Using the accelerated failure time model to analyze current status data with misclassified covariates

Baojiang Chen

Department of Biostatistics and Data Science,
University of Texas Health Science Center at Houston,
School of Public Health in Austin, Austin, Texas, 78701, U.S.A.
e-mail: baojiang.chen@uth.tmc.edu

Jing Qin

National Institute of Allergy and Infectious Diseases,
National Institutes of Health, MD 20892, U.S.A.
e-mail: jingqin@niaid.nih.gov

Ao Yuan

Department of Biostatistics, Bioinformatics and Biomathematics,
Georgetown University, Washington DC, 20057 USA
e-mail: ay312@georgetown.edu

Abstract: Current status data arise commonly in applications when there is only one feasible observation time to check if the failure time has occurred, but the exact failure time remains unknown. To accommodate the covariate effect on failure time, the accelerated failure time (AFT) model has been widely used to analyze current status data with the distribution of the failure time assumed to be specified or unspecified. In this paper, we consider a logistic regression with a misclassfied covariate from the current status observation scheme. A semiparametric AFT model was built to model current status data to eliminate the bias caused by this misclassification. This model is also robust to the misspecification of the failure time compared to the parametric AFT model, as we assume an unknown distribution of the failure time in the proposed model. Furthermore, incorporating the covariate effect on the failure time increases the flexibility of the model. Finally, we adapt the Expectation-Maximization algorithm for estimation, which guarantees the convergence of the estimate. Both theory and empirical studies show the consistency of the estimator.

MSC2020 subject classifications: Primary 62N99; secondary 62N02.
Keywords and phrases: AFT, current status data, EM algorithm, misclassification, pool adjacent violator algorithm (PAVA), semiparametric.

Received September 2019.

Contents

1. Introduction

In many applications, it is challenging to obtain an accurate measurement of a covariate. In general, the naive use of mismeasured covariates may lead to inconsistent estimators of covariate effects in a regression model (e.g., [6], [41], and [40]). There exists an extensive literature on methods to address the mismeasurement covariate problem, including likelihood-based methods (e.g., [31]), pseudo-likelihood methods (e.g., [23]), Bayesian methods (e.g., [15]), or the estimating equation method (e.g., [27]). To estimate the parameters of the mismeasurement distribution, all of these methods require either internal or external validation samples from replication studies ([6]).

There has been considerable work on methods for estimations under the framework of current status observation. Ayer et al. [1] discussed the nonparametric estimation of the survivor function using the pool adjacent violator algorithm (PAVA). Groeneboom, Maathuis and Wellner ([13, 14]) considered the current status data with competing risks, where they studied consistency, rates of convergence and limiting distribution of the maximum likelihood estimate (MLE). Rabinowitz, Tsiatis and Aragon [29] studied an accelerated failure time (AFT) model with interval-censored data. Jewell and Shiboski [19], Rossini and Tsiatis [30], Huang and Wellner [18], Lin, Oakes and Ying [24], and Shiboski [32], among others, considered the maximum likelihood inference on semiparametric transformation models for current status data. Maathuis and Hudgens [25] proposed nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. By inverting a Wald-type test for testing a null embedded proportional hazards model, Tian and Cai [33] introduced a novel method for an AFT model based on current status data or interval-censored data. Lam and Xue [21] and Cook et al. [9] developed semiparametric cure rate models to allow for the fact that not all individuals in the population are susceptible to such an event.

Zeng, Cook and Warkentin [43] considered a logistic regression analysis with a misclassified covariate from a current status observation scheme, from which they developed a likelihood-based approach for fitting regression models. Parametric and nonparametric estimates for the distribution of failure (seroconversion) time were developed. In particular, they only considered modeling the distribution of failure time without covariate effects. Because it is known that failure time may depend on the exposure of an antithrombotic drug, ignoring

the dependence of the failure time on these covariates can lead to invalid inference (See Section 6). Thus, in practice, it is desirable to accommodate some covariates when modeling the distribution of the failure time.

