Regression Calibration in Semiparametric Accelerated Failure Time Models

Menggang Yu[1],[*] and Bin Nan[2]

[1]Department of Medicine, Division of Biostatistics, Indiana University School of Medicine,
410 West 10th Street, Suite 3000, Indianapolis, Indiana 46202, USA
[2]Department of Biostatistics, University of Michigan, Ann Arbor, 1420 Washington Heights,
Ann Arbor, Michigan 48109, USA
[*]*email:* meyu@iupui.edu

SUMMARY. In large cohort studies, it often happens that some covariates are expensive to measure and hence only measured on a validation set. On the other hand, relatively cheap but error-prone measurements of the covariates are available for all subjects. Regression calibration (RC) estimation method (Prentice, 1982, *Biometrika* **69**, 331–342) is a popular method for analyzing such data and has been applied to the Cox model by Wang et al. (1997, *Biometrics* **53**, 131–145) under normal measurement error and rare disease assumptions. In this article, we consider the RC estimation method for the semiparametric accelerated failure time model with covariates subject to measurement error. Asymptotic properties of the proposed method are investigated under a two-phase sampling scheme for validation data that are selected via stratified random sampling, resulting in neither independent nor identically distributed observations. We show that the estimates converge to some well-defined parameters. In particular, unbiased estimation is feasible under additive normal measurement error models for normal covariates and under Berkson error models. The proposed method performs well in finite-sample simulation studies. We also apply the proposed method to a depression mortality study.

KEY WORDS: Accelerated failure time model; Measurement error; Regression calibration; Two-phase sampling; Validation data.

1. Introduction

In large cohort studies for failure time regression analysis, we often measure many variables. Among them some are easy to obtain while others may be costly to measure. One such example is a depression mortality cohort study conducted in two phases at the Indiana University School of Medicine (Callahan et al., 1998). Phase 1 of the study took place from January 1991 to June 1993 and the whole cohort was screened for depression, dementia, and alcoholism. Phase 2 of the study was conducted following phase 1 on a subsample of the cohort. Patients were selected based on simple random sampling stratified on their depression score measured by the Center for Epidemiologic Studies Depression scale (CES-D). During the second phase, patients were required to complete an extended interview that consisted of the Hamilton Depression Rating Scale (HAM-D) and part of the Sickness Impact Profile (SIP). HAM-D offers higher validity and reliability than CES-D in measuring response to treatment of depression and SIP measures the effect depressive symptoms had on physical and psychosocial functioning of study subjects. Both require more time and effort to collect than CES-D. A question of interest is to determine whether patients' depression and quality of life level are associated with mortality.

This example can be viewed as a survival analysis problem with covariates subject to measurement error. HAM-D and SIP scores are better measurements to reflect the true morbidity status but are only available for a subset of the cohort (called a validation subset) while CES-D score was measured on all subjects and viewed as a surrogate for HAM-D and SIP. The mortality information was also available on all subjects. The question of interest is to develop an effective way to deal with measurement error or surrogate covariates in the survival analysis framework.

In survival analysis, Cox model is a popular choice. Direct application of the Cox model to the above data detected significant effects of CES-D score, gender, race, and age on survival. However, the proportional hazard assumption is in question. Specifically, the corresponding p-values from the tests of the proportional hazard assumption based on Schoenfeld residuals (Grambsch and Therneau, 1994) are 0.090, 0.0005, 0.187, 0.333, and 0.019 for the four covariates and the overall model, respectively. Because the CES-D scores are considered as surrogates, we also fit the Cox model using HAMD&SIP, gender, race, and age as covariates where HAMD&SIP is sum of the HAM-D and SIP and used as index for a subject's morbidity level. We assign same weights to HAM-D and SIP in coming up with the index due to their similar ranges. For the subjects who do not have HAM-D and SIP scores, we use imputed values via a linear regression (see Section 5.2 for more details). Under such setting, these

p-values from the tests of the proportional hazard assumption now become 0.133, 0.007, 0.213, 0.314, and 0.026. Still the proportional hazard assumption is violated.

This leads us to consider alternative models for analysis. Semiparametric accelerated failure time (AFT) model is a good alternative (Kalbfleisch and Prentice, 2002, Chapter 7). The linear structure of the AFT model and the nice interpretation due to its direct modeling of survival time is especially appealing to medical investigators. Moreover, in our initial model fitting of the same data set, we found that the semiparametric AFT model appears to provide a good fit of the data. The empirical p-values for goodness-of-fit check are 0.692, 0.104, 0.828, and 0.254 for (imputed) HAMD&SIP, gender, race, and age respectively, based on 500 simulated martingale residual score processes (Lin, Robins, and Wei, 1996; Peng and Fine, 2006). Figure 1 provides plots of 50 realizations from the distributions of the score processes (dashed lines) together with the observed score processes (solid lines). Hence, we consider dealing with measurement error data under the semiparametric AFT model framework.

By viewing measurement error as a missing data problem, the inverse-probability-weighted (IPW) method can be applied for the semiparametric AFT model to deal with covariates subject to measurement error (Yu and Nan, 2006; and Nan, Kalbfleisch, and Yu, 2009). However, the IPW method uses outcome or failure time data from the validation subset only. As we have failure time information for all subjects in the cohort, this can lead to some loss of information. Augmented inverse-probability-weighted (AIPW) estimator has been proposed in the Cox model for censored data (Wang and Chen, 2001; Qi, Wang, and Prentice, 2005) and uncensored data (e.g., Wang and Wang, 2001). Although the class of AIPW estimators includes the semiparametric efficient estimator (Robins, Rotnitzky, and Zhao, 1994), implementation can be extremely challenging and usually involves solving integral equations (Nan, 2004). As a popular alternative, the usual AIPW estimating equation augments IPW equation by conditioning on observed data to reduce computational burden. As observed by Qi et al. (2005, Theorem 2–4) and Wang and Wang (2001, Theorem 1 and 2), such AIPW estimators are asymptotically equivalent to IPW estimators with estimated weights. The improvement in efficiency may be limited.

Corrected score method was proposed for the Cox model by Nakamura (1992) and studied by many others (e.g., Hu and Lin, 2002; Huang and Wang, 2000). These correction methods rely on approximating average terms among risk sets using measurement error data. Such methods may be hard to apply to the semiparametric AFT model setting where risk sets are defined on the residual scale that depends on observed failure time, mismeasured or missing covariates, and the true value of regression coefficients. The residual is defined as the difference between the failure time and its regression mean. In contrast, in the Cox regression, risk sets are usually defined by observed failure time and hence not subject to the influence of mismeasured or missing covariates.

We have also investigated the simulation and extrapolation (SIMEX) method due to its conceptual and computational simplicity (Carroll et al., 2006, Chapter 5). However when we apply it to semiparametric AFT models in a number of scenarios (results not shown), the method seems to have very

notable bias (>20% of the mean). We do not have a direct theoretical explanation for this occurrence of bias. Due to the connection between SIMEX method and the corrected score method (Carroll et al., 2006, Section 7.4), we think this may be caused by the way semiparametric AFT models work with risk sets and the fact that resulting estimating functions are nonsmooth (Tsiatis, 1990).

