error-prone continuous risk factor W, the basic equation (1) for classical likelihood models takes the form

$$u(w) = \int a(\tilde{w}|w)u^*(\tilde{w})d\tilde{w}, \tag{26}$$

where $a(\tilde{w}|w)$ is the conditional density of $\tilde{W}$ given W, the integral is over the entire range of $\tilde{W}$, and the arguments T_i, $\mathbf{Z}_i$, and $\boldsymbol{\theta}$ are suppressed. Analogous equations are obtained for the case of multiple error-prone continuous covariates and for the Cox model setting. The equation (26) is a Fredholm integral equation of the first kind. Such equations are discussed, for example, in Delves and Mohamed [33, Chapter 12], where numerical solution techniques are discussed. These techniques could be applied to the measurement error problem in suitable cases. However, as Delves and Mohamed indicate, such equations can sometimes be ill-conditioned and do not always have a solution. Thus, for example, for the logistic regression model with additive normal covariate error, Stefanski [34] showed that a corrected score function satisfying (1) does not exist.

One way around the problem is to carry out a mild discretization of the error-prone covariate: fine enough to reduce the bias satisfactorily but not so fine as to lead to numerical problems. This approach will not produce a strictly consistent estimator, but it is reasonable to expect that the bias will be small. This supposition is supported by Cochran's [35] classic work on subclassification, which indicates that the bulk of the information in a continuous variable can often be captured in a discretized version with 4-6 categories. We are currently exploring this discretization approach in more depth.

Alternatively, an attempt can be made to modify the corrected score approach so that it will work for the model under consideration. Huang and Wang [36] developed such a modification for logistic regression. In their work, the terms in the likelihood

score function were re-weighted to yield a new likelihood score function for which a corrected score satisfying (1) can be derived. These authors dealt only with independent additive error, which for the Cox model is already covered by existing corrected score methods [25, 37, 38]. Modification of the corrected score approach under other measurement error structures is an open problem.

Acknowledgments

This work was supported in part by a grant from the U. S. National Cancer Institute. We thank Els Goetghebeur and Malka Gorfine for helpful comments and Ruifeng Li for assistance in the data analysis for the example.

References

[1] Fuller WA. *Measurement Error Models*. Wiley: New York, 1987.

[2] Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models.* Chapman-Hall: London, 1995.

[3] Prentice R. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; **69**:331–342.

[4] Cox DR (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.

[5] Zhou H, Pepe M. Auxilliary covariate data in failure time regression. *Biometrika* 1995; **82**:139–149.

[6] Zhou H, Wang CY. Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B* 2000; **62**:657–665.

[7] Chen YH. Cox regression in cohort studies with validation sampling. *Journal of the Royal Statistical Society, Series B* 2002; **64**:51–62.

[8] Spiegelman D, McDermott A, Rosner, B. The regression calibration method for correcting measurement error bias in nutritional epidemiology. *American Journal of Clinical Nutrition* 1997; **65**(suppl):1179S–1186S.

[9] Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics* 1997; **53**:131–145.

[10] Zucker DM. A pseudo partial likelihood method for semi-parametric survival regression with covariate errors. *Journal of the American Statistical Association* 2005; **100**:1264–1277.

[11] Zucker DM, Spiegelman D. Inference for the proportional hazards model with misclassified discrete-valued covariates. *Biometrics* 2004; **60**:324–334.

[12] Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* 1998; **54**:1407–1419.

[13] Herring AH, Ibrahim JG. Likelihood-based methods for missing covariates in the Cox proportional hazards model. *Journal of the American Statistical Association* 2001; **96**:292–302.

[14] Nakamura T. Corrected score function of errors-in-variables models: methodology and application to generalized linear models. *Biometrika* 1990; **77**:127–137.

[15] Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics* 1992; **48**:829–838.

[16] Akazawa K, Kinukawa N, Nakamura T. A note on the corrected score function corrected for misclassification. *Journal of the Japan Statistical Society* 1998; **28**:115–123.

[17] Wu K, Willett WC, Fuchs CS, Colditz GA, Giovannucci EL. Calcium intake and risk of colon cancer in women and men. *Journal of the National Cancer Institute* 2002; **94**:437–446.

