Statistica Sinica **11**(2001), 1081-1104

RECALIBRATION BASED ON AN APPROXIMATE RELATIVE RISK ESTIMATOR IN COX REGRESSION WITH MISSING COVARIATES

C. Y. Wang*, Sharon X. Xie+ and Ross L. Prentice*

**Fred Hutchinson Cancer Research Center and +Penn Sylvania State University*

Abstract: We consider estimation in Cox (1972) regression with missing covariates. We focus on the situation when observable covariates and surrogates are continuous. To estimate the induced relative risk, we approximate it by a function of the conditional mean and the conditional variance of the missing variable, given observable covariate and surrogate variables in the risk set at each failure time. This approach may be considered as a higher order extension of the usual regression calibration method, and hence can be expected to reduce bias, especially when the relative risk or the estimating error from the surrogate variables is large. The proposed estimator arises from an approximate approach, so that the magnitude of any bias needs to be studied. Asymptotic distribution theory is developed and small sample performance is examined. We illustrate the method by an example from a medical study.

Key words and phrases: Cox regression, estimating equation, measurement error model, regression calibration, surrogate covariate.

1. Introduction

We consider analyses of failure time regression data with missing covariates. Let $(T_i, \delta_i), i = 1, \ldots, n$, be the observed failure times and noncensoring indicators for the ith study subject, where T is the minimum of the failure and the potential censoring times. We consider the Cox (1972) hazard regression model

$$\lambda(t; X_i, Z_i) = \lambda_0(t)\exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\},$$

where $\lambda_0(\cdot) \geq 0$ is an unspecified baseline hazard function, X_i is a covariate vector of dimension p which might be missing in some study subjects, Z_i is a vector of the covariates that is always observed and $\beta = (\beta_1', \beta_2')'$ are the regression coefficients to be estimated. Let W be a surrogate measure for X, so that the hazard function for T is independent of W given (X, Z). Here, we assume that a validation set is available where all variables (T, δ, X, Z, W) are observed. The nonvalidation set includes data on (T, δ, Z, W), but not X.

Prentice (1982) showed, under the usual independent censoring assumption, that given (Z, W) the hazard function is independent of the censorship and the induced hazard function given (Z, W) is

$$\lambda(t; Z, W) = \lambda_0(t)\exp\{\beta_2' Z(t)\}E[\exp\{\beta_1' X(t)\}|Y(t) = 1, Z(t), W(t)], \qquad (1)$$

where $Y(t) = I[T \geq t]$ is an at risk indicator. He proposed the regression calibration approach, when failure occurrence is rare and X given (Z, W) is normal. A detailed study of the regression calibration approach in failure time regression with missing covariate variables was recently presented in Wang, Hsu, Feng and Prentice (1997), including the possibility that the missingness mechanism may depend on the observed covariates. If the size of the relative risk is moderate, they found that this method has limited bias and is highly efficient in various situations. Their method is useful for both discrete and continuous covariate/surrogate variables. In the measurement error problem without a validation set, Clayton (1991) proposed a form of risk set calibration, which is restricted in that the slope of the regression line $E\{X(t)|Z(t), W(t), Y(t) = 1\}$ was held constant across risk sets (Carroll, Ruppert and Stefanski (1996, p.256)). Xie, Wang and Prentice (2001) proposed a more relaxed risk set calibration method in which $E\{X(t)|Z(t), W(t), Y(t) = 1\}$ is approximated by a linear function of (Z, W) and both the slope and the intercept vary across risk sets.

When the missingness mechanism depends on the observed covariate (Z, W), and Z and W are discrete, Zhou and Pepe (1995) proposed nonparametric estimation of the induced relative risk function $\exp\{\beta_2' Z(t)\}E[\exp\{\beta_1' X(t)\}|Y(t) = 1, Z(t), W(t)\}$ in (1). Based on this, they developed an estimated partial likelihood procedure. Paik and Tsai (1997) considered a similar approach for discrete (Z, W), but they proposed to further condition the associated conditional expectations on the discretized observed failure time if the missingness depends also on the observed failure time. Lin and Ying (1993) proposed an approximate partial likelihood estimator for the missing data problem but it does not accommodate the use of the surrogate covariate W and it may experience noteworthy bias if the missingness rate depends on Z; this was shown in some simulation comparisons of Wang, et al. (1997) and Paik and Tsai (1997). Robins, Rotnitzki and Zhao (1994, Section 8.3) proposed a class of inverse selection probability weighted estimators. The implementation of an augmented inverse probability weighted estimator was recently studied by Wang and Chen (2001).

The current research is motivated by (i) the goal of reducing the bias of the regression calibration approach when the absolute value of the relative risk parameter β_1 and the measurement error component of W may be large; and (ii) the need to approximate the induced relative risk when some components of (Z, W) are continuous. The proposed estimation procedure involves estimates of the nuisance parameters in the conditional mean and the conditional variance of

X given (Z, W) in each successive risk set, as follow-up time increases, based on the validation data, which are then used to estimate the induced relative risks when X is missing. In this paper, we allow the missingness to depend on (Z, W). The idea of the regression calibration estimator will be reviewed in Section 2. In Section 3, we describe the proposed method in detail. Asymptotic distribution theory is given in Section 4. Performance of the estimator is studied via simulations in Section 5. We illustrate the proposed estimator by an example from a medical study in Section 6. Section 7 provides a simple algorithm to approximate the proposed estimator when X is a scalar. This algorithm provides a practical illustration of implementing the method based on a replacement or recalibration model. Numerical results show that in general it has almost ignorable differences compared to the proposed method. Technical proofs are given in the Appendix.

2. Regression Calibration

Prentice (1982) noted that under normal measurement error and rare disease assumptions, the induced relative risk could be approximated as

$$\begin{aligned}&\exp\{\beta_2' Z\} E[\exp\{\beta_1 X\} | Y(t) = 1, Z, W] \\ &\approx \exp\{\beta_2' Z\} \exp\Big[\beta_1 E\{X|Z, W\} + \frac{1}{2}\beta_1' \mathrm{var}\{X|Z, W\}\beta_1\Big].\end{aligned}$$

This suggests that if the disease is rare and the variance of X given (Z, W) does not depend on (Z, W), then regression calibration (RC) may work well.

Wang, et al. (1997) studied the method and noted that it performed well as long as the relative risk parameter associated with the missing covariate variable is not too large. The RC estimator can be carried out by the following two steps:

(i) Estimate the missing X by a specified function $g(Z, W, \widehat{\alpha})$, which is modeled from the validation set parametrically and $\widehat{\alpha}$ is the associated estimated parameter.

(ii) Apply a Cox regression procedure by using covariate data (X, Z) in the validation set and $\{g(Z, W, \widehat{\alpha}), Z\}$ in the nonvalidation set.

They also considered a robust sandwich estimator for the covariance estimation, although a bootstrap procedure is an alternative. If we modify (i) by $g(Z, W, \widehat{\alpha}_t)$ for each risk set at time t, then one has a type of risk set regression calibration approach.

3. The Approximate Relative Risk Estimator

The estimation procedure proposed here is motivated by a series expansion. Observe that if $E[\exp\{\beta_1' X(t)\} | Y(t) = 1, Z(t), W(t)]$ exists, then

$$E[\exp\{\beta_1' X(t)\} | Y(t) = 1, Z(t), W(t)]$$

$$=\exp\Big[\beta_1' E\{X(t)|Z(t),W(t),Y(t)=1\}+(1/2)\beta_1'\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}\beta_1 + \sum_{r=3}^{\infty}\mathcal{E}_r\{Z(t),W(t),t\}\Big], \tag{2}$$

where $\mathcal{E}_r\{Z(t),W(t),t\}$ is a function of β_1 and the rth conditional cumulant of X given (Z,W) in the risk set at time t. In (2), $\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}$ denotes the conditional variance-covariance matrix if $X(t)$ is a vector. Expansion (2) is usually called a conditional cumulant generating function (Kendall and Stuart (1977, Chapter 3)). We note that if $E([X(t)-E\{X(t)|Z(t),W(t),Y(t)=1\}]^r|\ Z(t),W(t),Y(t)=1)$ is independent of $\{Z(t),W(t)\}$ for $r\geq 3$, then (2) becomes

$$\begin{aligned} &E\left[\exp\{\beta_1' X(t)\}|Y(t)=1,Z(t),W(t)\right] \\ &= c(t)\exp[\beta_1' E\{X(t)|Z(t),W(t),Y(t)=1\} \\ &\quad +(1/2)\beta_1'\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}\beta_1], \end{aligned} \tag{3}$$

for some $c(t)$, independent of $\{Z(t),W(t)\}$. For example, (3) holds when $X(t)$ given $\{Z(t),W(t),\ Y(t)=1\}$ is normal.

In the development of our method we assume (3) holds. Although higher order conditional cumulants could be considered for bias reduction purposes, it may be that any related bias reduction will be comparatively minor unless $|\beta_1|$ and the measurement error in W are very large. Furthermore, a higher order (>2) approximation would require a large validation sample size in each risk set since additional nuisance parameters would need to be estimated within the calibration procedure.