This article applies a regression calibration (RC) method to the semiparametric AFT model with covariates subject to measurement error. The RC method has the advantage of easy implementation. Usually a two-step process can be carried out. The first step is approximation, that is, to replace missing data by its conditional mean given surrogates and other available covariates. After this approximation, a standard analysis such as the logistic regression can be performed at the second step. Usually variance needs to be inflated to account for the approximation from the first step. In linear regression models without censoring, the RC method yields unbiased estimates for slope parameters when both true and surrogate covariates are normally distributed (Carroll et al., 2006, Chapter 4). This is because the difference between the true covariate and its conditional mean can be absorbed into the error part of linear regression models. As semiparametric AFT models are linear regression models with censored data (see (1) below), this implies that the RC method can also be an intuitively good candidate for them. Carroll et al. (2006, Chapter 4) has provided a comprehensive review of the RC method for analyzing data without censoring. In general, the RC method does not always lead to consistent estimates. So it is an approximate method. On the other hand, nice finite-sample performance has been observed in various applications that has contributed to its popularity especially among epidemiologists (Carroll et al., 2006, Chapter 4).

We also study the stratified random sampling scheme as in our motivating example for selecting the validation sample. This design results in neither independent nor identically distributed observations. Consequently, our development of asymptotic properties of the RC method rely on results for exchangeable random variables. Another sampling method, the i.i.d. Bernoulli sampling, may also be applied as in Wang et al. (1997). Although simpler and analogous results can be obtained from the i.i.d. Bernoulli sampling scheme, we consider the more general stratified random sampling scheme that is known to be more popular in practice.

2. Method

Let $T_i^{\dagger}$ and C_i be log-transformed failure and censoring times for subject i in a cohort of size N. We observe $T_i \equiv T_i^{\dagger} \wedge C_i$, a failure indicator $\Delta_i \equiv I\{T_i^{\dagger} \leq C_i\}$, and a d-dimensional vector of covariates $\mathbf{Z}_i$. The semiparametric AFT model postulates

$$T_i^{\dagger} = \boldsymbol{\beta}' \mathbf{Z}_i + e_i, \quad i = 1, \ldots, N, \tag{1}$$

where e_i's are independent and identically distributed with an unknown distribution F_0 and density function f_0. Censoring time C_i is assumed to be independent of e_i conditioning on $\mathbf{Z}_i$. This model has been discussed extensively in the literature (e.g., Kalbfleisch and Prentice, 2002, Chapter 7) when $\mathbf{Z}_i$ are measured without error.

Let $\epsilon_i^{\boldsymbol{\beta}} = T_i - \boldsymbol{\beta}' \mathbf{Z}_i$ be the observed residual for subject i, $N_i(t, \boldsymbol{\beta}) = I\left(\epsilon_i^{\boldsymbol{\beta}} \leq t, \Delta_i = 1\right)$ the residual counting
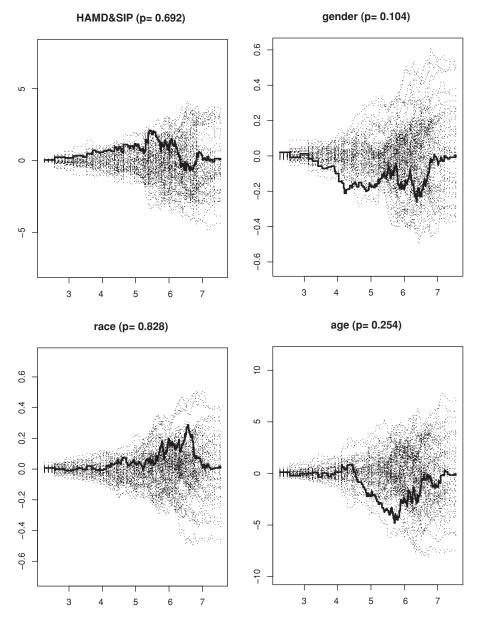
Figure 1. Goodness of fit for the depression mortality data.

process, and $Y_i(t, \boldsymbol{\beta}) = I\ (\epsilon_i^{\boldsymbol{\beta}} \geq t)$ the residual at-risk process. When $(\mathbf{Z}_i, T_i, \Delta_i)$ are observed for the entire cohort, Tsiatis (1990) introduced the following rank-based estimating function for $\boldsymbol{\beta}_0$,

$$S_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \int \rho_N(t, \boldsymbol{\beta}) \left\{ \mathbf{Z}_i - \bar{\mathbf{Z}}(t, \boldsymbol{\beta}) \right\}\ dN_i(t, \boldsymbol{\beta}), \quad (2)$$

where $\rho_N(t, \boldsymbol{\beta})$ is a weight function and $\bar{\mathbf{Z}}(t, \boldsymbol{\beta}) \equiv D_N^{(1)}(t, \boldsymbol{\beta})/D_N^{(0)}(t, \boldsymbol{\beta})$ with $D_N^{(0)}(t, \boldsymbol{\beta}) = N^{-1} \sum_{j=1}^{N} Y_j(t, \boldsymbol{\beta})$ and $D_N^{(1)}(t, \boldsymbol{\beta}) = N^{-1} \sum_{j=1}^{N} \mathbf{Z}_j Y_j(t, \boldsymbol{\beta})$.

As in the depression mortality study, we consider the situation when part of $\mathbf{Z}_i$ is subject to measurement error or missing. We also assume that there is an internal validation subset (Carroll et al., 2006, Chapter 4) on which $\mathbf{Z}_i$ are ac-

curately measured. Following the notation of Breslow and Wellner (2007), we focus on the situation when the validation subset is selected via the following two-phase sampling scheme. Let $\mathbf{Z} = (Z_1', Z_2')'$ with Z_1 subject to measurement error and Z_2 always observable. At the first phase, we observe $\mathbf{V} = (W, Z_2)' \in \mathcal{V}$ where W is a surrogate for Z_1. At the second phase, a subsample is selected and Z_1 is fully observed for these subjects. The subset is called a validation set because both Z_1 and W are observed. For subjects not in the validation set, Z_1 is missing. A popular selection scheme is stratified simple random sampling. Specifically, $\mathcal{V}$ is partitioned into J strata, or $\mathcal{V} \equiv \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_J$. Within each stratum $\mathcal{V}_j$ of size N_j, a simple random sample of size $n_j \leq N_j$ are randomly drawn without replacement for $1 \leq j \leq J$. We use a doubly subscripted notation $\eta_{j,i}$ to indicate the ith subject in stratum j, where $\eta_{j,i} = 1$ if the subject is selected

into the validation set and 0 otherwise. Note that the random variables $(\eta_{j,1}, \ldots, \eta_{j,N_j})$ are exchangeable and the J random vectors $(\eta_{j,1}, \ldots, \eta_{j,N_j})$ are independent. We further assume that censoring variable C is independent of $T^\dagger$ given observed covariate data $\mathbf{V}$. In the depression mortality study, Z_1 contains the HAM-D and SIP scores, Z_2 contains covariates such as gender, race, and age, whereas W contains CES-D scores. CES-D scores were stratified into two categories at the first phase: one with scores 16 or higher and the other with scores below 16. Simple random samples were then selected from the two categories in the second phase.