To overcome this problem, in this paper we consider a logistic regression analysis with a misclassified covariate from a current status observation scheme, in which a semiparametric AFT model for the failure time was built. In the semiparametric AFT model the distribution of the failure time is assumed to be unspecified. Few studies exist in the current status data literature regarding misclassified covariates that use a semiparametric AFT model. An adapted Expectation-Maximization (EM) algorithm ([10]) is developed using the PAVA to estimate the parameters in the AFT model. We have shown that, both in theory and in simulation studies, all estimators are consistent under some regularity conditions. This method is appealing in that it corrects the bias from covariate misclassification in the current status scheme. Furthermore, it accommodates the dependence of the failure time on covariates, leading to a more accurate and flexible model. Moreover, as a semiparametric method, it is robust to the misspecification of the distribution assumption of the failure time. Finally, this method is easily implemented in practice.

The remainder of this article is organized as follows. In Section 2, we introduce a motivating example. In Section 3, we introduce the notations and models. In Section 4, we describe an EM algorithm for the estimation of parameters. In Section 5, we study the asymptotic property of the estimator. Simulation studies are performed in Section 6 to assess the performance of the proposed method. Concluding remarks are made in Section 7.

2. Motivating example: A thrombosis study

Prophylaxis with antithrombotic heparin-based therapies is known to be highly effective in reducing the risk of thrombosis. In orthopedic surgery, administration of these therapies is now a standard practice ([38]). A study from four multi-center randomized trials in which patients underwent orthopedic surgery was conducted ([2, 11, 22, 34]). In this study, 3132 patients were randomized to receive enoxaparin in order to evaluate the treatment effect of enoxaparin vs. fondaparinux on reducing the risk of deep vein thrombosis (DVT) following hip or knee replacement. Some patients experienced serological responses when they underwent orthopedic surgery and were exposed to antithrombotic. The current research extends these results by investigating the association between this serological response and the subsequent risk of DVT.

Before surgery, patients are seronegative, and approximately 10 days following the surgery, antibodies develop among seroconverters. If all patients were tested approximately ten days after surgery, the seroconversion status would be known for all patients. However, patients were only tested at the time of hospital discharge, which in general is less than 10 days. According to Zeng, Cook and Warkentin [43], the median time from surgery to blood test is 5.9 days, and most seroconversions occur between 5 and 8 days. Therefore, negative test results may represent false negative classifications of the true seroconversion status

since these patients may have been tested before they developed the antibody response. Thus, to obtain a valid analysis, it is necessary to accommodate the timing of the test. In this situation, the seroconversion time is subject to a current status observation scheme. Because the seroconversion time may depend on exposure to an antithrombotic drug, it is desirable to accommodate some covariates when modeling the distribution of the seroconversion time. Here, we have developed a novel semiparametric method to analyze current status data with misclassified covariates.

3. Notation and models

Let $X_i = 1$ if individual i develops serological response and 0 otherwise, where $i = 1, \ldots, n$. Denote T_i as the serological response (failure) time if $X_i = 1$ and C_i is the random inspection time. We let $T_i = \infty$ if $X_i = 0$. Define $D_i = I(T_i \leq C_i)$. If $D_i = 1$, it means that the serological response occurs before or at the inspection time, hence $X_i = 1$; if $X_i = 0$ then $D_i = 0$; if $D_i = 0$, it is possible that the serological response time occurs after the blood test, hence D_i is a misclassified version of X_i. Let $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ be a vector of $p \times 1$ baseline covariates which are observed without measurement error.

For the serological response time T_i, we build a semiparametric AFT model for individuals who develop serological responses to accommodate for covariate effects. Specifically, we consider the model

$$\log T_i = \gamma_0 + \gamma^T Z_i + \epsilon_i \tag{3.1}$$

when $X_i = 1$, where γ_0 is an intercept, γ is an unknown $p \times 1$ vector of coefficients, ϵ_i has mean zero with an unspecified cumulative distribution function $G(s)$. It is noted that

$$
\begin{aligned}
P(D_i = 1 \mid X_i = 1, C_i, Z_i) &= P(T_i \leq C_i \mid X_i = 1, C_i, Z_i) \\
&= G(\log C_i - \gamma_0 - \gamma^T Z_i) \\
&\doteq F(\log C_i - \gamma^T Z_i), \tag{3.2}
\end{aligned}
$$

where $F(x) = G(x + \gamma_0)$. Since $G(\cdot)$ or $F(\cdot)$ is a nonparametric function, the model (3.1) is not identifiable unless we fix $F(\cdot)$ or $G(\cdot)$ at some point. As is commonly done in the literature (e.g., [37, 20]), we posit $\gamma_0 = 0$ so that the model (3.1) is identifiable.