Under (3), the induced hazard function (1) is

$$\lambda(t;Z,W)=c(t)\lambda_0(t)\exp\Big[\beta_1' E\{X(t)|Z(t),W(t),Y(t)=1\} + (1/2)\beta_1'\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}\beta_1+\beta_2' Z(t)\Big].$$

Let η_i be the indicator of subject i being in the validation set, *i.e.* X_i is not missing if $\eta_i=1$. Let $r\{\beta,X(t),\ Z(t)\}=\exp\{\beta_1' X(t)+\beta_2' Z(t)\}$; $r^*\{\beta,Z(t),W(t)\}=\exp[\beta_1' E\{X(t)|Z(t),\ W(t),Y(t)=1\}+(1/2)\beta_1'\mathrm{var}\{X(t)|Z(t),\ W(t),Y(t)=1\}\beta_1+\beta_2' Z(t)]$. The induced relative risk for subject i is

$$R_i(\beta,t)=[r\{\beta,Z_i(t),W_i(t)\}]^{\eta_i(t)}[c(t)r^*\{\beta,Z_i(t),W_i(t)\}]^{1-\eta_i(t)}. \tag{4}$$

As a result, the induced partial likelihood of the observed data under (3) is

$$\prod_{i=1}^{n}\Big[\frac{R_i(\beta,T_i)}{\sum_{j\in\mathcal{G}_i}R_j(\beta,T_i)}\Big]^{\delta_i}, \tag{5}$$

where $\mathcal{G}_i$ is the risk set at time T_i. Let $R_i^{(1)}(\beta,t)=(\partial/\partial\beta)R_i(\beta,t)$. Then given $r\{\beta,X(t),Z(t)\}$, $r^*\{\beta,Z(t),W(t)\}$ and $c(t)$, the induced estimating equation derived from (5) is

$$n^{-1/2}\sum_{i=1}^{n}\delta_i\Big\{\frac{R_i^{(1)}(\beta,T_i)}{R_i(\beta,T_i)}-\frac{\sum_{j=1}^{n}Y_j(T_i)R_j^{(1)}(\beta,T_i)}{\sum_{j=1}^{n}Y_j(T_i)R_j(\beta,T_i)}\Big\}=0.$$

However, $E\{X(t)|Z(t),W(t),Y(t)=1\}$, $\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}$ and $c(t)$ are not known and hence need to be estimated. Observe that the relationship between X_i and (Z_i,W_i) in the risk set can be examined from the validation data. We assume that there exists a known function $\mu\{Z(t),W(t),\mathcal{A}_t\}$ such that

$$E\{X(t)|Z(t),W(t),Y(t)=1\}=\mu\{Z(t),W(t),\mathcal{A}_t\},\tag{6}$$

where matrix $\mathcal{A}_t=(\alpha_{1t},\ldots,\alpha_{pt})$ and vector $\alpha_{jt}=(\alpha_{j0t},\ldots,\alpha_{jat})'$ is the parameter of the jth ($j=1.\ldots,p$) component of vector $E\{X(t)|Z(t),W(t),Y(t)=1\}$. For example, if $X(t)$ is a scalar ($p=1$) and $X(t)$ given $\{Z(t),W(t)\}$ is linear, then we may consider modeling $\mu\{Z(t),W(t),\mathcal{A}_t\}=\alpha_{0t}+\alpha_{1t}'Z(t)+\alpha_{2t}'W(t)$ for some $(\alpha_{0t},\alpha_{1t}',\alpha_{2t}')$. Write vector X as $(X_{(1)},\ldots,X_{(p)})'$. We also assume that there exists a known matrix function $\sigma^2\{Z(t),W(t),\Gamma_t\}$ such that the variance-covariance matrix of $X(t)$ given $\{Z(t),W(t),Y(t)=1\}$ is

$$\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}=\sigma^2\{Z(t),W(t),\Gamma_t\},\tag{7}$$

where $\Gamma_t=(\Gamma_{1t},\ldots,\Gamma_{pt})$ is a $(pb\times p)$ matrix, $\Gamma_{jt}=(\Gamma_{1jt}',\ldots,\Gamma_{pjt}')$ is a $(pb\times 1)$ vector and Γ_{ljt} is a $(b\times 1)$ vector for the parameter of $\mathrm{cov}\{X_{(l)}(t),X_{(j)}(t)|Z(t),W(t),Y(t)=1\}$ for some positive integer b. For example, if X is a scalar, then one may consider modeling $\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}\geq 0$ by a quadratic function of $\{Z(t),W(t)\}$. By the symmetry of $\mathrm{var}\{X(t)|Z(t),W(t),Y(t)=1\}$, Γ_t contains $bp(p+1)/2$ parameters.

Under (6), we may obtain a $n^{1/2}$ consistent estimate of $\mathcal{A}_t$ based on the validation data set in the risk set $\mathcal{G}_i$ such that $T_i\geq t$ under the constraint that the validation sample selection does not depend on the outcomes (T_i,δ_i). A convenient choice of the estimation procedure for α_{jt} ($j=1,\ldots,p$) arises by applying a least square (LS) estimator of the regression of $X_{(j)}(t)$ on $\{Z(t),W(t)\}$ in the risk set, assuming $X_{(j)}(t)$ is approximately a linear function of $Z(t)$ and $W(t)$. In general, we assume that there exists a $pa\times pa$ positive definite matrix Q_t and some $a\times p$ matrix $\mathcal{X}_i(t)$ such that

$$n^{1/2}\{\mathrm{vec}(\widehat{\mathcal{A}}_t-\mathcal{A}_t)\}=Q_t^{-1}n^{-1/2}\sum_{i=1}^{n}\eta_i(t)Y_i(t)\mathrm{vec}\{\mathcal{X}_i(t)\mathrm{diag}(e_{it})\}+o_p(1),\tag{8}$$

where $e_{it} = X_i(t) - \mu\{Z_i(t), W_i(t), \mathcal{A}_t)\}$ and $\text{diag}(e_{it})$ is a $p \times p$ diagonal matrix with the components of e_{it} as the diagonal elements. Observe that (8) holds when $\widehat{\mathcal{A}}$ is obtained from the LS estimator. If X is scalar then $\mathcal{X}_i(t) = \partial\mu\{Z(t), W(t), \alpha\}/\partial\alpha$, $Q_t = \text{plim}_{n\to\infty}\{n^{-1}\sum_{i=1}^n \eta_i Y_i(t)\mathcal{X}_i(t)\mathcal{X}_i'(t)\}$, where $\text{plim}_{n\to\infty}$ denotes the probability limit as $n \to \infty$.

Under (7), for the estimation of the covariance of $X_{(l)}(t)$ and $X_{(j)}(t)$ given $\{Z(t), W(t)\}$ in the risk set, Γ_{ljt} may be obtained from the LS estimates. Write e_{it} as $(e_{it(1)}, \ldots, e_{it(p)})'$. Similar to (8), we assume there exists a $bp^2 \times bp^2$ positive definite matrix H_t and some $b \times p^2$ matrix $\mathcal{Y}_i(t)$ such that

$$n^{1/2}\{\text{vec}(\widehat{\Gamma}_t - \Gamma_t)\} = H_t^{-1} n^{-1/2} \sum_{i=1}^n \eta_i(t) Y_i(t) \text{vec}\{\mathcal{Y}_i(t) F(e_{it})\}, \tag{9}$$

where $F(e_{it}) = \text{diag}\Big[\text{diag}\{e_{it(l)}e_{it(j)} - \widehat{e}_{it(l)}\widehat{e}_{it(j)},\ l = 1, \ldots, p\}, j = 1, \ldots, p\Big]$.