[18] Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.

[19] Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* Wiley: New York, 1980.

[20] Thomas DC. General relative-risk models for survival time and matched case-control analysis. *Biometrics* 1981; **37**:673–686.

[21] Breslow N, Day NE. *Statistical Methods in Cancer Research.* Volume 2: *The Design and Analysis of Cohort Studies.* Oxford University Press: Oxford, 1993.

[22] Cox DR. Partial likelihood. *Biometrika 1975;* **62**:269–276.

[23] Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1100–1120.

[24] Prentice RL, Self SG. Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Annals of Statistics* 1983; **11**:804–812.

[25] Huang Y, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. *Journal of the American Statistical Association* 2000; **95**:1209–1219.

[26] Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration in main study / internal validation study designs with an imperfect reference instrument. *Statistics in Medicine* 2001; **20**:139–160.

[27] Walter SD, Irwig LM. Estimation of test error rates, disease prevalence, and relative risk from misclassified data: a review. *Journal of Clinical Epidemiology* 1988; **41**:923–937.

[28] Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; **52**:797–810.

[29] Torrance-Rynard VL, Walter SD. Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine* 1997; **6**:2157–2175.

[30] Formann AK, Kohlmann T. Latent class analysis in medical research. *Statistical Methods in Medical Research* 1996; **5**:179–211.

[31] Armitage P, Doll R. Stochastic models for carcinogenesis. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability.* Volume 4: *Contributions to Biology and Problems of Medicine*, 19–38. University of California Press: Berkeley, California, 1961.

[32] Willett WC. *Nutritional Epidemiology*, 2nd ed. Oxford University Press: New York, 1998.

[33] Delves LM, Mohamed JL. (1985). *Computational Methods for Integral Equations.* Cambridge University Press: Cambridge, 1985.

[34] Stefanski L. Unbiased estimation of a nonlinear function of a normal mean with application to measurement-error models. *Communications in Statistics, Theory and Methods* 1989; **18**:4335–4358.

[35] Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968; **24**:295–313.

[36] Huang Y, Wang CY. Consistent function methods for logistic regression with errors in covariates. *Journal of the American Statistical Association* 2001; **95**:1209–1219.

[37] Kong FH, Gu M. Consistent estimation in Cox's proportional hazards model with covariate measurement errors. *Statistica Sinica* 1999; **9**:953–969.

[38] Hu C, Lin DY. Cox regression with covariate measurement error. *Scandinavian Journal of Statistics* 2002; **29**:637–655.

[39] Foutz RV. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* 1977; **72**:147–148.

[40] Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 1989; **84**: 1074–1078.

Appendix

Outline Proof of Consistency and Asymptotic Normality

We give here an outline derivation of the asymptotic properties of our estimator. Our goal is to indicate the main steps of the argument without dwelling on the technical details. For simplicity, we focus on the i.i.d. structural model case; our development will go through for the functional model case as well provided that $\{\mathbf{X}_i(t) : i = 1, 2, \ldots\}$ exhibits ergodic behavior suitably similar to that of an i.i.d. sequence. We assume the parameter space is a compact set $\mathcal{B}$ of $\boldsymbol{\beta}$ values of which the true value $\boldsymbol{\beta}_0$ is an interior point. We further assume that regularity conditions along the lines of Andersen and Gill [23] and Prentice and Self [24] (AG/PS) are in force over $\mathcal{B}$.

We take up first the issue of consistency. We are operating under the "asymptotic stability" conditions of AG/PS, which in the i.i.d. case follow from the functional law of large numbers in Andersen and Gill's Appendix III. Define

$$\begin{aligned} s_0(t, \boldsymbol{\beta}) &= E[Y_i(t)\psi(X_i(t), \boldsymbol{\beta})], \\ s_{1r}(t, \boldsymbol{\beta}) &= E[Y_i(t)\psi_r'(X_i(t), \boldsymbol{\beta})], \\ \chi_r(t, \boldsymbol{\beta}) &= E[Y_i(t)\xi_r(\mathbf{X}_i(t); \boldsymbol{\beta})\psi(\mathbf{X}_i(t), \boldsymbol{\beta}_0)]. \end{aligned}$$