Let Y_i denote the fully observed binary response of interest for individual i which takes the value 1 if individual i develops a DVT and 0 otherwise, $i = 1, \ldots, n$. Define $\pi_i = P(Y_i = 1 \mid X_i, Z_i)$, then we may consider building a logistic regression model:

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 X_i + \beta_2^T Z_i, \tag{3.3}$$

where $\beta = (\beta_0, \beta_1, \beta_2^T)^T$ is a vector of unknown regression parameters. Our primary interest is to estimate β.

If X_i is fully observed, we can then directly fit the model (3.3) to obtain an estimate β. However, the X_i value is not available when the inspection time is before the serological response time. A naive method is to use the current status indicator D_i instead of the true serological status X_i in the model (3.3); i.e., we fit the model:

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 D_i + \beta_2^T Z_i. \tag{3.4}$$

Simulation studies conducted in Section 6 demonstrate that the naive method using model (3.4) leads to substantive biases for β. Methods on handling measurement errors in the literature generally require internal or external validation samples or replication studies, which are not available in the thrombosis study. In this paper, we consider a simple and efficient estimation method for β under a current status observation scheme without needing the internal or external validation samples or replication studies by using a semiparametric AFT model.

To describe this method, we must specify a model for the probability of the serocoversion process X_i. Let $p_i = P(X_i = 1 \mid Z_i)$ denote the probability for the serocoversion process X_i. Similarly, we can build a logistic regression model for p_i

$$\log \frac{p_i}{1 - p_i} = \alpha_0 + \alpha_1^T Z_i. \tag{3.5}$$

Denote $\alpha = (\alpha_0, \alpha_1^T)^T$.

The observed data are $\{(Y_i, Z_i, D_i, C_i), i = 1, \ldots, n\}$. We assume that the inspection time C_i is independent of X_i and T_i given Z_i. Let $\theta = (\beta^T, \gamma^T, \alpha^T)^T$ be the vector of all the Euclidean parameters, where β is the parameter of interest, and γ and α are nuisance parameters.

4. Estimation procedure

With misclassified covariates, a popular technique to obtain the parameter estimation is to adapt the EM algorithm ([10]) by treating the true covariates as "missing data". The complete data density-mass function is

$$
\begin{aligned}
P(Y, X, D, Z, C) &= P(Y, X, D \mid Z, C)P(Z, C) \\
&= P(Y \mid X, D, Z, C)P(D \mid X, Z, C)P(X \mid Z, C)P(Z, C) \\
&= P(Y \mid X, Z)P(D \mid X, Z, C)P(X \mid Z)P(Z, C),
\end{aligned}
$$

where in the last step we used the fact that Y is independent of C, and that given (X, Z), D has no predictive value on Y ([43]), so that $P(Y \mid X, D, Z, C) = P(Y \mid X, Z)$, and that X is independent of C, so $P(X \mid Z, C) = P(X \mid Z)$. Note that $P(Z, C)$ contains no parameter of interest, so for inference of parameters of interest, we only need to consider the conditional density-mass function

$$P(Y, X, D \mid Z, C) = P(Y \mid X, Z)P(D \mid X, Z, C)P(X \mid Z).$$

Using the evaluations in the complete data conditional likelihood below, the conditional density of the observed incomplete data $(Y, D \mid Z, C)$ is

$$g(Y, D \mid Z, C, \theta, F) := P(Y, D \mid Z, C)$$

$$
\begin{aligned}
&= \; P(Y, D \mid X = 0, Z, C)P(X = 0 \mid Z, C) \\
&\quad + P(Y, D \mid X = 1, Z, C)P(X = 1 \mid Z, C) \\
&= \; P(Y \mid X = 0, D, Z)P(D \mid X = 0, Z, C)P(X = 0 \mid Z) \\
&\quad + P(Y \mid X = 1, D, Z)P(D \mid X = 1, Z, C)P(X = 1 \mid Z) \\
&= \; \Big(\frac{\exp(\beta_0 + \beta_2^T Z)}{1 + \exp(\beta_0 + \beta_2^T Z)}\Big)^Y \Big(\frac{1}{1 + \exp(\beta_0 + \beta_2^T Z)}\Big)^{1-Y} \frac{1}{1 + \exp(\alpha_0 + \alpha_1^T Z)} \\
&\quad + \Big(\frac{\exp(\beta_0 + \beta_1 + \beta_2^T Z)}{1 + \exp(\beta_0 + \beta_1 + \beta_2^T Z)}\Big)^Y \Big(\frac{1}{1 + \exp(\beta_0 + \beta_1 + \beta_2^T Z)}\Big)^{1-Y} \\
&\quad \times \{F(\log C - \gamma^T Z)\}^D \{1 - F(\log C - \gamma^T Z)\}^{1-D} \frac{\exp(\alpha_0 + \alpha_1^T Z)}{1 + \exp(\alpha_0 + \alpha_1^T Z)}.
\end{aligned}
$$