Assume that the conditional mean of Z_{1i} given Z_{2i} and W_i can be modeled by $E(Z_1 \,|\, Z_2, W) = g(Z_2, W, \boldsymbol{\alpha})$ for some nuisance parameter $\boldsymbol{\alpha}$ of dimension q. The nuisance parameter $\boldsymbol{\alpha}$ is usually unknown and can be estimated using the validation data. Now denote $\mathbf{Z}_{j,i}(\boldsymbol{\alpha}) = (Z_{1j,i}\,\eta_{j,i} + (1-\eta_{j,i})\,g(Z_{2j,i}, W_{j,i}, \boldsymbol{\alpha}), Z_{2j,i})'$. Let $\epsilon_{j,i}^{\boldsymbol{\beta}} = T_{j,i} - \beta_1' Z_{1j,i} - \beta_2' Z_{2j,i}, N_{j,i}(t, \boldsymbol{\beta}) = I\{\epsilon_{j,i}^{\boldsymbol{\beta}} \leq t, \Delta_{j,i} = 1\}, Y_{j,i}(t, \boldsymbol{\beta}) = I\{\epsilon_{j,i}^{\boldsymbol{\beta}} \geq t\}$ be the residual, its counting process, and at-risk process for subjects with complete data. Let $\epsilon_{j,i}^{\boldsymbol{\beta},\boldsymbol{\alpha}} = T_{j,i} - \beta_1' g(Z_{2j,i}, W_{j,i}, \boldsymbol{\alpha}) - \beta_2' Z_{2j,i}, N_{j,i}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = I\{\epsilon_{j,i}^{\boldsymbol{\beta},\boldsymbol{\alpha}} \leq t, \Delta_{j,i} = 1\}, Y_{j,i}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = I\{\epsilon_{j,i}^{\boldsymbol{\beta},\boldsymbol{\alpha}} \geq t\}$ be the residual, its counting process, and at-risk process when $Z_{1j,i}$ is replaced by $g(Z_{2j,i}, W_{j,i}, \boldsymbol{\alpha})$. Denote

$$D_N^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{N_j} \{\eta_{j,i} Y_{j,i}(t, \boldsymbol{\beta}) + (1-\eta_{j,i}) Y_{j,i}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})\}. \quad (3)$$

$$D_N^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{N_j} \{\eta_{j,i} \mathbf{Z}_{j,i} Y_{j,i}(t, \boldsymbol{\beta}) + (1-\eta_{j,i}) \mathbf{Z}_{j,i}(\boldsymbol{\alpha}) Y_{j,i}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})\}. \quad (4)$$

If $\boldsymbol{\alpha}$ is known, an RC-based estimating function similar to (2) can be formed as

$$\begin{aligned} S_N(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{N_j} \eta_{j,i} \int \big\{ & D_N^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) \mathbf{Z}_{j,i} \\ & - D_N^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) \big\} \mathrm{d}N_{j,i}(t, \boldsymbol{\beta}) \\ + \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{N_j} (1-\eta_{j,i}) \int \big\{ & D_N^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) \mathbf{Z}_{j,i}(\boldsymbol{\alpha}) \\ - D_N^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) \big\} & \mathrm{d}N_{j,i}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) \end{aligned} \quad (5)$$

where Gehan's weight, $D_N^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})/D_N^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})$, is used for the weighting process ρ_N.

We focus on the above Gehan weighted estimating function in this article mainly for notational simplicity. However, our theoretical results should hold for more general weight functions. Our choice is based on two advantages that Gehan weighted estimating function has. First, $S_N(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is a monotone function (Fygenson and Ritov, 1994; Yu and Nan, 2006), which is a very nice property to have for root finding. Second, the estimating function does not involve fractions.

Consequently the condition of truncating the stochastic integration (see (3.1) of Tsiatis, 1990) is not needed in our investigation of the asymptotic properties of $S_N(\boldsymbol{\beta}, \boldsymbol{\alpha})$. This not only prevents potential information loss due to truncation (Ying, 1993) but also provides less restriction, especially in the presence of measurement error data.

Since $\boldsymbol{\alpha}$ is unknown, we replace it in (5) by an estimate $\hat{\boldsymbol{\alpha}}$, which is usually obtainable using existing software based on the validation data. Consequently, we propose using $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) = 0$ as our estimating equation for $\boldsymbol{\beta}$, based on the RC method. To obtain an estimate for $\boldsymbol{\beta}$, we replace missing Z_1 with $g(Z_2, W, \hat{\boldsymbol{\alpha}})$ and perform a full cohort AFT model analysis to obtain a valid regression estimate $\hat{\boldsymbol{\beta}}_N$. Despite the nonsmooth nature of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ in $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_N$ can be readily obtained using existing computation method in full cohort AFT models (e.g., Lin and Geyer, 1992; Jin et al., 2003; and Yu and Nan, 2006).

3. Asymptotic Properties

In this section, we explore asymptotic properties of our proposed estimating function $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ and its corresponding estimate $\hat{\boldsymbol{\beta}}_N$. We show that $\hat{\boldsymbol{\beta}}_N$ converges to some well-defined parameter that may or may not be equal to the true parameter. When the magnitude of regression coefficient or the measurement error is small, however, bias is in general very limited (see the comment above Proposition 2). Improvement in efficiency on the other hand can be quite remarkable compared with analysis using only validation data. This will be demonstrated numerically in Section 5.

In Proposition 1, we show uniform convergence of $D_N^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}), D_N^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})$, and $S_N(\boldsymbol{\beta}, \boldsymbol{\alpha})$ to their limits:

$$\begin{aligned} d^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{j=1}^{J} \nu_j \big[& p_j P_{0|j}\{\epsilon^{\boldsymbol{\beta}} \geq t\} \\ & + (1-p_j) P_{0|j}\{\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}} \geq t\} \big] \end{aligned} \quad (6)$$

$$\begin{aligned} d^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{j=1}^{J} \nu_j \big[& p_j P_{0|j}\{\mathbf{Z}I(\epsilon^{\boldsymbol{\beta}} \geq t)\} \\ & + (1-p_j) P_{0|j}\{\mathbf{Z}(\boldsymbol{\alpha}) I(\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}} \geq t)\} \big] \end{aligned} \quad (7)$$

$$\begin{aligned} s(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{j=1}^{J} \nu_j \big(& \big[p_j P_{0|j}\{\mathbf{Z} d^{(0)}(\epsilon^{\boldsymbol{\beta}}; \boldsymbol{\beta}, \boldsymbol{\alpha})\Delta\} \\ & + (1-p_j) P_{0|j}\{\mathbf{Z}(\boldsymbol{\alpha}) d^{(0)}(\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}}; \boldsymbol{\beta}, \boldsymbol{\alpha})\Delta\} \big] \\ & - \big[p_j P_{0|j}\{d^{(1)}(\epsilon^{\boldsymbol{\beta}}; \boldsymbol{\beta}, \boldsymbol{\alpha})\Delta\} \\ & + (1-p_j) P_{0|j}\{d^{(1)}(\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}}; \boldsymbol{\beta}, \boldsymbol{\alpha})\Delta\} \big] \big). \end{aligned} \quad (8)$$

Here $p_j = n_j/N_j, \nu_j$ is the limit of N_j/N, and $P_{0|j}(\cdot) = E(\cdot \,|\, \mathbf{V} \in \mathcal{V}_j)$ is the conditional expectation in stratum j. Under regularity assumptions including the uniqueness of $\boldsymbol{\beta}^*$ that solves $s(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = 0$ at the true $\boldsymbol{\alpha}_0$, we then have the convergence of $\hat{\boldsymbol{\beta}}_N$ to $\boldsymbol{\beta}^*$. In Proposition 2, we show the asymptotic linearity property of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ and the asymptotic normality of $\hat{\boldsymbol{\beta}}_N$.