After estimating the parameters $\mathcal{A}_t$'s and Γ_t's, $c(t)$ in (4) needs to be estimated. We note that when the selection probability is a function of $Z(t)$ and $W(t)$, we have

$$\begin{aligned}
&E\Big[r\{\beta, X(t), Z(t)\} | Y(t) = 1, \eta(t) = 1\Big] \\
&= E\Big(E\Big[r\{\beta, X(t), Z(t)\} | Y(t) = 1, \eta(t) = 1, Z(t), W(t)\Big] \Big| Y(t) = 1, \eta(t) = 1\Big) \\
&= c(t) E\Big[r^*\{\beta, Z(t), W(t)\} | Y(t) = 1, \eta(t) = 1\Big].
\end{aligned}$$

Therefore, $c(t)$ may be estimated by

$$\widehat{c}(t) = \frac{\sum_{j=1}^n Y_j(t)\eta_j(t) r\{\beta, X_j(t), Z_j(t)\}}{\sum_{j=1}^n Y_j(t)\eta_j(t)\widehat{r}^*\{\beta, Z_j(t), W_j(t)\}},$$

where $\widehat{r}^*\{\beta, Z_j(t), W_j(t)\}$ is defined as $r^*\{\beta, Z_j(t), W_j(t)\}$ but with estimated $\mathcal{A}_t$'s and Γ_t's. Plugging the $\widehat{\mathcal{A}}_t$, $\widehat{\Gamma}_t$ and $\widehat{c}(t)$ into (4), we write $R(\beta, t)$ as $R(\beta,\ t,\ \widehat{\mathcal{A}}_t,\ \widehat{\Gamma}_t)$ and $R^{(1)}(\beta, t)$ as $R^{(1)}(\beta,\ t,\ \widehat{\mathcal{A}}_t,\ \widehat{\Gamma}_t)$. We propose an approximate relative risk (ARR) estimator, $\widehat{\beta} = (\widehat{\beta}_1', \widehat{\beta}_2')'$, which solves

$$n^{-1/2} \sum_{i=1}^n \delta_i \Big\{ \frac{R_i^{(1)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i})}{R_i(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i})} - \frac{S^{(1)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i})}{S^{(0)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i})} \Big\} = 0, \tag{10}$$

where

$$\begin{aligned}
S^{(0)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i}) &= n^{-1} \sum_{j=1}^n Y_j(T_i) R_j(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i}); \\
S^{(1)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i}) &= n^{-1} \sum_{j=1}^n Y_j(T_i) R_j^{(1)}(\beta, T_i, \widehat{\mathcal{A}}_{T_i}, \widehat{\Gamma}_{T_i}).
\end{aligned}$$

For further insight into this modeling and notation, see the special case described in Section 5.

Generally, our proposed estimator is a refined approach of the RC estimator. We note that (6) and (7) are additional assumptions on the conditional linear regression for $X(t)$. This time-dependent conditional linear regression model at each risk set can generally be properly fitted and examined since $\{X(t), Z(t), W(t)\}$ are available from the validation data. Therefore, model misspecification can be minimized in applications by simple data analyses. Further cautions regarding this issue will be given in the Discussion section.

4. Asymptotic Distribution Theory

Before we present the large sample theory, we first explain the dimensionality issue regarding nuisance parameters. When (Z, W) are discrete, to implement the estimated partial likelihood (EPL) of Zhou and Pepe (1995), at a failure time t we need to estimate the induced relative risk by

$$\begin{aligned}&\widehat{r}^*\{\beta, Z_i(t), W_i(t)\}\\ &= \frac{\sum_{j=1}^n \exp\{\beta_1' X_j(t)\} I[T_j \geq t, Z_j(t) = Z_i(t), W_j(t) = W_i(t)]}{\sum_{j=1}^n I[T_j \geq t, Z_j(t) = Z_i(t), W_j(t) = W_i(t)]} \exp\{\beta_2' Z_i(t)\},\end{aligned}$$

if X_i is missing. In addition, $E\{(\partial/\partial\beta) r(\beta, X_i, Z_i) | Z_i, W_i, T_i \geq t\}$ is involved in the EPL estimating equation and hence needs to be estimated similarly. If the support of (Z, W) has q discrete values, then it is essentially the case that there are $2q$ nuisance parameters at a failure time. For continuous (Z, W), to implement our proposed ARR estimator, we apply the LS estimates of $\mathcal{A}_t$ and Γ_t which satisfy (8) and (9). Therefore, similar to the EPL estimator, the dimension of all nuisance parameters goes to infinity if n does. Note that, in the EPL estimator of Zhou and Pepe (1995), Theorem 3.1 of Andersen and Gill (1982) was applied to show that the convergence of $\widehat{r}^*\{\beta, Z_i(t), W_i(t)\}$ is uniform in t. This is basically a type of uniform convergence of the Law of Large Numbers on the time axis. In our problem, the uniform convergence of the LS estimates of $\mathcal{A}_t$ and Γ_t follows similarly. Therefore, our ARR estimator can be considered as an extension of the EPL estimator to continuous covariates, and a more precise approximation than the RC estimator. On the other hand, infinite dimensionality of nuisance parameters occurs if one instead applies a maximum likelihood estimator (Chen and Little (1999)): Let $\mathcal{L}$ denote likelihood, the maximum likelihood estimator is based on the likelihood that

$$\frac{\partial}{\partial\beta}\log\mathcal{L}(T_i, \delta_i | Z_i, W_i) = E\{\frac{\partial}{\partial\beta}\log\mathcal{L}(T_i, \delta_i | X_i, Z_i) | T_i, \delta_i, Z_i, W_i\}.$$

Here the conditional expectation can be carried out by noting

$$\mathcal{L}(X_i|T_i,\delta_i,Z_i,W_i)=\frac{\mathcal{L}(T_i,\delta_i|X_i,Z_i)\mathcal{L}(X_i|Z_i,W_i)}{\int_x \mathcal{L}(T_i,\delta_i|X=x,Z_i)\mathcal{L}(X=x|Z_i,W_i)dx}.$$

Therefore, infinite dimensionality occurs since $\mathcal{L}(T_i,\delta_i|X_i,Z_i)$ contains the unknown baseline hazard $\Lambda_0(t)$ which has to be estimated.

4.1. Notation and asymptotic distribution theory

Let $\{T_i,\delta_i,X_i,Z_i,W_i,\eta_i)$, $i=1,\ldots,n$, be i.i.d. random samples of the underlying random variables $\{T,\delta,X,Z,W,\eta\}$. Let $N_i(t)=I(\delta_i=1,T_i\le t)$ be the failure time counting process, and let

$$M_i(t)=N_i(t)-\int_0^t Y_i(u)R_i(\beta,u,\mathcal{A}_u,\Gamma_u)\lambda_0(u)du$$

be the associated counting process martingale. Let $\mathcal{T}_i(t)$ be a $p\times ap$ matrix such that the elements of the jth row are zeros except that the $\{(j-1)a+1,\ldots,ja\}$th elements are equal to $\partial\mu_{(j)}\{Z(t),W(t),\mathcal{A}_t\}/(\partial\alpha_{ju})$, where $\mu_{(j)}(Z,W,\mathcal{A}_t)$ is the jth element of $\mu\{Z(t),W(t),\mathcal{A}_t\}$. Let $\mathcal{V}_i(t)$ be a $p^2\times bp^2$ matrix such that the elements of the (lj)th row $(l,j=1,\ldots,p)$ are zeros except that the $\{(lj-1)b+1,\ldots,ljb\}$th elements are equal to $\partial\sigma^2_{(l,j)}\{Z(t),W(t),\Gamma_t\}/(\partial\Gamma_{jlt})$, where $\sigma^2_{(l,j)}\{Z(t),W(t),\Gamma_t\}$ is the (l,j)th element of $\sigma^2\{Z(t),W(t),\Gamma_t\}$. For a vector a, we write aa' as $a^{\otimes 2}$. Define

$$\begin{aligned}
&s^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)=E\{Y(t)R(\beta,t,\mathcal{A}_t,\Gamma_t)\};\\
&s^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)=E\{Y(t)R^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)\};\\
&s^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)=E\{Y(t)R^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)\},\\
&R^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)=\frac{\partial^2}{\partial\beta^2}R_i(\beta,t,\mathcal{A}_t,\Gamma_t);\\
&e_t=X(t)-\mu\{Z(t),W(t),\mathcal{A}_t)\},\quad \widehat{e}_t=X(t)-\mu\{Z(t),W(t),\widehat{\mathcal{A}}_t\};\\
&F(e_t)=\operatorname{diag}\Big[\operatorname{diag}\{e_{t(l)}e_{t(j)}-\widehat{e}_{t(l)}\widehat{e}_{t(j)},\ l=1,\ldots,k\},j=1,\ldots,k\Big];\\
&C_t=E\Big[(1-\eta)Y(t)\Big\{\frac{R^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R(\beta,t,\mathcal{A}_t,\Gamma_t)}-\frac{S^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}\\
&\qquad R(\beta,t,\mathcal{A}_t,\Gamma_t)\beta_1'\mathcal{T}(t)\Big];\qquad(11)\\
&D_t=E\Big[(1-\eta)Y(t)\Big\{\frac{R^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R(\beta,t,\mathcal{A}_t,\Gamma_t)}-\frac{s^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{s^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}R(\beta,t,\mathcal{A}_t,\Gamma_t)\\
&\qquad\times(1/2)\{\operatorname{vec}(\beta_1\beta_1')\}'\mathcal{V}(t)\Big];\qquad(12)
\end{aligned}$$

$$G(\beta) = \int_0^1 \Big[s^{(2)}(\beta, u, \mathcal{A}_u, \Gamma_u) - \frac{s^{(1)}(\beta, u, \mathcal{A}, \Gamma)^{\otimes 2}}{s^{(0)}(\beta, u, \mathcal{A}, \Gamma)} \Big] \lambda_0(u) du; \tag{13}$$

$$\begin{aligned} V(\beta) = E\Big[\int_0^1 \Big\{ \frac{R^{(1)}(\beta, u, \mathcal{A}_u, \Gamma_u)}{R(\beta, u, \mathcal{A}_u, \Gamma_u)} - \frac{s^{(1)}(\beta, u, \mathcal{A}_u, \Gamma_u)}{s^{(0)}(\beta, u, \mathcal{A}_u, \Gamma_u)} \Big\} dM(u) \\ - \eta \int_0^1 Y(u) C_u Q_u^{-1} \mathrm{vec}\{\mathcal{X}(u) \mathrm{diag}(e_u)\} \lambda_0(u) du \\ - \eta \int_0^1 Y(u) D_u H_u^{-1} \mathrm{vec}\{\mathcal{Y}(u) F(e_u)\} \lambda_0(u) du \Big]^{\otimes 2}. \end{aligned} \tag{14}$$

The following conditions are assumed to establish the large sample theory.