Let $g(Y, D, Z, C \mid \theta, F) = g(Y, D \mid Z, C, \theta, F)h(Z, C)$ be the joint density of the observed data, and $(\hat{\theta}_n, \hat{F}_n)$ be the maximum likelihood estimate (MLE) of the true (θ_0, F_0). Then

$$
(\hat{\theta}_n, \hat{F}_n) = \arg\max_{(\theta, F)} \prod_{i=1}^n g(Y_i, D_i \mid Z_i, C_i, \theta, F) = \arg\max_{(\theta, F)} \prod_{i=1}^n g(Y_i, Z_i, D_i, C_i, \mid \theta, F).
$$

(4.1)

The last equality is because $g(Y_i, Z_i, D_i, C_i, \mid \theta, F)$ and $g(Y_i, D_i \mid Z_i, C_i, \theta, F)$ are equivalent in terms of maximization over (θ, F). i.e., $(\hat{\theta}_n, \hat{F}_n)$ is the MLE of (θ_0, F_0) under the full likelihood of the observed data. However, parameter estimation based on the above incomplete data mixture model is not easy. A common practice is to compute $(\hat{\theta}_n, \hat{F}_n)$ via the iterative EM algorithm based on the complete data likelihood, which is a non-mixture model but with missing data.

The complete data likelihood for $\{(Y_i, X_i, D_i), i = 1, \ldots, n\}$ given $\{(Z_i, C_i), i = 1, \ldots, n\}$ is

$$
\begin{aligned}
L_n(\theta, F) &= \prod_{i=1}^n P(Y_i, X_i, D_i \mid Z_i, C_i) \\
&= \prod_{i=1}^n P(Y_i \mid X_i, Z_i; \theta) \times \prod_{i=1}^n P(D_i \mid X_i, Z_i, C_i; \theta) \times \prod_{i=1}^n P(X_i \mid Z_i; \theta) \\
&:= L_{1,n}(\beta) \times L_{2,n}(\gamma, F) \times L_{3,n}(\alpha),
\end{aligned}
$$

where

$$
L_{1,n}(\beta) = \prod_{i=1}^n P(Y_i \mid X_i, Z_i; \theta),
$$

with $P(Y_i \mid X_i, Z_i; \theta)$ specified in (3.3),

$$
L_{2,n}(\gamma, F) = \prod_{i=1}^n P(D_i \mid X_i, Z_i, C_i; \theta),
$$

and

$$L_{3,n}(\alpha) = \prod_{i=1}^{n} P(X_i \mid Z_i; \theta).$$

Note that $X_i = 0$ implies $D_i = 0$, thus $P(D_i = 0 \mid X_i = 0, Z_i, C_i) = 1$. Since $P(D_i = 1 \mid X_i = 0, Z_i, C_i) = 0$, to make $L_2(\gamma, F)$ non-zero, we define $0^0 = 1$, so that $P(D_i = 1 \mid X_i = 0, Z_i, C_i)$ is written as $F(\log C_i - \gamma^T Z_i)^{X_i} = 0^0 = 1$. Using (3.2), the second part is $L_{2,n}(\gamma, F) = \prod_{i=1}^{n} P(D_i = 1 \mid X_i, Z_i, C_i)^{D_i} \times P(D_i = 0 \mid X_i, Z_i, C_i)^{1-D_i} = \prod_{i=1}^{n} \{F(\log C_i - \gamma^T Z_i)\}^{X_i D_i} \{1 - F(\log C_i - \gamma^T Z_i)\}^{X_i(1-D_i)}$, which involves the unknown parameter γ and unknown nonparametric function F. Let $n_1 = \sum_{i=1}^{n} X_i$, and $S_1 = \{i : X_i = 1, 1 \le i \le n\}$ with $|S_1| = n_1$, then $L_{2,n}(\gamma, F)$ is re-witten as

$$L_{2,n}(\gamma, F) = \prod_{i \in S_1} \{F(\log C_i - \gamma^T Z_i)\}^{D_i} \{1 - F(\log C_i - \gamma^T Z_i)\}^{1-D_i}.$$