The following regularity conditions are needed for our propositions:

C1. The covariate **Z** has bounded support.

C2. The number of strata, J, is finite and fixed. For any stratum j, $v_j > 0$ and $\sqrt{N}(N_j/N - \nu_j) = O_p(1)$.

C3. For any given set $\mathbf{A} \subset \mathbf{R}^q$,

  i. The conditional mean of $E(Z_1 \,|\, Z_2, \, W) = g(Z_2, W, \boldsymbol{\alpha})$ is continuous with respect to $\boldsymbol{\alpha} \in \mathbf{A}$ and uniformly bounded.

  ii. The class $\{g(Z_2, \, W, \, \boldsymbol{\alpha}), \, \boldsymbol{\alpha} \in \mathbf{A}\}$ forms a P-Donsker class.

C4. The true value $\boldsymbol{\alpha}_0$ is an interior point of $\mathbf{A}$ such that $\sqrt{N}\{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\}$ is asymptotically normal with mean 0.

C5. For each stratum j, both $P_{0|j}\{\epsilon^{\boldsymbol{\beta}} \geq t\}$ and $P_{0|j}\{\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}} \geq t\}$ are continuous in $\boldsymbol{\beta}, t$, and $\boldsymbol{\alpha}$. The limit functions $d^{(0)}(\epsilon^{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\alpha}), d^{(0)}(\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\alpha}), d^{(1)}(\epsilon^{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\alpha})$, and $d^{(1)}(\epsilon^{\boldsymbol{\beta},\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ are continuously differentiable with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

C6. The solution to $s(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = 0$ for $\boldsymbol{\beta}$ is unique and is an interior point of $\mathbf{B}$, where $\mathbf{B} \subset \mathbf{R}^d$ is a compact set of possible parameter values.

Condition **C1** commonly holds. Condition **C2** is easily satisfied in our case, where the strata are determined by **V** and N_j follows a binomial distribution with success probability $\nu_j = P(\mathbf{V} \in \mathcal{V}_j)$. For Condition **C3**, suppose $g(Z_2, W, \boldsymbol{\alpha}) = h(\alpha_0 + \alpha_Z Z_2 + \alpha_W W)$, where $h(\cdot)$ is monotone and continuously differentiable, then from Lemma 2.6.15 and 2.6.18 (part viii) of van der Vaart and Wellner (1996), we know that $\{h(\alpha_0 + \alpha_Z Z + \alpha_W W): \alpha_0, \alpha_Z, \alpha_W \in R\}$ forms a VC-subgraph class. If a VC-subgraph class has an envelope function with finite second moment, it is P-Donsker. This includes generalized linear models for the conditional mean of $E(Z_1 \,|\, Z_2, W) = g(Z_2, W, \boldsymbol{\alpha})$. Condition **C4** consists of expected properties for estimates of $\boldsymbol{\alpha}$ using validation data based on stratified simple random sampling, for which various methods have been proposed. Specifically, we have the following asymptotic representation under the stratified simple random sampling (Breslow and Wellner, 2007):

$$(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = N^{-1}\mathbf{H}^{-1}\sum_{j=1}^{J}\left\{p_j^{-1}\sum_{i=1}^{N_j}\eta_{j,i}e_{j,i}\mathbf{V}_{j,i}\right\} + o_p(N^{-1/2}), \quad (9)$$

where $\mathbf{H}$ is the Hessian matrix, $e_{j,i} = Z_{1j,i} - g(Z_{2j,i}, W_{j,i}, \boldsymbol{\alpha})$ is the residual and $\mathbf{V}_{j,i} = (Z_{2j,i}, W_{j,i})$. Condition **C5** is likely to hold in situations when both the failure time and censoring time are continuously distributed, given the continuity condition **C3.i** for g. As a result of **C5**, $s(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is continuously differentiable in both its arguments. Condition **C6** is essential for the consistency of $\hat{\boldsymbol{\beta}}_N$.

Proposition 1 provides convergence results of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ and $\hat{\boldsymbol{\beta}}_N$ and Proposition 2 provides the asymptotic linearity of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ in an $N^{-1/2}$ neighborhood of $\boldsymbol{\beta}^*$ and the asymptotic normality of $\hat{\boldsymbol{\beta}}_N$. We delegate the proofs of the propositions to the Web Appendix due to their length.

PROPOSITION 1: *Assume that the number of strata J is finite. Let $d^{(0)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha}), d^{(1)}(t, \boldsymbol{\beta}, \boldsymbol{\alpha})$, and $s(\boldsymbol{\beta}, \boldsymbol{\alpha})$ be defined as in (6), (7), and (8) correspondingly.*

1. *Under Conditions **C1**–**C5**, we have*

$$\sup_{\boldsymbol{\beta}\in\mathbf{B},\boldsymbol{\alpha}\in\mathbf{A},t\in R}\left|D_N^{(0)}(t,\boldsymbol{\beta},\boldsymbol{\alpha}) - d^{(0)}(t,\boldsymbol{\beta},\boldsymbol{\alpha})\right| = o_{p^*}(1),$$
$$(10)$$

$$\sup_{\boldsymbol{\beta}\in\mathbf{B},\boldsymbol{\alpha}\in\mathbf{A},t\in R}\left|D_N^{(1)}(t,\boldsymbol{\beta},\boldsymbol{\alpha}) - d^{(1)}(t,\boldsymbol{\beta},\boldsymbol{\alpha})\right| = o_{p^*}(1).$$
$$(11)$$

Here p^* denotes in outer probability. See Chapter 1 of van der Vaart and Wellner (1996) for the definition of outer probability.