Using arguments of AG/PS, including appeal to the asymptotic stability conditions, we find that $U_r(\boldsymbol{\beta})$ converges uniformly over $\mathcal{B}$ to

$$u_r(\boldsymbol{\beta}) = \int \left[\chi_r(t, \boldsymbol{\beta}) - \left(\frac{s_{1r}(t, \boldsymbol{\beta})}{s_0(t, \boldsymbol{\beta})} \right) s_0(t, \boldsymbol{\beta}_0) \right] \lambda_0(t) dt.$$

In view of equations (9), (10), and (11) of our Sec. 3, the same arguments yield the result that $U_r^*(\boldsymbol{\beta})$ converges uniformly over $\mathcal{B}$ to $u_r(\boldsymbol{\beta})$. We thus have the following:

1. The function $\mathbf{U}^*(\boldsymbol{\beta})$, being continuous over $\mathcal{B}$, is therefore uniformly continuous over $\mathcal{B}$.

2. The function $\mathbf{U}^*(\boldsymbol{\beta})$ converges uniformly to $\mathbf{u}(\boldsymbol{\beta})$ over $\mathcal{B}$.

3. As can be seen by inspection (recalling the definition $\xi_r(\mathbf{x};\boldsymbol{\beta}) = \psi_r'(\mathbf{x};\boldsymbol{\beta})/\psi(\mathbf{x};\boldsymbol{\beta})$), $\mathbf{u}(\boldsymbol{\beta}_0) = \mathbf{0}$.

Suppose now that $\boldsymbol{\beta}_0$ is the *only* zero point of $\mathbf{u}(\boldsymbol{\beta})$ in $\mathcal{B}$. This will hold, for example, if $\psi(\mathbf{x};\boldsymbol{\beta}) = \Psi(\mathbf{x}^T\boldsymbol{\beta})$ for a function $\Psi(u)$ that satisfies the conditions stated in Prentice and Self's paper (the classical case $\Psi(u) = e^u$ is covered). Then, given (1)-(3) above and the compactness of the parameter space, convergence of $\hat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_0$ follows by standard subsequence arguments. In the case where $\boldsymbol{\beta}_0$ is not necessarily the only zero point of $\mathbf{u}(\boldsymbol{\beta})$, so that there may be multiple roots to the corrected score equation, the arguments of Foutz [39] imply that there exists a unique sequence of roots that converges to $\boldsymbol{\beta}_0$.

We now discuss the asymptotic normality of $\hat{\boldsymbol{\beta}}$. By Taylor expansion we may write

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{D}(\tilde{\boldsymbol{\beta}})\mathbf{U}(\boldsymbol{\beta}_0),$$

where $\tilde{\boldsymbol{\beta}}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, and hence converges to $\boldsymbol{\beta}_0$. Hence, as in AG/PS, $\mathbf{D}(\tilde{\boldsymbol{\beta}})$ converges to the limiting value of $\mathbf{D}(\boldsymbol{\beta}_0)$, which exists by virtue of the asymptotic stability conditions. It now remains only to show that $n^{\frac{1}{2}}\ \mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically normal.

We first recall the expression (8) for the corrected score function:

$$U_r^*(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\xi_{ir}^*(T_i,\boldsymbol{\beta}) - \frac{e_{1r}^*(T_i)}{e_0^*(T_i)}\right).$$

Since $U_r^*(\boldsymbol{\beta}_0)$ does not have exactly expectation zero, the martingale approach of AG/PS cannot be applied to derive the asymptotic distribution of $\mathbf{U}^*(\boldsymbol{\beta}_0)$. Instead, we follow the approach of Lin and Wei [40]. From this point forward, all quantities involving $\boldsymbol{\beta}$ (including those in which the dependence is suppressed from the notation) are evaluated at the true value $\boldsymbol{\beta}_0$, except in (27), which presents a definition for general $\boldsymbol{\beta}$. We use counting process notation, based on the definition $N_i(t) = I(T_i \leq t, \delta_i = 1)$.