The third part $L_{3,n}(\alpha) = \prod_{i=1}^{n} P(X_i = 1 \mid Z_i)^{X_i} \{1 - P(X_i = 1 \mid Z_i)\}^{1-X_i} = \prod_{i=1}^{n} p_i^{X_i} (1 - p_i)^{1-X_i}$, with p_i specified in (3.5).

In practice, the X is subject to missingness, so the observed data are (Y, Z, D, C). In the EM algorithm, we iterate the E and M steps until convergence. Specifically, the estimation of the parameter θ and F is updated at the $(t+1)$st through the following two steps:

E-step: Compute

$$Q(\theta, F; \hat{\theta}^{(t)}, \hat{F}^{(t)}) = E[\log\{L_n(\theta, F; Y, X, Z, D, C)\} \mid Y, Z, D, C; \hat{\theta}^{(t)}, \hat{F}^{(t)}].$$

M-step: Maximize $Q(\theta, F; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to θ and F to obtain $\hat{\theta}^{(t+1)}$ and $\hat{F}^{(t+1)}$.

The E-step requires the calculation of the conditional expectation given the observed data, which can be written to three parts

$$Q(\theta, F; \hat{\theta}^{(t)}, \hat{F}^{(t)}) = Q_1(\beta; \hat{\theta}^{(t)}, \hat{F}^{(t)}) + Q_2(\gamma, F; \hat{\theta}^{(t)}, \hat{F}^{(t)}) + Q_3(\alpha; \hat{\theta}^{(t)}, \hat{F}^{(t)}),$$
$$(4.2)$$

where each component represents the conditional expectation of the logarithm of each of the three parts in the complete data likelihood. Since these three parts do not share the same parameters, we can update estimates $\hat{\beta}^{(t+1)}$, $\hat{\gamma}^{(t+1)}$, $\hat{F}^{(t+1)}$ and $\hat{\alpha}^{(t+1)}$ by maximizing the Q-functions separately. For updating $\hat{\beta}^{(t+1)}$, it is noted that

$$Q_1(\beta; \hat{\theta}^{(t)}, \hat{F}^{(t)})$$
$$= \sum_{i=1}^{n} \Big[D_i \log\{P(Y_i \mid X_i = 1, Z_i)\}$$
$$+ (1 - D_i) \Big[\sum_{X_i=0}^{1} (\hat{w}_i^{(t)})^{X_i} (1 - \hat{w}_i^{(t)})^{1-X_i} \log\{P(Y_i \mid X_i, Z_i)\} \Big] \Big], \quad (4.3)$$

where $\hat{w}_i^{(t)} = P(X_i = 1 \mid Y_i, D_i = 0, Z_i, C_i; \hat{\theta}^{(t)}, \hat{F}^{(t)})$. It is easy to specify that $\hat{w}_i^{(t)} = A/B$, where $A = P(Y_i \mid X_i = 1, D_i = 0, Z_i; \hat{\beta}^{(t)})P(D_i = 0 \mid X_i = 1, Z_i, C_i; \hat{\gamma}^{(t)}, \hat{F}^{(t)})P(X_i = 1 \mid Z_i; \hat{\alpha}^{(t)})$ and $B = \sum_{X_i=0}^{1} P(Y_i \mid X_i, D_i = 0, Z_i; \hat{\beta}^{(t)})P(D_i = 0 \mid X_i, Z_i, C_i; \hat{\gamma}^{(t)}, \hat{F}^{(t)})P(X_i \mid Z_i; \hat{\alpha}^{(t)})$. Maximizing $Q_1(\beta; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to β can be easily achieved using standard statistical software by creating a pseudo-data set similar to [43].