2. *Let $\|\cdots\|$ denote the super-norm. Let $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbf{A}$ such that $\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| \to 0$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbf{B}$ such that $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| \to 0$ as $N \to \infty$. Under Conditions **C1**–**C5**,*

$$\sup_{\boldsymbol{\beta}\in\mathbf{B}}\sqrt{N}\,|\,\{S_N(\boldsymbol{\beta},\boldsymbol{\alpha}_1) - S_N(\boldsymbol{\beta},\boldsymbol{\alpha}_2)\}$$
$$-\{s(\boldsymbol{\beta},\boldsymbol{\alpha}_1) - s(\boldsymbol{\beta},\boldsymbol{\alpha}_2)\}\,| = o_{p^*}(1), \quad (12)$$

$$\sup_{\boldsymbol{\alpha}\in\mathbf{A}}\sqrt{N}\,|\,\{S_N(\boldsymbol{\beta}_1,\boldsymbol{\alpha}) - S_N(\boldsymbol{\beta}_2,\boldsymbol{\alpha})\}$$
$$-\{s(\boldsymbol{\beta}_1,\boldsymbol{\alpha}) - s(\boldsymbol{\beta}_2,\boldsymbol{\alpha})\}\,| = o_{p^*}(1). \quad (13)$$

3. *If further the condition **C6** holds, then the solution $\hat{\boldsymbol{\beta}}_N$ of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) = o_{p^*}(N^{-1/2})$ converges to $\boldsymbol{\beta}^*$ with probability 1 where $\boldsymbol{\beta}^* \in \mathbf{B}$ solves $s(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) = 0$.*

Note that $\boldsymbol{\beta}^*$ may not be equal to $\boldsymbol{\beta}_0$. In such case, the estimate resulting from the RC method can be at the best viewed as an approximation. In the following two important cases of measurement error models, however, consistent estimation (i.e., $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$) is obtainable.

The first case is under the general Berkson measurement error model that specifies the conditional distribution of Z_{1i} given W_i and Z_{2i}. One such example assumes a linear structure between Z_1 and $(W, Z_2), Z_1 = \psi_0 + \psi_1'W + \psi_2'Z_2 + U$, where the error U is independent of (W, Z_2). Note that in this case, $T_i^\dagger$ and imputed covariate vector $(g(Z_{2i}, W_i, \boldsymbol{\alpha}), Z_{2i})$ still follow a semiparametric AFT model. This can be seen by writing $T^\dagger = \beta_1 E[Z_1 \,|\, W, Z_2] + \beta_2 Z_2 + \beta_1 U + e$. Hence, if we denote $e^* = \beta_1 U + e$, then e^* is independent of $E[Z_1 \,|\, W, Z_2]$ and Z_2. This implies $\beta_1 = \beta_1^*$ and $\beta_2 = \beta_2^*$.

The second case is under the additive measurement error model when both Z and its surrogate W are jointly normally distributed. The additive measurement error model specifies a linear structure between W and $(Z_1, Z_2), W = \gamma_0 + \gamma_1'Z_1 + \gamma_2'Z_2 + U$, where the error U is independent of (Z_1, Z_2). In this case, we can rewrite $T^\dagger = \beta_1 E[Z_1 \,|\, W, Z_2] + \beta_2 Z_2 + \beta_1(Z_1 - E[Z_1 \,|\, W, Z_2]) + e$. Then, due to the joint normality and the fact that $(Z_1 - E[Z_1 \,|\, W, Z_2])$ is uncorrelated with both $E[Z_1 \,|\, W, Z_2]$ and Z_2, we conclude that $(Z_1 - E[Z_1 \,|\, W, Z_2])$ is also independent of both $E[Z_1 \,|\, W, Z_2]$ and Z_2. As a result, we can denote $e^* = \beta_1(Z_1 - E[Z_1 \,|\, W, Z_2]) + e$. Hence $T^\dagger$ still is linearly related to $E[Z_1 \,|\, W, Z_2]$ and Z_2, and e^* can be treated as an error term independent of $E[Z_1 \,|\, W, Z_2]$ and Z_2. Again we have $\beta_1 = \beta_1^*$ and $\beta_2 = \beta_2^*$.

In general cases, we may have asymptotic bias. However such bias is likely to be small when the variation of the term $\beta_1(Z_1 - E[Z_1 \,|\, W, Z_2])$ is relatively small comparing to the variation of e. This can happen if either $|\beta_1|$ is small or the measurement error is small.

PROPOSITION 2:   *Assume conditions* **C1**–**C6** *hold.*

1. *For any* $\boldsymbol{\alpha} \in \mathbf{A}$, $S_N(\boldsymbol{\beta}, \boldsymbol{\alpha})$ *is asymptotically linear in an* $N^{-1/2}$ *neighborhood of* $\boldsymbol{\beta}^*$. *In other words, if* $\sqrt{N}\,|\boldsymbol{\beta}_N - \boldsymbol{\beta}^*| = O_{p^*}(1)$, *then*

$$\sqrt{N}\{S_N(\boldsymbol{\beta}_N, \boldsymbol{\alpha}) - S_N(\boldsymbol{\beta}^*, \boldsymbol{\alpha})\}$$
$$= \sqrt{N}(\boldsymbol{\beta}_N - \boldsymbol{\beta}^*)\dot{s}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}) + o_{p^*}(1), \qquad (14)$$

where $\dot{s}_{\boldsymbol{\beta}}(\cdot, \cdot)$ *denotes the partial derivative of s with respect to* $\boldsymbol{\beta}$.

2. $\sqrt{N}\{S_N(\boldsymbol{\beta}_N, \hat{\boldsymbol{\alpha}}) - S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})\}$ *is asymptotically equivalent to* $\sqrt{N}\{S_N(\boldsymbol{\beta}_N, \boldsymbol{\alpha}_0) - S_N(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0)\}$ *in the sense that*

$$\sqrt{N}\,|\,\{S_N(\boldsymbol{\beta}_N, \hat{\boldsymbol{\alpha}}) - S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})\}$$
$$- \{S_N(\boldsymbol{\beta}_N, \boldsymbol{\alpha}_0) - S_N(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0)\}\,|$$
$$= o_{p^*}(\sqrt{N}\,|\boldsymbol{\beta}_N - \boldsymbol{\beta}^*\|).$$

As a result, $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ *is asymptotically linear in an* $N^{-1/2}$ *neighborhood of* $\boldsymbol{\beta}^*$. *That is, for any* $\boldsymbol{\beta}_N \in \mathbf{B}$ *such that* $\sqrt{N}\,|\boldsymbol{\beta}_N - \boldsymbol{\beta}^*| = O_{p^*}(1)$,

$$\sqrt{N}\{S_N(\boldsymbol{\beta}_N, \hat{\boldsymbol{\alpha}}) - S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})\}$$
$$= \sqrt{N}(\boldsymbol{\beta}_N - \boldsymbol{\beta}^*)\dot{s}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0) + o_{p^*}(1). \qquad (15)$$

3. *If* $\dot{s}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}_0)$ *is nonsingular, then* $\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}^*)$ *is asymptotically normal with mean* 0.

4. Variance Estimation

From Proposition 2, we see that estimation of the asymptotic variance of $\hat{\boldsymbol{\beta}}_N$ requires estimation of the asymptotic slope of $S_N(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$, which involves estimation of both $f_0(\cdot)$ and censoring distribution. It is known to be hard to estimate (e.g., Lin and Geyer, 1992; Jin et al., 2003). Here, we adopt the method proposed by Huang (2002). Similar algorithms have been used in Nan, Yu, and Kalbfleisch (2006) and Yu and Nan (2006). Numerical stability in variance estimation has been observed from these papers. Specifically, we follow the following three steps.