(A1) Given $\{Z(t), W(t)\}$, $t \in [0,1]$, the induced hazard function $\lambda(t; Z, W)$ is independent of the censorship.

(A2) The baseline hazard satisfies $\int_0^1 \lambda_0(t) dt < \infty$.

(A3) $\mathrm{pr}\{Y(t) = 1 | Z(t), W(t)\} > 0$ for any $\{Z(t), W(t)\}$ in the support of the density of (Z, W) and $t \in [0,1]$.

(A4) There exists an open $\mathcal{B}$ containing the true β, such that $R^{(2)}(\beta, t, \mathcal{A}_t, \Gamma_t)$ exists and is continuous on $\mathcal{B}$ and $t \in [0,1]$. Furthermore, $G(\beta)$ is positive definite.

(A5) For any $j \geq 3$, $E\left([X(t) - E\{X(t)|Z(t), W(t), T \geq t\}]^j | Z(t), W(t), T \geq t\right)$ is independent of $\{Z(t), W(t)\}$.

(A6) For any $t \in [0,1]$, $E\{X(t)|Z(t), W(t), Y(t)=1\}$ satisfies (6) and $\mathrm{var}\{X(t)|Z(t), W(t), Y(t)=1\}$ satisfies (7). Furthermore, $\widehat{\mathcal{A}}_t$ and $\widehat{\Gamma}_t$ satisfy (8) and (9), respectively.

(A7) The selection probability of the validation set does not depend on (T, δ) (though it may depend on the observed covariates (Z, W), and it may be time-dependent).

Proposition 1. *Under Conditions* (A1)-(A7)*, if the ARR estimator solving* (10) *is* $\widehat{\beta}$*, then* $n^{1/2}(\widehat{\beta} - \beta)$ *is asymptotically normally distributed with mean zero and asymptotic covariance* $G^{-1}(\beta) V(\beta) \{G^{-1}(\beta)\}'$.

The proof is given in the Appendix. Condition (A1) asserts that, conditional on the observed data, the subjects censored at time t can be regarded as representatives of the subjects that survive up to t and are still at risk. Conditions (A2) and (A3) ensure that when the sample size is large the relative risk can be

approximated in the risk set at any time $t \in [0,1]$. The unbiasedness of the estimating equation (10) is easily seen because it is obtained from the induced partial likelihood. By (A4), the consistency can be shown. We note that this presentation of consistency is restricted by Conditions (A5) and (A6). The asymptotic normality of $\widehat{\beta}$ is typically retained under departure from these assumptions, but asymptotic bias may result. One sufficient condition for (A5) and (A6) is that $\beta_1 = 0$ and any rth moment, $r \geq 3$, of $X|(Z,W)$ is independent of (Z,W); this includes a normal distribution with possibly a heteroscedastic variance function.

4.2. Robust covariance estimation

Let $\mathcal{A} = \{\mathcal{A}_{T_i}, i = 1,\ldots,n\}$, $\Gamma = \{\Gamma_{T_i}, i = 1,\ldots,n\}$. To estimate the covariance we note that under Conditions (A1)-(A7), $G(\beta)$ can be consistently estimated by $G_n(\widehat{\beta}, \widehat{\mathcal{A}}, \widehat{\Gamma})$, where

$$G_n(\beta,\mathcal{A},\Gamma) = -n^{-1}\sum_{i=1}^{n}\Big[\frac{R_i^{(2)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}{R_i(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})} - \Big\{\frac{R_i^{(1)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}{R_i(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}\Big\}^{\otimes 2} - \frac{S^{(2)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}{S^{(0)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})} + \Big\{\frac{S^{(1)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}{S^{(0)}(\beta,T_i,\mathcal{A}_{T_i},\Gamma_{T_i})}\Big\}^{\otimes 2}\Big]. \tag{15}$$

Denote the left side of (10) by $U_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$. To estimate $V(\beta)$, it can be shown, as in the Appendix, that

$$\begin{aligned} &U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma}) \\ &= n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Big\{\frac{R_i^{(1)}(\beta,u,\mathcal{A}_u,\Gamma_u)}{R_i(\beta,u,\mathcal{A}_u,\Gamma_u)} - \frac{S^{(1)}(\beta,u,\mathcal{A}_u,\Gamma_u)}{S^{(0)}(\beta,u,\mathcal{A}_u,\Gamma_u)}\Big\} \\ &\{dN_i(u) - Y_i(u)R_i(\beta,u,\mathcal{A}_u,\Gamma_u)\lambda_0(u)du\} \\ &-n^{-1/2}\sum_{i=1}^{n}\int_0^1 \eta_i(u)Y_i(u)C_uQ_u^{-1}\mathrm{vec}\{\mathcal{X}_i(u)\mathrm{diag}(e_{iu})\}\lambda_0(u)du \\ &-n^{-1/2}\sum_{i=1}^{n}\int_0^1 \eta_i(u)Y_i(u)D_uH_u^{-1}\mathrm{vec}\{\mathcal{Y}_i(u)F(e_{iu})\}\lambda_0(u)du + o_p(1). \end{aligned} \tag{16}$$

Write (16) as $U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma}) = n^{-1/2}\sum_{i=1}^{n}\phi_i(\beta,\mathcal{A}_{T_i},\Gamma_{T_i}) + o_p(1)$, which is a linearization as independent terms. Let $V_n(\beta,\mathcal{A},\Gamma) = n^{-1}\sum_{i=1}^{n}\phi_i(\beta,\mathcal{A}_{T_i},\Gamma_{T_i})\phi_i'(\beta,\mathcal{A}_{T_i},\Gamma_{T_i})$. Note that $V_n(\beta,\mathcal{A},\Gamma) \to V(\beta)$ in probability. To estimate $\phi_i(\beta,\mathcal{A}_{T_i},\Gamma_{T_i})$, we note that both Q_u and H_u can be obtained by the related sample averages in the risk set of the validation set. For example, if X is scalar and $\mu(Z,W,\mathcal{A}_t)$ is linear in (Z,W), then $\widehat{Q}_u = n^{-1}\sum_{i=1}^{n}\eta_iY_i(u)\{1,Z_i'(u),W_i'(u)\}'\{1,Z_i'(u),W_i'(u)\}$. To estimate C_u and D_u, we need only further estimate the baseline cumulative

hazard function by

$$d\widehat{\Lambda}(t) = \sum_{i=1}^{n} \Big\{ \frac{1}{nS^{(0)}(\beta, t, \widehat{\mathcal{A}}_t, \widehat{\Gamma}_t)} \Big\} dN_i(t).$$

5. Simulation Study

To understand the moderate sample size performance of the proposed ARR estimator, we present results from a simulation study.

5.1. Univariate covariate with surrogate

We generated data from a hazard function given by $\lambda(t; X) = \lambda_0 \exp(\beta X)$, where $\lambda_0 \equiv 1$ and X is a time-independent covariate from a standard normal distribution. We consider continuous surrogate variables such that $W_i = X_i + \sigma e_i$, where e_i is a standard normal random variable and independent of X_i, and here σ controls the magnitude of the measurement error. We compare the following estimators:

- Complete case (CC) estimator: This is a complete-case estimator which applies a usual Cox regression estimation procedure to the validation set.
- Regression calibration (RC) estimator: This estimator applies time-indepen -dent $\mathcal{A}$'s to estimate missing X values and then solves a usual partial likelihood score equation with the observed X in the validation set and estimated X in the non-validation set. Here we model $E(X|W)$ by $\alpha_0 + \alpha_1 W$, and hence $\mathcal{A} = (\alpha_0, \alpha_1)'$ is estimated by the LS estimator in the validation set.
- Approximate relative risk (ARR) estimator: The proposed estimator of this paper which solves (10). At each risk set, we model $E(X|W, Y(t) = 1)$ by $\alpha_{0t} + \alpha_{1t} W$, $\mathrm{var}\{X|W, Y(t) = 1\}$ by $\gamma_{0t} + \gamma_{1t} W + \gamma_{2t} W^2$, and we restrict the estimator $\widehat{\sigma}^2(W, \Gamma_t)$ to be positive. A Newton-Raphson algorithm is applied to solve the estimating equation.