For updating $\hat{\gamma}^{(t+1)}$ and $\hat{F}^{(t+1)}$, the conditional expectation $Q_2(\gamma, F; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ takes the form

$$
\begin{aligned}
Q_2(\gamma, F; \hat{\theta}^{(t)}, \hat{F}^{(t)}) &= E[\log\{L_{2,n}(\gamma, F)\} \mid Y, D, Z, C; \hat{\theta}^{(t)}] \\
&= \sum_{i=1}^{n} D_i \log\{F(\log C_i - \gamma^T Z_i)\} \\
&\quad + (1 - D_i)\hat{w}_i^{(t)} \log\{1 - F(\log C_i - \gamma^T Z_i)\}.
\end{aligned}
$$

Maximizing $Q_2(\gamma, F; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to γ and F may not be easy, as F is a nonparametric non-decreasing function. Here we describe a two-stage estimation procedure:

Stage 1: For a given γ, we maximize $Q_2(\gamma, F; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to F under the constraint that $F(a_1) \leq \ldots \leq F(a_n)$ if $a_1 \leq \ldots \leq a_n$, $a_i = \log C_i - \gamma^T Z_i$. This maximization can be easily achieved by regressing $D_i/(D_i + (1 - D_i)\hat{w}_i^{(t)})$ on $\log C_i - \gamma^T Z_i$ with weight $D_i + (1 - D_i)\hat{w}_i^{(t)}$ using the PAVA. Denote the estimate $\hat{F}(\gamma)$.

Stage 2: We maximize $Q_2(\gamma, \hat{F}(\gamma); \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to γ to obtain $\hat{\gamma}^{(t+1)}$. This maximization can be achieved using existing software, such as the optim() function in R ([28]).

Finally, for updating $\hat{\alpha}^{(t+1)}$, the conditional expectation $Q_3(\alpha; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ takes the form

$$
\begin{aligned}
Q_3(\alpha; \hat{\theta}^{(t)}, \hat{F}^{(t)}) &= E[\log\{L_{3,n}(\alpha)\} \mid Y, D, Z, C; \hat{\theta}^{(t)}, \hat{F}^{(t)}] \\
&= \sum_{i=1}^{n} \left[\hat{w}_i^{(t)} \log p_i + (1 - \hat{w}_i^{(t)}) \log(1 - p_i)\right].
\end{aligned}
$$

Maximizing $Q_3(\alpha; \hat{\theta}^{(t)}, \hat{F}^{(t)})$ with respect to α can be easily achieved as it has the form of a binary log-likelihood with data $\hat{w}_i^{(t)}$.

In fact, our model is equivalent to a mixture model. It is well known that there is no general convergence theorem on the EM algorithm, and generally, it converges to a local maxima ([39], Theorem 3). If the underlying model has the concave property, the EM algorithm converges to the global MLE. On the other hand, if the underlying model is not concave, only local convergence is possible. Thus in an application, one needs to apply the EM algorithm with different starting values to get possible different local stationary points and compare the log-likelihood at these stationary points to find the global maxima.

Our algorithm is a semiparametric version of the EM algorithm. The semiparametric and nonparametric EM algorithms are used widely in the literature, such as in [26, 5, 16, 12], and see the argument there for the convergence of such algorithm ([12], p.67-68). Chen, Zhang and Davidian [8] applied the EM algorithm to a semiparametric random effects model, while Bordes, Chauveau and Vandekerknove [4] applied the EM algorithm to a semiparametric mixture model, using simulation studies to justify the convergence of the algorithm. Finally, Balan and Putter [3] developed an R-package of EM algorithm for semiparametric shared frailty models.

5. Consistency of the estimators

Let $g_0(y, z, d, c \mid \theta, F) = g_0(y, d \mid z, c, \theta, F)h_0(z, c)$ be the true joint mass-density of (Y, Z, D, C), Θ be the domain of θ and $\mathcal{F}$ be the collection of all the distribution functions F of T, and P be the probability measure of F_0. To show the consistency of the estimators, we need the following regularity conditions:

(C0). The inspection time C_i is independent of X_i and T_i given Z_i.
(C1). The range of (β, α) is bounded.
(C2). The support of Z is bounded.

We show the convergence of the estimators based on the following theorems:

Theorem 5.1. *Assume (C0)-(C2), then as $n \to \infty$,*

$$\hat{\theta}_n \overset{a.s.}{\to} \theta_0, \quad \text{and} \quad \int \left|\hat{F}_n(x) - F_0(x)\right| dF_0(x) \overset{a.s.}{\to} 0,$$

Condition (C0) is commonly assumed for this type of problem, such as in [43].