1. Decompose the estimated asymptotic variance of $N^{-1/2}S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})$ into $\mathbf{CC}^T$, where $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_d)$ and $\mathbf{C}^T$ is the transpose of $\mathbf{C}$;
2. Solve equations $N^{-1/2}S_N(\widetilde{\boldsymbol{\beta}}_j, \hat{\boldsymbol{\alpha}}) = \mathbf{c}_j$ for $\widetilde{\boldsymbol{\beta}}_j$, $j = 1, \ldots, d$.
3. Let $\mathbf{D} = (\widetilde{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_N, \ldots, \widetilde{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_N)$. Then, $N\mathbf{DD}^T$ is a consistent variance estimator of $N^{1/2}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}^*)$.

For step 2, an effective hybrid Newton method proposed by Yu and Nan (2006) can be applied. Step 1 requires estimation of the asymptotic variance of $N^{-1/2}S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})$. We delegate the estimation procedure to Web Appendix B. The essence is to express $\sqrt{N}S_N(\boldsymbol{\beta}^*, \hat{\boldsymbol{\alpha}})$ asymptotically as a sum of mutually independent empirical processes. Then, its asymptotic variance can be estimated by using sample estimates corresponding to these empirical processes.

5. Numerical Examples

5.1 *Simulations*

For simplicity, we consider one-dimensional Z and its surrogate W. Sample size N is chosen to be 300, 1000, or 3000. Validation sample size n is chosen to be 90 or 120 when $N = 300$,

200, or 400 when $N = 1000$, and 400 or 800 when $N = 3000$. We generate 500 data sets under each of these scenarios.

The AFT model is $\log T = \beta Z + e$ where the true $\beta_0 = 1$. Error distribution is from uniform $UNIF(-3, 3)$. Censoring time C is distributed according to a shifted exponential distribution with its shift and scale chosen so that the censoring rate is approximately 70%. For the covariate Z, we consider two distributions: standard normal and Bernoulli. Under the first setting when Z is continuous, we generate the surrogate W according to $W = Z + U$ such that $U \sim N(0, \sigma_W)$. Under the second setting when Z is binary, we generate the surrogate W so that it is a misclassification of Z via concordance probability $p_W = P(W = 1\,|\,Z = 1) = P(W = 0\,|\,Z = 0)$. The conditional mean is estimated via least square regression under the first setting and via sampling frequency under the second setting.

The validation sets are selected via stratified simple random sampling. The stratification is based on whether W is less than 0.5 or not under the first scenario, and whether W is 0 or not under the second scenario. We perform four analyses on each simulated data set to compare their performances under various scenarios. The analyses are: full cohort analysis (Full) where all data are used; validation data only analysis (VA) where only observations in the validation subset are used; naive analysis (Naive) where missing Zs are replaced by their surrogates W; and the RC analysis (RC).

Results are listed in Tables 1 and 2. We let σ_W and p_W vary to investigate how the quality of the surrogate affects the estimation. From the simulation results, we see that full cohort estimates are the most efficient as one would expect. VA estimates are consistent but not very efficient unless the validation size is large. The consistency of the VA method is due to the fact that sampling is based only on covariates that do not alter the conditional distribution of the outcome. The Naive method clearly leads to severe bias. On the other hand, RC estimates have very limited bias and very impressive efficiency improvement, particularly when the validation size increases or when measurement error or classification error diminishes. It also has smaller standard error than the VA analysis under all scenarios.

We also carry out simulation studies to compare the IPW and RC methods under the same settings. From Table 3, we observed a significant amount of reduction in empirical mean square errors of the RC estimates in almost all cases. So the RC method can be an attractive alternative to the IPW method.

5.2 *The Depression Mortality Study*

We analyze the depression mortality study introduced at the beginning of the article. Among 3767 patients screened at the phase 1 of the study, 3155 patients scored below 16 on the CES-D. Among them a simple random sample of 100 were selected. Among 612 patients who scored 16 or higher on the CES-D, 300 were selected. At the phase 2 of the study, the 400 selected patients were required to complete HAM-D and SIP. There are total 3756 subjects in the analysis data set with seven cases deleted due to missing information. The total number of death is 810 and the median follow-up time is 72 months. We study the impact of depression, age, gender, and race on patient survival. In the original study (Callahan et al., 1998), more covariates were considered. Here, we use a subset

Table 1
Simulation results based on 500 data sets under the model $\log T = \beta Z + \epsilon$ (with $\beta_0 = 1$) where Z is continuous

	Full	$\sigma_W = 0.5$			$\sigma_W = 1$			$\sigma_W = 2$		
		VA	Naive	RC	VA	Naive	RC	VA	Naive	RC
$N = 300, n = 90$										
Avg $\hat{\beta}$	0.996	1.009	0.840	0.992	0.994	0.577	0.989	0.994	0.251	0.973
SE $\hat{\beta}$	0.210	0.373	0.358	0.233	0.378	0.302	0.279	0.359	0.207	0.320
Avg $\hat{\sigma}(\hat{\beta})$	0.209	0.386	0.365	0.231	0.391	0.312	0.274	0.398	0.211	0.338
90% CP	88.8%	90.8%	87.4%	86.8%	90.2%	61.0%	88.0%	91.4%	3.2%	89.4%
95% CP	96.0%	95.8%	91.6%	93.4%	94.8%	72.0%	94.4%	96.8%	6.4%	93.4%
$N = 300, n = 120$										
Avg $\hat{\beta}$	1.012	1.004	0.884	1.014	1.007	0.641	1.021	1.032	0.302	1.031
SE $\hat{\beta}$	0.211	0.318	0.320	0.229	0.346	0.279	0.270	0.338	0.193	0.300
Avg $\hat{\sigma}(\hat{\beta})$	0.211	0.335	0.325	0.229	0.340	0.282	0.265	0.338	0.198	0.308
90% CP	88.4%	89.4%	86.8%	90.2%	86.6%	63.6%	87.8%	89.0%	3.0%	89.8%
95% CP	95.4%	95.4%	93.0%	94.6%	93.4%	74.2%	95.0%	93.2%	6.2%	95.0%
$N = 1000, n = 200$										
Avg $\hat{\beta}$	0.996	1.018	0.970	0.996	1.017	0.834	0.994	0.998	0.551	0.999
SE $\hat{\beta}$	0.109	0.236	0.243	0.114	0.231	0.235	0.129	0.255	0.207	0.167
Avg $\hat{\sigma}(\hat{\beta}_N)$	0.112	0.251	0.255	0.116	0.255	0.242	0.127	0.260	0.204	0.155
90% CP	91.8%	91.8%	90.8%	90.2%	92.4%	83.4%	90.0%	89.8%	28.0%	86.2%
95% CP	96.0%	96.4%	95.4%	96.6%	97.4%	88.6%	95.0%	95.4%	44.0%	94.6%
$N = 1000, n = 400$										
Avg $\hat{\beta}$	1.005	0.983	0.969	1.005	1.005	0.866	1.004	0.992	0.626	1.003
SE $\hat{\beta}$	0.108	0.179	0.171	0.110	0.176	0.164	0.117	0.175	0.145	0.134
Avg $\hat{\sigma}(\hat{\beta}_N)$	0.112	0.175	0.179	0.115	0.178	0.173	0.123	0.180	0.152	0.140
90% CP	91.8%	88.8%	91.8%	91.8%	92.8%	82.2%	90.6%	91.2%	20.4%	91.6%
95% CP	95.8%	93.0%	95.6%	94.8%	95.6%	90.0%	95.0%	95.8%	30.6%	96.4%
$N = 3000, n = 400$										
Avg $\hat{\beta}$	1.000	1.005	0.971	1.001	1.011	0.824	0.999	1.008	0.533	0.999
SE $\hat{\beta}$	0.063	0.169	0.180	0.065	0.169	0.170	0.075	0.183	0.144	0.099
Avg $\hat{\sigma}(\hat{\beta})$	0.064	0.174	0.178	0.066	0.178	0.168	0.073	0.180	0.140	0.091
90% CP	91.2%	90.6%	89.2%	89.4%	91.2%	71.2%	88.8%	88.8%	5.0%	87.4%
95% CP	96.8%	96.2%	94.4%	96.0%	96.0%	82.2%	94.2%	94.8%	8.8%	92.0%
$N = 3000, n = 800$										
Avg $\hat{\beta}$	1.001	0.996	0.968	0.999	1.001	0.843	0.998	0.993	0.574	0.999
SE $\hat{\beta}$	0.060	0.126	0.123	0.063	0.126	0.118	0.071	0.122	0.099	0.085
Avg $\hat{\sigma}(\hat{\beta})$	0.064	0.122	0.125	0.065	0.124	0.119	0.072	0.125	0.102	0.085
90% CP	92.6%	88.8%	90.8%	92.4%	89.4%	61.8%	90.0%	92.6%	0.4%	90.0%
95% CP	97.4%	93.6%	95.4%	96.6%	93.4%	72.6%	96.0%	95.8%	1.4%	96.0%