Under this data generating device, we applied the LS estimates for $\mathcal{A}_t$ and Γ_t which satisfy (8) and (9), such that $\mathcal{X}_i = (1, W_i)'$, $Q_t = \mathrm{plim}_{n\to\infty}\{n^{-1}\sum_{i=1}^{n} \eta_i Y_i(t) \mathcal{X}_i \mathcal{X}_i'\}$; $\mathcal{Y}_i = (1, W_i, W_i^2)'$, $H_t = \mathrm{plim}_{n\to\infty}\{n^{-1}\sum_{i=1}^{n} \eta_i Y_i(t) \mathcal{Y}_i \mathcal{Y}_i'\}$. The corresponding $\mathcal{T}$ in (11) and $\mathcal{V}$ in (12) are $\mathcal{T}_i = (1, W_i)$ and $\mathcal{V}_i = (1, W_i, W_i^2)$. A total of 500 replicates was generated in each simulation configuration. In the following tables, "bias" means the average of $\widehat{\beta} - \beta$, where $\widehat{\beta}$ is an estimator of β. The "SD" denotes the square root of the sample variance of the 500 estimates of $\widehat{\beta}$.

At each risk set, we estimate $\mathcal{A}_t$ and Γ_t by the least squares estimates using data $\{X(t), W(t)\}$ from the validation set. In order to have reasonable calibration, we need enough data points for $\{X(t), W(t)\}$. This is to fulfill Conditions (A2) and (A3) of Section 4.1 so that $\mathcal{A}_t$ and Γ_t can be estimated satisfactorily. Therefore, we exclude the estimated score for subject i from the estimating equation (10) if the number of validation subjects in the risk set is less than six. Alternatively, one could retain the contribution of a subject i to the score in (10), but not update $\widehat{\mathcal{A}}_t$ and $\widehat{\Gamma}_t$, if the risk set size falls below a certain value.

In Table 1, we consider the relative risk parameter $\beta = \ln(2)$ and $\beta = \ln(4)$, respectively, to represent moderate and large relative risks. We consider sample sizes $n = 100$ and 300 with 50% missing X; measurement error magnitude $\sigma = 0.5$ or 1, and the censoring percentage $\tau = 0.25, 0.5$, or 0.75. It can be seen that when $\beta = \ln(2)$ the RC estimator works very well with limited biases and high relative efficiency, even when the measurement error is large. However, the bias problem of the RC estimator becomes severe when $\beta = \ln(4)$. Observe that the bias problem of the RC estimator and the ARR estimator is an increasing function of the measurement error σ and a decreasing function of the censoring percentage τ. It is seen that the ARR estimator has the smaller bias. The bias reduction arises from the greater flexibility in approximating the induced relative risk which in turn involves additional nuisance parameter estimation. Observe that, although X given W is normal, X given $\{W, Y(t) = 1\}$ is not. In this case, (A.5) and (A.6) do not hold since $\beta \neq 0$. Large β will increase the bias, but that from the ARR estimator can be seen to be much less than that from the RC under these simulation conditions.

5.2. Bivariate covariate without surrogate

Now consider bivariate covariates (X, Z). Assume that the hazard function is $\lambda(t; X, Z) = \exp(\beta_1 X + \beta_2 Z)$. Covariate X and Z are both from a uniform $(-\sqrt{3}, \sqrt{3})$ distribution and $\mathrm{corr}(X, Z) = 0.25$. The censoring percentage is 50%. In addition to CC, RC, ARR, we also calculate the approximate partial-likelihood (APL) estimator of Lin and Ying (1993). The APL estimator solves

$$n^{-1/2} \sum_{i=1}^{n} \delta_i \begin{pmatrix} \eta_i & 0 \\ 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} X_i \\ Z_i \end{pmatrix} - \frac{S^{(1)}_{APL}(\beta, T_i)}{S^{(0)}_{APL}(\beta, T_i)} \right\} = 0,$$

where $S^{(m)}_{APL}(\beta, T_i) = n^{-1} \sum_{j=1}^{n} \eta_j I[T_j \geq T_i] \begin{pmatrix} X_j \\ Z_j \end{pmatrix}^m r\{\beta, X_j(T_i), Z_j(T_i)\}$, $m = 0, 1$. Now, we consider a slightly relaxed assumption on the validation data. Assume that the selection probability of the validation set follows the model that $\mathrm{pr}(\eta = 1 | Z, T) = \{1 + \exp(-\gamma_0 - \gamma_1 Z - \gamma_2 T)\}^{-1}$, where γ's are given

Table 1. Simulation with moderate and large relative risk parameters $\beta = \ln(2), \ln(4)$, respectively. n is the total sample size and τ is the censoring percentage. Surrogate $W = X + \sigma e$ where e is a standard normal random variable. Results were from 500 replicates.

					σ=0.5		$\sigma = 1$	
β	τ	n		CC	RC	ARR	RC	ARR
ln(2)	0.25	100	bias	0.005	−0.013	0.001	−0.029	−0.004
			SD	0.213	0.142	0.151	0.160	0.178
		300	bias	0.001	−0.014	−0.001	−0.031	−0.003
			SD	0.103	0.081	0.085	0.086	0.097
	0.50	100	bias	0.009	0.009	0.002	−0.020	0.002
			SD	0.261	0.166	0.173	0.184	0.205
		300	bias	0.005	−0.005	0.004	−0.017	0.003
			SD	0.122	0.091	0.094	0.096	0.106
	0.75	100	bias	0.005	0.000	0.008	−0.013	0.006
			SD	0.368	0.234	0.239	0.258	0.273
		300	bias	−0.012	0.002	0.007	−0.008	0.007
			SD	0.176	0.128	0.131	0.138	0.147
ln(4)	0.25	100	bias	0.019	−0.105	−0.031	−0.204	−0.060
			SD	0.270	0.179	0.208	0.194	0.245
		300	bias	0.004	−0.113	−0.037	−0.212	−0.042
			SD	0.131	0.105	0.117	0.112	0.138
	0.50	100	bias	0.027	−0.075	−0.020	−0.152	−0.040
			SD	0.317	0.202	0.226	0.215	0.269
		300	bias	0.010	−0.080	−0.028	−0.154	−0.031
			SD	0.152	0.118	0.127	0.120	0.149
	0.75	100	bias	0.047	−0.047	−0.007	−0.121	−0.025
			SD	0.466	0.263	0.285	0.278	0.343
		300	bias	0.028	−0.046	−0.013	−0.112	−0.013
			SD	0.198	0.147	0.158	0.147	0.187

Note: CC is the complete case estimator; RC is the regression calibration estimator and ARR is the proposed approximate relative risk estimator.

Table 2. Bivariate covariate without surrogate. Note that $\text{pr}(\eta = 1|Z, T) = \{1+\exp(-\gamma_0-\gamma_1 Z-\gamma_2 T)\}^{-1}$. The censoring percentage was 50%. Covariate X and Z were from a uniform $(-\sqrt{3}, \sqrt{3})$ distribution such that $\text{corr}(X, Z) = 0.25$.

			β_1=ln(2)				β_2=ln(2)			
γ	n		CC	APL	RC	ARR	CC	APL	RC	ARR
(0,0,0)	200	bias	0.013	0.014	−0.016	0.000	0.008	0.023	−0.041	0.013
		SD	0.180	0.181	0.163	0.173	0.167	0.172	0.124	0.129
	400	bias	0.017	0.017	−0.017	0.009	0.005	0.012	−0.043	0.007
		SD	0.119	0.120	0.105	0.116	0.110	0.112	0.080	0.086
(−1,1,0)	200	bias	0.017	0.067	−0.017	−0.013	0.017	−0.674	−0.095	0.010
		SD	0.206	0.211	0.186	0.189	0.246	0.240	0.146	0.158
	400	bias	0.020	0.061	−0.018	0.000	0.014	−0.659	−0.097	−0.005
		SD	0.143	0.142	0.124	0.128	0.161	0.150	0.098	0.107
(−1,1,1)	200	bias	0.047	0.079	−0.053	−0.023	0.095	−0.532	−0.068	0.011
		SD	0.191	0.188	0.152	0.178	0.227	0.205	0.133	0.142
	400	bias	0.041	0.073	−0.058	−0.030	0.097	−0.522	−0.071	0.004
		SD	0.130	0.124	0.102	0.112	0.151	0.132	0.087	0.097
			β_1=ln(3)				β_2=ln(3)			
γ	n		CC	APL	RC	ARR	CC	APL	RC	ARR
(0,0,0)	200	bias	0.012	0.021	−0.097	0.012	0.001	0.032	−0.180	−0.007
		SD	0.197	0.204	0.158	0.193	0.187	0.206	0.139	0.160
	400	bias	0.030	0.033	−0.091	0.019	0.013	0.024	−0.176	−0.003
		SD	0.135	0.137	0.104	0.134	0.132	0.136	0.089	0.111
(−1,1,0)	200	bias	0.024	−0.051	−0.109	−0.022	0.013	−0.672	−0.275	0.011
		SD	0.235	0.210	0.179	0.211	0.271	0.251	0.165	0.206
	400	bias	0.037	−0.047	−0.105	0.001	0.030	−0.650	−0.272	0.021
		SD	0.159	0.139	0.119	0.143	0.182	0.160	0.111	0.147
(−1,1,1)	200	bias	0.063	0.019	−0.145	−0.028	0.098	−0.533	−0.227	0.021
		SD	0.219	0.190	0.151	0.190	0.252	0.223	0.151	0.180
	400	bias	0.075	−0.001	−0.142	−0.030	0.118	−0.514	−0.226	0.003
		SD	0.132	0.130	0.104	0.129	0.170	0.143	0.097	0.121