With interval censoring data (type I or current status data) for the Cox proportional hazards model, the regression parameters can be efficiently estimated with rate $n^{1/2}$ (see, for example, [17]). However, using the same data for the standard semiparametric accelerated failure time model, the asymptotic distribution for the estimated regression parameters is an open problem: it is not clear whether or not $\hat{\theta}_n$ is $n^{1/2}$ convergence, or whether it has a normal limiting distribution (see [18], Section 3.3.2). The reason may be because for the current status model, the nonparametric maximum likelihood estimate $\hat{F}_n$ of F has a convergence rate only of $n^{1/3}$, and the process $\{n^{1/3}(\hat{F}_n(\cdot) - F_0(\cdot)) : 0 \le t \le \tau\}$ is *not tight* in $D[0, \tau]$, and *does not converge weakly as a process* ([18], Section 2.2). Hence the profile log-likelihood $l_n(\theta, \hat{F}_n(\cdot \mid \theta))$ is not smooth enough as a function of θ. Chen et al. [7] encountered another similar challenge in density ratio models. Our model is more complicated than the standard AFT model, hence the asymptotic distribution of $\hat{\theta}_n$ in our model is currently not clear. In application, we suggest using a bootstrap method to obtain the variability of the proposed estimators.

Next we study the possibility of rate $\sqrt{n}$ estimability of our model. Let the original data be $\mathbf{V}^0 \sim Q \in \mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$. In practice, sometimes the original data $\mathbf{V}^0$ is not completely observed; instead only $\mathbf{V} = T(\mathbf{V}^0)$ is observed, for some known map T. The resulting model for $\mathbf{V}$ is $\mathcal{P} = \{P = Q_\theta T^{-1} : \theta \in \Theta\}$. Denote $L_2^0(Q) = \{g : \int g(\mathbf{v})Q(d\mathbf{v}) = 0, \quad \int g^2(\mathbf{v})Q(d\mathbf{v}) < \infty\}$, and denote $L_2^0(P)$ similarly. Let $a(\mathbf{v}^0) = \mathbf{i}(\mathbf{v}^0 \mid \theta, Q)$ be the score for θ in model Q, and $\mathbf{i}(\mathbf{v} \mid \theta, P)$ be that in model P, by Proposition 1.1 in [12],

$$\mathbf{i}(\mathbf{v} \mid \theta, P) = E\big(a(\mathbf{V}^0) \mid T(\mathbf{V}^0) = \mathbf{v}\big).$$

For missing data, the above observation is extended to the following notion of *score operator* $\mathbf{i} : L_2^0(Q) \to L_2^0(P)$ (see [12]), where

$$\mathbf{i}a = E(a(\mathbf{V}^0) \mid T(\mathbf{V}^0) = \mathbf{v}), \quad \forall\, a \in L_2^0(Q).$$

It is a bounded linear operator and has an adjoint $\mathbf{i}^* : L_2^0(P) \to L_2^0(Q)$ determined by

$$< b, \mathbf{i}a >_{L_2(P)} = < \mathbf{i}^*b, a >_{L_2(Q)}, \quad \forall\, a \in L_2^0(Q), \quad b \in L_2^0(P),$$

where $< b, g >_{L_2(P)} = \int b(\mathbf{v})g(\mathbf{v})P(d\mathbf{v})$ and $< b, g >_{L_2(Q)} = \int b(\mathbf{v}^0)g(\mathbf{v}^0)Q(d\mathbf{v}^0)$. Also, $\mathbf{i}^*b = E(b(\mathbf{V}) \mid \mathbf{V}^0), \forall b \in L_2^0(P)$. The *information operator* is defined as $\mathbf{i}^*\mathbf{i} : L_2^0(Q) \to L_2^0(Q)$.

The information operator can provide a guideline for the rate $\sqrt{n}$ estimability of the model parameters. If it is boundedly invertible, the parameter is likely to be $\sqrt{n}$ estimable; otherwise, it may not. Yuan, Xu and Zheng [42] used this method to evaluate the rate $\sqrt{n}$ estimability of parameters in some incomplete data models.

Below we investigate model (4.1) for the rate $\sqrt{n}$ estimability for functionals of F, by evaluating its information operator.