of variables to illustrate our method. There exists a clearly linear relationship between CES-D and HAMD&SIP (sum of HAM-D and SIP) scores among the 400 validation subjects. The estimated correlation coefficient is 0.58. The estimated reliability ratio is 0.653 (note that the reliability ratios in our simulation study under the normal covariates are 0.8, 0.5, and 0.2, corresponding to σ_W of 0.5, 1, and 2).

In our analysis, we first regress HAMD&SIP on CES-D using the 400 validation subjects. The resulted intercept and slope estimates are (9.78, 0.52) with standard errors (0.84, 0.04). Then, we apply the RC method. Both the results of RC and VA methods are listed in Table 4. In the VA analysis, gender is found to be significantly associated with mortality, whereas in the RC analysis, age is found to be significantly associated with the mortality. Note that the coefficients of HAMD&SIP, FEMALE, and AGE have the same direction as those based on the VA analysis, but the sign of the coefficient of BLACK is reversed.

6. Discussion

We have proposed a RC method for estimating parameters in semiparametric AFT models with covariates subject to measurement error. The method performs well in simulations and has clear advantage over ad hoc approaches. As our asymptotic results are derived under very general assumptions between the true covariate vector Z_1 and its surrogate W, this method is also applicable for survival data with missing covariates as long as reliable calibration between missing and observed covariates is feasible. In this case, the RC method is a mean imputation method.

Table 2
Simulation results based on 500 *data sets under the model* $\log T = \beta Z + \epsilon$ *(with* $\beta_0 = 1$*) where Z is binary*

		$p_W = 0.9$			$p_W = 0.7$			$p_W = 0.5$		
	Full	VA	Naive	RC	VA	Naive	RC	VA	Naive	RC
$N = 300, n = 90$										
Avg $\hat{\beta}$	1.001	1.042	0.235	1.020	0.933	0.531	0.944	1.003	0.866	0.983
SE $\hat{\beta}$	0.448	0.835	0.812	0.807	0.798	0.821	0.702	0.723	0.858	0.511
Avg $\hat{\sigma}(\hat{\beta})$	0.455	0.864	0.779	0.861	0.823	0.807	0.731	0.780	0.840	0.536
90% CP	91.2%	89.8%	71.4%	90.4%	91.6%	82.4%	91.0%	90.6%	87.0%	91.0%
95% CP	94.6%	94.0%	78.6%	94.0%	96.0%	89.0%	95.6%	95.4%	93.8%	96.0%
$N = 300, n = 120$										
Avg $\hat{\beta}$	1.036	1.004	0.382	0.995	1.050	0.660	1.065	0.999	0.898	1.030
SE $\hat{\beta}$	0.444	0.719	0.7%	0.712	0.721	0.704	0.675	0.665	0.719	0.511
Avg $\hat{\sigma}(\hat{\beta})$	0.456	0.730	0.678	0.729	0.714	0.705	0.659	0.675	0.728	0.520
90% CP	91.4%	88.8%	74.4%	90.0%	88.0%	82.0%	89.6%	89.0%	88.0%	89.2%
95% CP	96.0%	94.4%	83.0%	95.0%	93.4%	89.4%	94.6%	94.6%	94.2%	95.6%
$N = 1000, n = 200$										
Avg $\hat{\beta}$	1.005	1.015	0.823	1.016	1.035	0.445	1.032	1.024	0.152	1.001
SE $\hat{\beta}$	0.227	0.522	0.526	0.297	0.526	0.523	0.443	0.532	0.469	0.531
Avg $\hat{\sigma}(\hat{\beta}_N)$	0.237	0.495	0.535	0.288	0.525	0.518	0.434	0.543	0.507	0.547
90% CP	92.4%	89.2%	86.8%	88.2%	89.8%	70.4%	89.2%	90.6%	46.2%	90.8%
95% CP	95.8%	94.8%	93.0%	93.2%	95.0%	79.8%	94.0%	95.2%	60.0%	95.6%
$N = 1000, n = 400$										
Avg $\hat{\beta}$	1.005	1.005	0.887	1.003	1.018	0.610	1.017	1.029	0.363	1.028
SE $\hat{\beta}$	0.232	0.334	0.348	0.266	0.351	0.370	0.335	0.391	0.348	0.389
Avg $\hat{\sigma}(\hat{\beta}_N)$	0.238	0.348	0.380	0.269	0.371	0.372	0.344	0.381	0.362	0.383
90% CP	92.2%	90.2%	91.2%	90.0%	92.8%	71.6%	90.2%	88.8%	44.6%	88.8%
95% CP	98.0%	95.0%	95.8%	96.0%	95.0%	80.8%	95.4%	94.4%	60.0%	95.0%
$N = 3000, n = 400$										
Avg $\hat{\beta}$	0.990	0.997	0.774	0.986	0.996	0.388	0.980	0.957	0.086	0.945
SE $\hat{\beta}$	0.125	0.354	0.370	0.164	0.355	0.372	0.285	0.342	0.373	0.343
Avg $\hat{\sigma}(\hat{\beta})$	0.136	0.348	0.374	0.170	0.370	0.364	0.277	0.378	0.356	0.383
90% CP	94.0%	89.6%	84.8%	91.4%	91.0%	49.2%	89.8%	92.6%	19.4%	93.0%
95% CP	97.4%	94.6%	90.0%	95.2%	96.4%	61.4%	94.0%	97.8%	27.4%	96.4%
$N = 3000, n = 800$										
Avg $\hat{\beta}$	0.997	0.999	0.849	0.995	1.009	0.518	1.003	0.996	0.221	0.995
SE $\hat{\beta}$	0.134	0.247	0.268	0.156	0.278	0.257	0.221	0.260	0.253	0.261
Avg $\hat{\sigma}(\hat{\beta})$	0.136	0.245	0.265	0.161	0.259	0.259	0.226	0.265	0.253	0.268
90% CP	90.4%	91.2%	82.8%	90.8%	86.8%	41.0%	92.2%	90.4%	9.4%	90.0%
95% CP	96.0%	94.8%	90.0%	95.6%	92.6%	54.2%	95.4%	95.4%	13.2%	95.0%