in Table 2 and T is the observed failure time. When $\gamma = (0, 0, 0)$, the missingness is completely at random and the validation sampling rate is 50%. When $\gamma = (-1, 1, 0)$, the missingness depends also on Z; this is the case when the validation set is stratified on Z. There were 69 % missing X values. We also consider the case when the missingness depends on the observed failure time, with $\gamma = (-1, 1, 1)$, leading to 61% missing X's. We consider the case when $\beta = \{\ln(2), \ln(2)\}'$ and $\beta = \{\ln(3), \ln(3)\}'$, respectively.

Results from Table 2 show that the ARR estimator is quite satisfactory for both $\beta_1 = \ln(2)$ and $\beta_1 = \ln(3)$, while the RC estimator has a bias problem especially when $\beta_1 = \ln(3)$. Although the APL estimator is consistent when the missingness is completely at random, it has a serious bias problem when the missingness depends on Z or (Z, T). The CC analysis has a bias problem if the missingness depends on the observed failure time, but the bias is minimal for small β. While not reported in Table 2, our simulations indicate that the bias problem of the RC estimator is primarily due to increasing β_1, although increasing β_2 also has some effect.

5.3. Sandwich and Bootstrap covariance estimation

In Table 3 we evaluate the robust sandwich estimation for the variance of the ARR estimator, given in Section 4.2, and a bootstrap procedure. Bootstrap covariance estimation is an attractive alternative since the asymptotic covariance formula may be hard to program. We consider resampling $B = 30$ and $B = 60$ times, each bootstrap sample consisting of a random sample of size n from $\{N_i, Y_i, X_i, Z_i, \eta_i\}, i = 1, \ldots, n$. Data were generated similar to the setting of Table 1, but with $\beta = \ln(3)$, here, we consider the case when the censoring percentage was 50%. We consider two missingness mechanisms: (i) the missingness is completely at random with 50% missing X's; (ii) the missingness depends on the observed failure time with $\text{pr}(\eta = 1|W, T) = \{1 + \exp(2 - 3T)\}^{-1}$, leading to 64% missing X's. In the table, "mean(SE)" denotes the average of the 200 standard error estimates. We also calculate the 95% coverage probabilities.

It is seen that the sandwich estimator and the bootstrap estimator perform equally well in most cases, and the coverage probabilities are close to the nominal value. The CC analysis has a bias problem when the missingness depends on the observed failure time, although the coverage probability for $n = 100$ is still good. However, the coverage probability for the CC analysis is about 84% when we increase the sample size to $n = 400$. From our computations, it appears that calculating the sandwich formula of the standard error of the ARR estimator takes about the same time as bootstrapping 30 resamples. Therefore, a bootstrap procedure is a comparatively practical approach to estimating the standard error of the ARR estimator.

Table 3. Simulation study on covariance estimation with $\beta = \ln(3)$, n is the total sample size. The censoring percentage was 50%. Surrogate $W = X + .5e$ where e is a standard normal random variable.

n		CC	ARR	ARR bootstrap	
				$B = 30$	$B = 60$
$\text{pr}(\eta = 1 \mid W, T) = 0.5$					
100	bias	0.048	0.012	—	—
	SD	0.292	0.206	—	—
	mean(SE)	0.268	0.196	0.211	0.211
	95 % cov. prob.	0.946	0.932	0.932	0.938
200	bias	0.028	0.000	—	—
	SD	0.183	0.141	—	—
	mean(SE)	0.180	0.137	0.143	0.145
	95 % cov. prob.	0.958	0.934	0.930	0.940
$\text{pr}(\eta = 1 \mid W, T) = \{1 + \exp(2 - 3T)\}^{-1}$					
100	bias	0.210	0.035	—	—
	SD	0.470	0.228	—	—
	mean(SE)	0.445	0.225	0.245	0.245
	95 % cov. prob.	0.970	0.968	0.964	0.956
200	bias	0.181	0.020	—	—
	SD	0.304	0.160	—	—
	mean(SE)	0.290	0.155	0.160	0.158
	95 % cov. prob.	0.928	0.934	0.934	0.950

Note: For $\widehat{\beta}$ being the CC or the ARR estimator, "bias" denotes the average of $\widehat{\beta} - \beta$, "SD" denotes the square root of the sample variance of the 200 estimates. We used bootstrap only to estimate standard error and hence "bias" and "SD" for the last two columns were filled with "—".

6. Illustration

We consider an example from the Studies of Left Ventricular Dysfunction (SOLVD, 1991). The failure time is the time from randomization to the trial to death and the covariate of interest is the left ventricular ejection fraction (EF). Subjects were patients with a diagnosis of congestive heart failure. We

consider n = 308 subjects of the study in our illustration. Among them there were $n_v = 162$ subjects in the validation set where they had their EFs measured from a radionucleotide technique. A related echocardiographic measure (W) for EF was ascertained for all subjects of the study. Among these 308 subjects, 93 deaths were observed during the follow-up of the trial. Values of X and W were standardized to have mean zero and variance one.

We first examined the missingness mechanism of the data by modeling the selection probabilities of X_i (EF). We ran a logistic regression analysis with outcome η_i, covariates (T_i, W_i), $i = 1, \ldots, n$. The parameter estimate for T_i was 0.125 (SE 0.120), and that for W_i was -0.0012 (SE 0.0003). The selection probabilities depend strongly on W but not significantly on the observed failure time. Note that in this case, the CC analysis is still consistent but inefficient. The consistency of the CC analysis under this missingness mechanism can be easily shown since the hazard function given (Z, X) is the same as that given $(Z, X, \eta = 1)$. In Table 4, results from the CC, RC, ARR estimators are presented. Note that when we fitted a linear model of X_i on W_i, the standard deviation of the error was about 0.5 which indicates a moderate measurement error. Similar to the findings from the upper portion of Table 1, the results from RC and ARR are similar because both the relative risk parameter (β) and the measurement error variance are moderate in size. All the estimates indicate that reduced left ventricular EFs increases the risk of mortality. The relative risk estimate from CC is smaller than the others perhaps because of the moderate dependence of the selection probabilities on the observed failure time. The main difference in the statistical inference is that the CC analysis does not indicate a significant relationship between EF and mortality while the other estimators do.

Table 4. Analyses of the risk of reduced ventricular ejection fractions and heart failure mortality.

	CC	RC	ARR
ER	-0.210	-0.310	-0.274
(SE)	(0.143)	(0.131)	(0.130)

7. Recalibration Based on A Simple ARR Algorithm

The proposed ARR estimator can be approximated by the following algorithm, which has a recalibration explanation. For simplicity, consider X as a

univariate continuous variable. In this case, $\mathcal{A}_t = \alpha_t$ and $\Gamma_t = \gamma_t$ are both vectors. We note that

$$\begin{aligned}&\exp\left[\beta_1\mu\{Z(t),W(t),\alpha_t\} + (1/2)\beta_1^2\sigma^2\{Z(t),W(t),\gamma_t\}\right]\\ &= \exp\left(\beta_1\left[\mu\{Z(t),W(t),\alpha_t\} + (1/2)\beta_1\sigma^2\{Z(t),W(t),\gamma_t\}\right]\right).\end{aligned}$$

Therefore, we consider a simple ARR algorithm below.

(i) Obtain an initial estimator, say the complete case estimator or the ordinary regression calibration estimator. Denote the estimate of β_1 by $\tilde{\beta}_1$.

(ii) In the validation set, for each uncensored failure time t, we model functions μ and σ^2 and then obtain $\widehat{\alpha}_t$, $\widehat{\gamma}_t$. For each missing $X(t)$ value, replace it by $X^*(t) = \mu\{Z(t),W(t),\widehat{\alpha}_t\} + (1/2)\tilde{\beta}_1\sigma^2\{Z(t),W(t),\widehat{\gamma}_t\}$.

(iii) Apply a Cox regression procedure using the true covariate value $X(t)$ in the validation sample, and the replacement value $X^*(t)$ otherwise.

Table 5. Difference of ARR and Simple ARR algorithm estimators. n is the total sample size and τ is the censoring percentage. Surrogate $W = X + \sigma e$ where e is a standard normal random variable.