We define

$$\begin{aligned}\bar{N}(t) &= \frac{1}{n}\sum_{i=1}^{n} N_i(t),\\ \mathcal{N}(t) &= E[N_i(t)].\end{aligned}$$

We have, from the law of large numbers along with (9) and (10), that $\bar{N}(t) \to \mathcal{N}(t)$, $e_0^*(t) \to s_0(t)$, and $e_{ir}^*(t) \to s_{1r}(t)$ as $n \to \infty$. Here the dependence on $\boldsymbol{\beta}$ is suppressed from the notation, and, as mentioned above, evaluation is at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. As seen from Andersen and Gill [23, Appendix III], the convergence is uniform in t. In the development below, the symbol $\doteq$ will denote equality up to negligible terms.

We can write

$$\begin{aligned}U_r^*(\boldsymbol{\beta}_0) &= \frac{1}{n}\sum_{i=1}^{n}\int \xi_{ir}^*(t,\boldsymbol{\beta}_0)dN_i(t) - \int \frac{e_{1r}^*(t)}{e_0^*(t)}d\bar{N}(t)\\ &= \frac{1}{n}\sum_{i=1}^{n}\int \xi_{ir}^*(t,\boldsymbol{\beta}_0)dN_i(t) - \int \frac{e_{1r}^*(t)}{e_0^*(t)}d\mathcal{N}(t) - \int \frac{s_{1r}(t)}{s_0(t)}d(\bar{N}-\mathcal{N})(t)\\ &\quad - \int \left[\frac{e_{1r}^*(t)}{e_0^*(t)} - \frac{s_{1r}(t)}{s_0(t)}\right] d(\bar{N}-\mathcal{N})(t)\\ &\doteq \frac{1}{n}\sum_{i=1}^{n}\int \xi_{ir}^*(t,\boldsymbol{\beta}_0)dN_i(t) - \int \frac{e_{1r}^*(t)}{e_0^*(t)}d\mathcal{N}(t) - \int \frac{s_{1r}(t)}{s_0(t)}d(\bar{N}-\mathcal{N})(t).\end{aligned}$$

In addition,

$$\begin{aligned}\frac{e_{1r}^*(t)}{e_0^*(t)} &= \frac{e_{1r}^*(t)}{s_0(t)} + e_{1r}^*(t)\left[\frac{1}{e_0^*(t)} - \frac{1}{s_0(t)}\right]\\ &= \frac{e_{1r}^*(t)}{s_0(t)} - \left[\frac{e_{1r}^*(t)}{e_0^*(t)s_0(t)}\right](e_0^*(t) - s_0(t))\\ &\doteq \frac{1}{s_0(t)}\left[e_{1r}^*(t) - \frac{s_{1r}(t)}{s_0(t)}(e_0^*(t) - s_0(t))\right].\end{aligned}$$

Thus, substituting and re-arranging, we obtain

$$\begin{aligned}U_r^*(\boldsymbol{\beta}_0) &\doteq \frac{1}{n}\sum_{i=1}^{n}\int \xi_{ir}^*(t,\boldsymbol{\beta}_0)dN_i(t) - \int \frac{s_{1r}(t)}{s_0(t)}d(\bar{N}-\mathcal{N})(t)\\ &\quad - \int \frac{1}{s_0(t)}\left[e_{1r}^*(t) - \frac{s_{1r}(t)}{s_0(t)}(e_0^*(t) - s_0(t))\right] d\mathcal{N}(t)\\ &= \frac{1}{n}\sum_{i=1}^{n}\int \left(\xi_{ir}^*(t,\boldsymbol{\beta}_0) - \frac{s_{1r}(t)}{s_0(t)}\right) dN_i(t) - \int \left(\frac{e_{1r}^*(t)}{s_0(t)} - \frac{s_{1r}(t)}{s_0(t)}\frac{e_0^*(t)}{s_0(t)}\right) d\mathcal{N}(t)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\int\left(\xi^*_{ir}(t,\boldsymbol{\beta}_0)-\frac{s_{1r}(t)}{s_0(t)}\right)dN_i(t)\\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\int\left(\frac{Y_i(t)\eta_{ir}(t,\boldsymbol{\beta}_0)}{s_0(t)}-\frac{s_{1r}(t)}{s_0(t)}\frac{Y_i(t)\psi^*_i(t,\boldsymbol{\beta}_0)}{s_0(t)}\right)d\mathcal{N}(t)\\
&= \frac{1}{n}\sum_{i=1}^{n}\Upsilon_{ir}(\boldsymbol{\beta}_0),
\end{aligned}
$$