Due to the limitation of space, we focus in this article on additive measurement error models. Multiplicative errors also arise frequently in practice. While a simple log transformation can convert multiplicative errors to additive errors, sometimes special treatment is needed. We refer readers to Section 4.5 of Carroll et al. (2006) for performing RC methods with multiplicative error models. Extension to semiparametric AFT models may be carried out similarly and warrants further investigation.

The projection technique studied by Chen (2002) for the Cox model and by Jiang and Zhou (2007) for the additive hazards model should be applicable in our setting. The projection estimator is generally consistent and is guaranteed to have greater efficiency than the estimator based on validation data only. To apply this technique, there is a need to study the asymptotic behavior of misspecified AFT models. There are also some computationally challenging issues because the resulting estimator and its variance involves (products) derivatives of nonsmooth estimating functions under both the true AFT model and a misspecified AFT model. This approach is currently under our investigation.

While RC is convenient and often relatively easy to implement, correctly specified maximum likelihood methods to address covariate measurement error have been shown (in other models) to be much more efficient than RC methods (e.g., Spiegelman, Rosner, and Logan, 2000). However, development in our setting can require some work due to the semiparametric nature of the AFT model. This approach warrants further investigation, likely in the context of nonparametric maximum-likelihood estimation.

Table 3
Simulation results comparing the IPW and RC methods

| | $\sigma_W = 0.5$ | | $\sigma_W = 1$ | | $\sigma_W = 2$ | |
Z continuous	IPW	RC	IPW	RC	IPW	RC
$N = 1000, n = 200$						
Avg $\hat\beta$	1.008	0.996	1.007	0.994	1.010	0.999
SE $\hat\beta$	0.263	0.114	0.262	0.129	0.267	0.167
$N = 1000, n = 400$						
Avg $\hat\beta$	0.988	1.005	0.984	1.004	0.980	1.003
SE $\hat\beta$	0.182	0.110	0.175	0.117	0.191	0.134
$N = 3000, n = 400$						
Avg $\hat\beta$	0.998	1.001	0.996	0.999	1.001	0.999
SE $\hat\beta$	0.173	0.065	0.179	0.075	0.182	0.099
$N = 3000, n = 800$						
Avg $\hat\beta$	0.991	0.999	0.995	0.998	1.000	0.999
SE $\hat\beta$	0.127	0.063	0.130	0.071	0.137	0.085
	$p_W = 0.5$		$p_W = 0.7$		$p_W = 0.9$	
Z binary	IPW	RC	IPW	RC	IPW	RC
$N = 1000, n = 200$						
Avg $\hat\beta$	1.010	1.001	1.006	1.032	0.998	1.016
SE $\hat\beta$	0.590	0.531	0.574	0.443	0.544	0.297
$N = 1000, n = 400$						
Avg $\hat\beta$	1.002	1.028	1.007	1.017	0.977	1.003
SE $\hat\beta$	0.390	0.389	0.380	0.335	0.368	0.266
$N = 3000, n = 400$						
Avg $\hat\beta$	1.015	0.945	0.999	0.980	1.005	0.986
SE $\hat\beta$	0.370	0.343	0.362	0.285	0.352	0.164
$N = 3000, n = 800$						
Avg $\hat\beta$	0.983	0.995	0.996	1.003	0.995	0.995
SE $\hat\beta$	0.274	0.261	0.262	0.221	0.257	0.156

Table 4
Results for the depression mortality study

| | Validation only analysis | | | RC analysis | | |
	Estimate	SE	p-value	Estimate	SE	p-value
HAMD&SIP	−0.007	0.011	0.514	−0.021	0.047	0.652
FEMALE	0.991	0.235	<.0001	0.680	0.697	0.330
BLACK	−0.103	0.259	0.691	0.297	0.795	0.709
AGE	−0.031	0.017	0.070	−0.036	0.005	<.0001

6. Supplementary Materials

Web Appendices referenced in Sections 3 and 4 are available under the Paper Information link at the *Biometrics* website http://www.biometrics.tibs.org.

References

Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics* **34,** 86–102.

Callahan, C. M., Wolinsky, F. D., Stump, T. E., Nienaber, N. A., Hui, S. L., and Tierney, W. M. (1998). Mortality, symptoms, and functional impairment in late-life depression. *Journal of General Internal Medicine* **13,** 746–752.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*, 2nd edition. London: Chapman and Hall.

Chen, Y. H. (2002). Cox regression in cohort studies with validation sampling. *Journal of the Royal Statistical Society, Series B* **64,** 51–62.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Annals of Statistics* **22,** 732–746.

Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81,** 515–526.

Hu, C. and Lin D. Y. (2002). Cox regression with covariate measurement error. *Scandinavian Journal of Statistics* **29,** 637–655.

Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association* **97,** 318–327.

Huang, Y. and Wang C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association* **45,** 1209–1219.

Jiang, J. and Zhou, H. (2007). Additive hazard regression with auxilary covariates. *Biometrika* **94,** 359–369.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90,** 341–353.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. New York: John Wiley & Sons, Inc.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semi-parametric linear regression with censored data. *Journal of Computational and Graphical Statistics* **1,** 77–90.

Lin, D. Y., Robins, J. M., and Web, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83,** 381–393.

Nan, B. (2004). Efficient estimation for case-cohort studies. *Canadian Journal of Statistics* **32,** 403–419.

Nan, B., Kalbfleisch, J. D., and Yu, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missimg data. To appear in *Annals of Statistics*.

Nan, B., Yu, M., and Kalbfleisch, J. D. (2006). Censored linear regression for case-cohort studies. *Biometrika* **93,** 747–762.

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

Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association* **101,** 1085–1093.

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

Qi, L., Wang, C. Y., and Prentice, R. L. (2005). Weigthed estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* **100,** 1250–1263.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846–866.

Spiegelman, D., Rosner, B., and Logan, R. (2000). Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association* **95,** 51–61.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18,** 354–372.

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

Wang, C. Y. and Chen, H. Y. (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.

Wang, S. and Wang, C. Y. (2001). A note on kernel assisted estimators in missing covariate regression. *Statistics & Probability Letters* **55,** 439–449.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* **21,** 76–99.

Yu, M. and Nan, B. (2006). A hybrid Newton-type method for censored survival data using double weights in linear models. *Lifetime Data Analysis* **12,** 345–364.