			σ=0.5			$\sigma = 1$		
			β			β		
τ	n		ln(2)	ln(4)	ln(6)	ln(2)	ln(4)	ln(6)
0.2	100	mean	0.000	0.000	0.002	−0.002	−0.007	−0.012
		SD	0.009	0.025	0.042	0.024	0.070	0.123
	300	mean	0.000	0.000	0.003	0.000	0.003	0.011
		SD	0.003	0.010	0.017	0.010	0.034	0.058
0.8	100	mean	0.001	0.004	0.012	0.003	0.009	0.018
		SD	0.012	0.034	0.065	0.046	0.122	0.204
	300	mean	0.000	0.001	0.004	0.000	0.005	0.010
		SD	0.006	0.016	0.027	0.019	0.057	0.091

As will be clear later, there is no need to iterate procedures (i)-(iii) since little improvement can be seen. The simple ARR algorithm described above is slightly different from the ARR estimator solving (10): it uses X_i^* which is not exactly the same as $R_i^{(1)}(\beta,t,\alpha,\gamma)/R_i(\beta,t,\alpha,\gamma)$ when $\eta_i = 0$; and it ignores

the use of $c(t)$ in $S^{(0)}(\beta, t, \alpha, \gamma)$ and $S^{(1)}(\beta, t, \alpha, \gamma)$. As a result, it may have a slightly larger bias under some extreme cases. Table 5 presents the mean and the standard deviation of the differences (from 500 replicates) between them. Data were generated similar to those of Table 1, with 50% missing X and we consider various parameters. It can be seen that they have very small differences, and if the relative risk relationship is not extreme then the difference is trivial. It is easier to implement this simple ARR approximation, as compared to solving (10). To estimate the standard error, we may use the bootstrap procedure described earlier.

8. Discussion

In Cox regression the problem of the regression parameter estimation with missing covariates is of considerable practical importance. However, this problem becomes more complicated if the observable covariate or surrogate variables are continuous and the relative risk parameter for X is large. For example, the implementation of an EM-type maximum likelihood estimator (Chen and Little (1999)) is not straightforward and will involve the estimation of the cumulative baseline hazard function. In this case, the RC estimator is easy to implement, but may be substantially biased in extreme circumstances. Here, we have proposed a second order approximation of the induced relative risk which evidently has less bias in estimating β. Our method utilizes the first two conditional moments, given observed covariate/surrogate variables in the risk set, and can be considered as a higher order approximation than the usual RC estimator. From the viewpoint of methodology, the proposed estimator is a refined RC estimator. This can be treated as an approximation of the maximum likelihood estimator (Chen and Little (1999)), which requires the model assumption of X given (Z, W). Although bias may arise from this approximation under some extreme cases, it does not need either numerical integration or estimation of the cumulative hazard function.

The consistency result of the estimator is based on Conditions (A5) and (A6) which do not hold in general and hence more generally $\widehat{\beta}$ will be consistent for a β_* that solves $E\{U_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})\} = 0$. Indeed, the bias of the ARR estimator may be appreciable when β_1 is very large. For example, under the same setting as Tables 2 and 3, when $\beta_1 = \ln(6)$, biases of the ARR estimator may be larger than 0.1 with unsatisfactory coverage probabilities. Nevertheless, this value represents an extreme relative risk relationship. Even so, approximation using the first two conditional moments could be better than approximation using higher (> 2) conditional moments since the approximation has to be done in all risk sets, even risk sets of moderate size. We note that, to avoid a finite sample problem in estimating nuisance parameters $\mathcal{A}_t$ and Γ_t, the estimating score for subject i

should be excluded from the estimating equation (10) if the size of the validation subjects in the risk set is small (or recalibration can be discontinued once the risk set size falls below a specific value). Our findings in the simulation study indicate that if $X(t)$ given $\{Z(t), W(t)\}$ is approximately linear, then applying the linear conditional mean and the quadratic conditional variance of $\{Z(t), W(t)\}$ to estimate the induced relative risk will in general lead to good estimation of β.

Estimation of the variance of X given $(Z, W, T \geq t)$ could be an immense task when X is not univariate. For univariate X, our unreported simulation study indicates that the proposed estimator works well for bivariate (Z, W) as long as we exclude a subject from the estimating equation when the size of the risk set is less than 10. It is true that when β_1 is larger than ln(6), the bias of the ARR is not negligible, but the relative bias is still small. One also should note that the larger SE of the ARR estimator (than say that from the RC) is not due to the estimation of the variance functions. This is indeed a phenomenon in measurement error problems in nonlinear regression: reducing bias comes at the cost of increasing SE if $|\beta_1|$ is large. To illustrate this phenomenon, consider binary W and univariate X, where the EPL estimator of Zhou and Pepe (1995) is consistent. When β_1 is not large, EPL is similar to RC and they are both better than CC. The APL estimator of Lin and Ying is the same as CC. However, when β_1 is as large as ln(4), the EPL could even be less efficient than the CC if the correlation of X and W is weak and the censoring percentage is low.

Acknowledgements

The research was supported by US National Institutes of Health grants CA 53996 (Wang, Prentice), AG 15026 (Wang). The authors are grateful to the SOLVD investigators and Haibo Zhou for access to study data, and to an anonymous referee for helpful comments.

Appendix: Technical Proofs

We assume that $(N_i, Y_i, X_i, Z_i, \eta_i)$ are identically independently distributed and the regularity conditions (A1)-(A7) have been made. Let $G_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) = -n^{-1/2}(\partial/\partial\beta)U_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$, which can be easily seen to be (15). By direct calculations, $G_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) = -\mathcal{M}(\beta, 1) - \mathcal{A}(\beta, 1) + o_p(1)$, where

$$\mathcal{M}(\beta,1) = n^{-1}\sum_{i=1}^{n}\int \Big[\frac{R_i^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R_i(\beta,t,\mathcal{A}_t,\Gamma_t)} - \Big\{\frac{R_i^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R_i(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}^{\otimes 2} - \frac{S^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)} + \Big\{\frac{S^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}^{\otimes 2}\Big] dM_i(t),$$

$$\mathcal{A}(\beta,1)=n^{-1}\sum_{i=1}^{n}\int\Big[\frac{R_i^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R_i(\beta,t,\mathcal{A}_t,\Gamma_t)}-\Big\{\frac{R_i^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R_i(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}^{\otimes 2}$$
$$-\frac{S^{(2)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}+\Big\{\frac{S^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}\Big\}^{\otimes 2}\Big]$$
$$Y_i(t)R_i(\beta,t,\mathcal{A}_t,\Gamma_t)\lambda_0(t)dt.$$

It can be shown that $\mathcal{M}(\beta,1)$ converges to 0 in probability since it is a local square integrable martingale with variance process converging to 0. Also by simple algebra it can be shown that $\mathcal{A}(\beta,1)$ converges to $G(\beta)$, which was defined in (13). Therefore, we have shown that $n^{-1/2}(\partial/\partial\beta)U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})$ converges to $G(\beta)$ in probability. By the Inverse Function Theorem (Rudin (1964)), or theory on M-estimates (Huber (1981), Chapter 3), by Condition (A4) it can be shown that $\widehat{\beta}\to\beta$ in probability.

We now derive the asymptotic distribution of $n^{1/2}(\widehat{\beta}-\beta)$. By a Taylor expansion of $U_n(\widehat{\beta},\widehat{\mathcal{A}},\widehat{\Gamma})$, we have that $0=U_n(\widehat{\beta},\widehat{\mathcal{A}},\widehat{\Gamma})=U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})-G_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})n^{1/2}(\widehat{\beta}-\beta)+o_p(1)$. Hence,

$$n^{1/2}(\widehat{\beta}-\beta)=G_n^{-1}(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})+o_p(1). \tag{17}$$

We first linearize $U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})$ into a sum of independent random variables. Let

$$L_i(\beta,t,\mathcal{A}_t,\Gamma_t)=\frac{R_i^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{R_i(\beta,t,\mathcal{A}_t,\Gamma_t)}-\frac{S^{(1)}(\beta,t,\mathcal{A}_t,\Gamma_t)}{S^{(0)}(\beta,t,\mathcal{A}_t,\Gamma_t)}.$$

Then $U_n(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})=U_{1n}(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})+U_{2n}(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})$, where

$$U_{1n}(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})=n^{-1/2}\sum_{i=1}^{n}\int_0^1 L_i(\beta,u,\widehat{\mathcal{A}}_u,\widehat{\Gamma}_u)dM_i(u);$$
$$U_{2n}(\beta,\widehat{\mathcal{A}},\widehat{\Gamma})=n^{-1/2}\sum_{i=1}^{n}\int_0^1 L_i(\beta,u,\widehat{\mathcal{A}}_u,\widehat{\Gamma}_u)R_i(\beta,u,\mathcal{A}_u,\Gamma_u)Y_i(u)\lambda_0(u)du.$$