where

$$
\Upsilon_{ir}(\boldsymbol{\beta})=\delta_i\left[\xi^*_{ir}(T_i,\boldsymbol{\beta})-\frac{e^*_{1r}(T_i)}{e^*_0(T_i)}\right]-\int Y_i(t)\left[\frac{\eta_{ir}(t,\boldsymbol{\beta})}{s_0(t)}-\frac{s_{1r}(t)}{s_0(t)}\frac{\psi^*_i(t,\boldsymbol{\beta})}{s_0(t)}\right]d\mathcal{N}(t). \quad (27)
$$

From this result it follows immediately from the classical central limit theorem for i.i.d. random vectors that $n^{\frac{1}{2}}\ \mathbf{U}^*(\boldsymbol{\beta}_0)$ is asymptotically mean-zero multivariate normal. It is straightforward to see that the asymptotic covariance matrix of $n^{\frac{1}{2}}\ \mathbf{U}^*(\boldsymbol{\beta}_0)$ can be estimated consistently by the expression (14) evaluated at $\hat{\boldsymbol{\beta}}$.

Finally, we turn to the cumulative hazard estimator (16). This estimator can be written as

$$
\hat{\Lambda}_0(t)=\int_0^t\frac{d\bar{N}(u)}{e^*_0(u,\hat{\boldsymbol{\beta}})}\ .
$$

Using arguments similar to the above, we find that

$$
\hat{\Lambda}_0(t)-\Lambda_0(t)\doteq-\mathbf{a}^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{n}\sum_{i=1}^{n}\int_0^t\frac{dN_i(u)}{s_0(u,\boldsymbol{\beta}_0)}-\frac{1}{n}\sum_{i=1}^{n}\int_0^t\frac{Y_i(u)\psi^*_i(u,\boldsymbol{\beta}_0)}{s_0(u,\boldsymbol{\beta}_0)}\,d\mathcal{N}(u),
$$

with the r-th element of $\mathbf{a}$ given by

$$
a_r=\int_0^t\frac{s_{1r}(u,\boldsymbol{\beta}_0)}{s_0(u,\boldsymbol{\beta}_0)^2}\,d\mathcal{N}(u).
$$

Using the approximation $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\doteq\mathbf{D}(\boldsymbol{\beta})^{-1}\mathbf{U}^*(\boldsymbol{\beta}_0)$, and defining $\mathbf{c}=-\mathbf{D}(\boldsymbol{\beta})^{-1}\mathbf{a}$, we obtain

$$
\hat{\Lambda}_0(t)-\Lambda_0(t)\doteq\frac{1}{n}\sum_{i=1}^{n}\Upsilon^*_i(\boldsymbol{\beta}_0), \quad (28)
$$

where

$$
\Upsilon^*_i(\boldsymbol{\beta})=\sum_{r=1}^{p}c_r\Upsilon_{ir}(\boldsymbol{\beta})+\int_0^t s_0(u,\boldsymbol{\beta})^{-1}\,dN_i(u)-\int_0^t\frac{Y_i(u)\psi^*_i(u,\boldsymbol{\beta})}{s_0(u,\boldsymbol{\beta})}\,d\mathcal{N}(u).
$$

Once again, asymptotic normality of the estimator and consistency of the proposed variance estimator (17) are immediately apparent. When, as in Sec. 4, we estimate the

parameters $\boldsymbol{\omega}$ that determine the classification probabilities, the representation (28) becomes

$$\hat{\Lambda}_0(t) - \Lambda_0(t) \doteq \frac{1}{n}\sum_{i=1}^{n} \Upsilon_i^*(\boldsymbol{\beta}_0) + \mathbf{h}^T(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0),$$

where the ν-th element of $\mathbf{h}$ is given by

$$h_\nu = [\dot{\mathbf{U}}^*(\boldsymbol{\beta}, \boldsymbol{\omega})^T \mathbf{c}(\boldsymbol{\beta})]_\nu - \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{Y_i(u)}{s_0(u, \boldsymbol{\beta})} \sum_{l=1}^{K} [\dot{B}_\nu^{(i)}]_{k(i,u)l} \psi(\mathbf{w}_l, \mathbf{Z}_i(u); \boldsymbol{\beta})\, d\mathcal{N}(u).$$