By noting that each element of $\widehat{\mathcal{A}}_t$ is a linear summation of X_i values in the risk set of the validation set, it can be shown that $\sup_{t\in[0,1]}\parallel\mathrm{vec}(\widehat{\mathcal{A}}_t-\mathcal{A}_t)\parallel\to 0$ *a.s.* (Andersen and Gill (1982), Theorem 3.1). Similarly, $\sup_{t\in[0,1]}\parallel\mathrm{vec}(\widehat{\Gamma}_t-\Gamma_t)\parallel\to 0$ *a.s.* Therefore, by the Lenglart Inequality (Fleming and Harrington (1991), Chapter 4),

$$n^{-1/2}\sum_{i=1}^{n}\int_0^1\{L_i(\beta,u,\widehat{\Gamma}_u,\widehat{\Gamma}_u)-L_i(\beta,u,\mathcal{A},\Gamma)\}dM_i(u)\xrightarrow{p}0.$$

Hence,

$$U_{1n}(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) = n^{-1/2} \sum_{i=1}^{n} \int_0^1 L_i(\beta, u, \mathcal{A}_u, \Gamma_u) dM_i(u) + o_p(1)$$
$$= U_{1n}(\beta, \mathcal{A}, \Gamma) + o_p(1).$$

We now linearize $U_{2n}(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$. By a Taylor expansion as in Zhou and Pepe (1995) $(x/y = x_0/y_0 + (x - x_0)/y_0 - (y - y_0)x_0/y_0^2 + O\{(x - x_0)^2 + (y - y_0)^2\})$, following some calculations we have

$$U_{2n}(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$$
$$= -n^{-1/2} \sum_{i=1}^{n} \int_0^1 L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) \{R_i(\beta, u, \widehat{\mathcal{A}}_u, \widehat{\Gamma}_u) - R_i(\beta, u, \mathcal{A}_u, \Gamma_u)\}$$
$$\lambda_0(u) du + o_p(1).$$

By (4), we write $U_2(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) = A_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) + B_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) + o_p(1)$, where

$$A_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$$
$$= -n^{-1/2} \sum_{i=1}^{n} \int_0^1 \{1 - \eta_i(u)\} L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) \{\widehat{c}(u) - c(u)\}$$
$$r^*\{\beta, Z_i(u), W_i(u)\} \lambda_0(u) du,$$

which can be shown to be $o_p(1)$ by noting that $\sup_{t \in [0,1]} \parallel \widehat{c}(t) - c(t) \parallel \to 0$ *a.s.*, using the uniform convergence of $\widehat{\mathcal{A}}_t$ and $\widehat{\Gamma}_t$ above. Further,

$$B_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$$
$$= -n^{-1/2} \sum_{i=1}^{n} (1 - \eta_i) \int_0^1 L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) \Big(\exp[\beta_1' \mu\{Z_i(u), W_i(u), \widehat{\mathcal{A}}_u\}$$
$$+ (1/2)\beta_1' \sigma^2\{Z_i(u), W_i(u), \widehat{\Gamma}_u\}\beta_1] - \exp[\beta_1' \mu\{Z_i(u), W_i(u), \mathcal{A}_u\}$$
$$+ (1/2)\beta_1' \sigma^2\{Z_i(u), W_i(u), \Gamma_u\}\beta_1] \Big) e^{\beta_2' Z_i(u)} c(u) \lambda_0(u) du + o_p(1)$$
$$= -n^{-1/2} \sum_{i=1}^{n} (1 - \eta_i) \int_0^1 L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) r^*\{\beta, Z_i(u), W_i(u)\}$$
$$\times \Big[\beta_1' \mathcal{T}_i \mathrm{vec}(\widehat{\mathcal{A}}_u - \mathcal{A}_u) + (1/2)\{\mathrm{vec}(\beta_1 \beta_1')\}' \mathcal{V}_i(u) \mathrm{vec}(\widehat{\Gamma}_u - \Gamma_u) \Big]$$
$$c(u) \lambda_0(u) du + o_p(1).$$

The last equality was obtained from a Taylor expansion on $\exp[\beta_1' \mu\{Z_i(u), W_i(u), \widehat{\mathcal{A}}_u\} + (1/2)\beta_1' \sigma^2\{Z_i(u), W_i(u), \widehat{\Gamma}_u\}\beta_1]$. By (8) and (9),

$$U_{2n}(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$$

$$
\begin{aligned}
&= -\int_0^1 \Big\{ n^{-1} \sum_{i=1}^{n} \{1-\eta_i(u)\} L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) r^*\{\beta, Z_i(u), W_i(u)\} \beta_1' \mathcal{T}_i(u) \Big\} \\
&\qquad \times Q_u^{-1} n^{-1/2} \sum_{j=1}^{n} \eta_j(u) Y_j(u) \mathrm{vec}\{\mathcal{X}_j(u) \mathrm{diag}(e_{ju})\} c(u) \lambda_0(u) du \\
&\quad - \int_0^1 \Big\{ n^{-1} \sum_{i=1}^{n} \{1-\eta_i(u)\} L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) r^*\{\beta, Z_i(u), W_i(u)\} (1/2) \\
&\quad \{\mathrm{vec}(\beta_1 \beta_1')\}' \mathcal{V}_i(u) \Big\} \\
&\qquad \times H_u^{-1} n^{-1/2} \sum_{j=1}^{n} \eta_j(u) Y_j(u) \mathrm{vec}\{\mathcal{Y}_j(u) F(e_{ju})\} c(u) \lambda_0(u) du + o_p(1).
\end{aligned}
$$

Let

$$
\begin{aligned}
C_{un} &= n^{-1} \sum_{i=1}^{n} \{1-\eta_i(u)\} L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) c(u) r^*\{\beta, Z_i(u), W_i(u)\} \beta_1' \mathcal{T}_i(u); \\
D_{un} &= n^{-1} \sum_{i=1}^{n} \{1-\eta_i(u)\} L_i(\beta, u, \mathcal{A}_u, \Gamma_u) Y_i(u) c(u) r^*\{\beta, Z_i(u), W_i(u)\} (1/2) \\
&\qquad \{\mathrm{vec}(\beta_1 \beta_1')\}' \mathcal{V}_i(u).
\end{aligned}
$$

Then it can be easily seen that $C_{un} \to C_u$ in probability and $D_{un} \to D_u$ in probability, where C_u and D_u were defined in (11). Hence,

$$
\begin{aligned}
U_{2n}(\beta, \mathcal{A}, \Gamma) = &-n^{-1/2} \sum_{j=1}^{n} \int C_u Q_u^{-1} \eta_j(u) Y_j(u) \mathrm{vec}\{\mathcal{X}_j(u) \mathrm{diag}(e_{ju})\} \lambda_0(u) du \\
&-n^{-1/2} \sum_{j=1}^{n} \int D_u H_u^{-1} \eta_j(u) Y_j(u) \mathrm{vec}\{\mathcal{Y}_j(u) F(e_{ju})\} \lambda_0(u) du + o_p(1).
\end{aligned}
$$

As a result, we have shown $U_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma}) = \sum_{i=1}^{n} \phi_i(\beta, \mathcal{A}, \Gamma)$ as given in (16). By the Central Limit Theorem, $U_n(\beta, \widehat{\mathcal{A}}, \widehat{\Gamma})$ is asymptotically normally distributed with mean 0 and variance $V(\beta)$ given by (14). Finally, by (17), the proof of Proposition 1 is completed.

References

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

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

Chen, H. Y. and Little, R. J. A. (1999). Proportional hazards regression with missing covariates. *J. Amer. Statist. Assoc.* **94**, 896-908.

Clayton, D. G. (1991). Models for the analysis of cohort and case–control studies with inaccurately measured exposure. In: *Statistical Models for Longitudinal Studies of Health* (Dwyer, J. H., Feinleib, M., Lipsert, P., et al. eds.), pp. 301–333. Oxford University Press, New York.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

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

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, 4th ed., Vol 1. Hafner, New York.

Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *J. Amer. Statist. Assoc.* **88**, 1341-1349.

Paik, M. C. and Tsai, W. Y. (1997). On using the Cox proportional hazards model with missing covariates. *Biometrika* **84**, 579–593.

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

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846–866.

Rudin, W (1964). *Principles of Mathematical Analysis*. McGraw-Hill, New York.

The SOLVD investigators (1991). Effect of enalaprial on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *New England Journal of Medicine* **325**, 293–302.

Wang, C. Y. and H. Y. Chen (2001). Augmented inverse probability weighted estimator for Cox missing covariate regression. *Biometrics***57**, 414-419.

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

Xie, S. X., Wang, C. Y. and Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *J. Roy. Statist. Soc. Ser. B*, in press.

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

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A.

E-mail: cywang@fhcrc.org

E-mail: rprentic@fhcrc.org

Department of Health Evaluation Sciences, Penn Sylvania State University, College of Medicine, Hershey, PA 17033, U.S.A.

E-mail: sxie@psu.edu

(Received May 1999; accepted April 2001)