Table 1

Simulation Results for the Case of a Single Binary Covariate
Classification Rates Assumed Known
Sample Size = 2,000, Unexposed Cumulative Incidence = 25%

Percent Exposed	Error Rate	True RR	Percent Bias In Estimated Log RR			Standard Deviation – CSCORE		MSE Ratio	95% CI Coverage
			Naive Cox	CSCORE	FWMLE	Empirical	Mean of Estimates		
5 %	1 %	1.5	-15.89	-1.93	-1.92	0.196	0.194	1.00	95.12
5 %	1 %	2.0	-14.00	-0.46	-0.43	0.176	0.173	1.00	95.10
5 %	5 %	1.5	-48.45	-4.73	-4.77	0.264	0.255	1.00	94.90
5 %	5 %	2.0	-46.08	-2.19	-2.16	0.231	0.218	1.00	93.90
5 %	10 %	1.5	-66.02	-6.64	-6.99	0.349	0.338	1.03	94.76
5 %	10 %	2.0	-64.16	-2.87	-2.90	0.313	0.281	0.99	93.62
5 %	20 %	1.5	-82.24	-17.22	-23.81	0.591	0.620	1.22	93.07
5 %	20 %	2.0	-80.92	-7.20	-9.87	0.531	0.482	1.04	91.87
25 %	1 %	1.5	-3.03	-0.06	-0.02	0.096	0.096	1.00	94.92
25 %	1 %	2.0	-3.01	-0.15	-0.10	0.090	0.089	1.00	94.46
25 %	5 %	1.5	-14.09	0.02	0.06	0.106	0.105	1.00	94.56
25 %	5 %	2.0	-13.89	-0.28	-0.22	0.100	0.097	1.00	94.62
25 %	10 %	1.5	-26.61	-0.01	0.04	0.123	0.120	1.00	94.90
25 %	10 %	2.0	-26.14	-0.33	-0.30	0.112	0.110	1.00	94.60
25 %	20 %	1.5	-48.41	-0.54	-0.49	0.167	0.164	1.00	94.96
25 %	20 %	2.0	-47.42	-0.18	-0.15	0.156	0.149	0.99	94.10
40 %	1 %	1.5	-2.43	-0.36	-0.32	0.088	0.087	1.00	94.62
40 %	1 %	2.0	-2.11	-0.03	0.01	0.081	0.082	1.00	95.22
40 %	5 %	1.5	-10.03	0.39	0.43	0.095	0.094	1.00	94.58
40 %	5 %	2.0	-10.35	0.02	0.07	0.091	0.089	1.00	94.90
40 %	10 %	1.5	-20.63	0.01	0.05	0.108	0.106	1.00	94.74
40 %	10 %	2.0	-20.40	0.28	0.30	0.102	0.101	1.00	95.04
40 %	20 %	1.5	-41.04	-0.34	-0.33	0.140	0.142	1.00	95.58
40 %	20 %	2.0	-40.66	0.29	0.34	0.135	0.135	1.00	94.90

Legend:

RR = Relative Risk
CSCORE = Corrected Score Estimator
FWMLE = Full Weibull Maximum Likelihood Estimator
MSE Ratio = ratio of mean square error of FWMLE to that of CSCORE

Table 2

Simulation Results for the Corrected Score Estimate for a Single Binary Covariate
Classification Rates Estimated From an External Replicate Measures Sample of Size 250
Unexposed Cumulative Incidence = 25%

Sample Size	Percent Exposed	Error Rate	True RR	% Bias In $\hat{\beta}$	Empirical Std Dev	Mean of SD Estimates	95% CI Coverage
2,000	5 %	1 %	1.5	-1.10	0.206	0.204	93.14
	5 %	1 %	2.0	0.22	0.185	0.185	93.72
	5 %	5 %	1.5	-1.96	0.293	0.292	93.30
	5 %	5 %	2.0	2.54	0.278	0.270	94.17
	5 %	10 %	1.5	-1.31	0.441	0.491	94.30
	5 %	10 %	2.0	4.48	0.412	0.439	94.87
	5 %	20 %	1.5	13.13	0.670	1.663	96.83
	5 %	20 %	2.0	11.30	0.637	1.758	95.37
	25 %	1 %	1.5	0.36	0.097	0.097	94.90
	25 %	1 %	2.0	0.23	0.091	0.090	94.70
	25 %	5 %	1.5	0.34	0.109	0.107	94.54
	25 %	5 %	2.0	0.47	0.101	0.100	95.02
	25 %	10 %	1.5	0.57	0.123	0.124	94.90
	25 %	10 %	2.0	0.83	0.117	0.117	95.44
	25 %	20 %	1.5	1.51	0.183	0.181	95.24
	25 %	20 %	2.0	3.39	0.187	0.179	95.56
	40 %	1 %	1.5	1.03	0.087	0.087	95.30
	40 %	1 %	2.0	0.59	0.083	0.083	94.84
	40 %	5 %	1.5	0.36	0.095	0.095	95.12
	40 %	5 %	2.0	0.54	0.091	0.091	94.78
	40 %	10 %	1.5	0.63	0.106	0.108	95.24
	40 %	10 %	2.0	0.66	0.106	0.104	94.98
	40 %	20 %	1.5	2.08	0.150	0.149	94.96
	40 %	20 %	2.0	1.79	0.148	0.148	95.72
1,000	5 %	1 %	1.5	-4.07	0.296	0.297	91.58
	5 %	1 %	2.0	0.68	0.263	0.263	92.42
	5 %	5 %	1.5	-7.94	0.426	0.442	91.78
	5 %	5 %	2.0	0.56	0.382	0.383	92.30
	5 %	10 %	1.5	-12.20	0.600	0.790	91.48
	5 %	10 %	2.0	-1.38	0.546	0.641	92.79
	5 %	20 %	1.5	-15.98	2.805	2.130	96.76
	5 %	20 %	2.0	-12.77	1.776	1.913	95.04
	25 %	1 %	1.5	-0.47	0.137	0.137	94.56
	25 %	1 %	2.0	0.55	0.126	0.127	95.46
	25 %	5 %	1.5	0.25	0.153	0.151	94.76
	25 %	5 %	2.0	0.89	0.141	0.141	95.16
	25 %	10 %	1.5	0.20	0.176	0.175	94.80
	25 %	10 %	2.0	0.33	0.165	0.163	94.24
	25 %	20 %	1.5	0.34	0.253	0.255	94.78
	25 %	20 %	2.0	2.91	0.241	0.239	95.18
	40 %	1 %	1.5	0.33	0.121	0.124	95.32
	40 %	1 %	2.0	0.22	0.116	0.117	95.58
	40 %	5 %	1.5	0.80	0.135	0.135	94.90
	40 %	5 %	2.0	0.68	0.130	0.129	94.86
	40 %	10 %	1.5	0.50	0.154	0.153	95.20
	40 %	10 %	2.0	0.60	0.147	0.146	94.88
	40 %	20 %	1.5	2.92	0.210	0.211	95.12
	40 %	20 %	2.0	1.93	0.202	0.204	95.84

RR = Relative Risk

Table 3
Estimated Coefficients and Standard Errors for the Nurses Health Study of the Relationship Between Dietary Calcium Intake and Distal Colon Cancer Incidence

	High Calcium		BMI of 22 to < 25		BMI of 25 to < 30		BMI of 30+		Aspirin Use	
Method	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Cox	-0.3448	0.1694	0.6837	0.2240	0.5352	0.2395	0.5729	0.2876	-0.4941	0.1954
CS0	-0.7121	0.3690	0.7124	0.2247	0.5776	0.2419	0.6157	0.2892	-0.4994	0.1955
CS1	-0.7121	0.3832	0.7124	0.2249	0.5776	0.2423	0.6157	0.2896	-0.4994	0.1955

Legend:

Cox = classical Cox regression analysis
CS0 = corrected score method, observed classification matrix taken as known
CS1 = corrected score method, accounting for uncertainty in the classification matrix