Theorem 5.2. *For model (4.1), the information operator $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible.*

Thus many functionals of the model (4.1), including θ, may not be $\sqrt{n}$ estimable. For the MLE $\hat{\theta}_n$, although its asymptotic distribution and exact convergence rate are unknown, we can get an upper bound on its convergence rate, as shown in the next Theorem. Let $\|\hat{\theta}_n - \theta_0\| = \big(\sum_{j=1}^d (\hat{\theta}_{n,j} - \theta_{0,j})^2\big)^{1/2}$, where $d = \dim(\theta)$, and $\|\hat{F}_n - F_0\| = \int |\hat{F}_n(x) - F_0(x)|dF_0(x)$.

Theorem 5.3. *Assume (C0)-(C2), then*

$$\|\hat{\theta}_n - \theta_0\| + \|\hat{F}_n - F_0\| = O_p\big(n^{-1/3}\big).$$

For presentational continuity, we have relegated all the long proofs of Theorems 5.1 to 5.3 to the Appendix.

6. Simulation studies

In this section, we assess the performance of the proposed method by comparing the following methods: (i) the naïve method by using a logistic regression model and including the current status indicator D_i as a covariate (*Naive*), i.e., we fit the model (3.4); (ii) the proposed method by fitting a semiparametric AFT model for the seroconversion time (*Proposed*); (iii) the method by fitting a correctly specified parametric AFT model for the seroconversion time, which serves as a benchmark for comparisons (*Parametric*); and (iv) the method developed by [43] using a nonparametric estimate of the seroconversion time distribution (*ZCW*).

We posit that $Z_i \sim \text{Bin}(1, 0.5)$, $X_i \sim \text{Bin}(1, p_i)$, $p_i = \text{expit}(\alpha_0 + \alpha_1 Z_i)$, $Y_i \sim \text{Bin}(1, \pi_i)$ with $\pi_i = \text{expit}(\beta_0 + \beta_1 X_i + \beta_2 Z_i)$, where $\text{expit}(a) = \exp(a)/(1 + \exp(a))$. The underlying times to seroconversion for individuals who are susceptible are simulated under the AFT model

$$\log T_i = \gamma_0 + \gamma_1 Z_i + \epsilon_i.$$

We consider two distributions: (a) T_i follows a log-normal distribution with $E[\log T_i] = \gamma_0 + \gamma_1 Z_i$, and $\text{Var}(\log T_i) = \sigma^2$; and (b) T_i follows a Weibull distribution with a shape a, and a scale $b_i = \exp(-\gamma_0 - \gamma_1 Z_i)$. For the inspection process, we posit that C_i follows the Gamma distribution with a shape 1, and a rate ξ. In the simulation, we posit that $\alpha_0 = 0, \alpha_1 = 1, \beta_0 = 0, \beta_1 = 1, \beta_2 = 1$, $\gamma_0 = 1$, $\gamma_1 = 1$ or 0.1, $\sigma^2 = 1$, $a = 1$, and we posit different values of ξ to induce different proportions of patients who develop serological response before inspection time, i.e., $P(T < C \mid X = 1)$. We consider a sample size $n = 2000$ and 200, and for each model configuration, we conduct 2000 simulations.

Tables 1 and 2 report the empirical biases and standard deviations where T follows log-normal distribution for $\gamma_1 = 1$ and $\gamma_1 = 0.1$, respectively. To evaluate the efficiency of the proposed method compared to the parametric method, we also report the relative efficiency (RE) for the proposed method, which is defined as the ratio of the empirical standard deviations between the proposed estimate and the parametric estimate. The probability $P(T < C \mid X = 1) = 0.2$ and 0.4 by assuming $\xi = 0.5$ and 0.2. It can be seen that the naive analysis yields substantial biases for all three parameters. The ZCW's method also yields estimators with substantial biases for a larger γ_1 value (e.g., $\gamma_1 = 1$), especially for β_1, β_2 and α_1; as is expected, the ZCW's method yields estimators with smaller biases for a smaller γ_1 value (e.g., $\gamma_1 = 0.1$). The biases of the naive and ZCW's estimators also depend on the probability of $P(T < C \mid X = 1)$ (or ξ), where as $P(T < C \mid X = 1)$ decreases (or ξ increases), the bias increases. When $P(T < C \mid X = 1)$ is large or ξ is small, the ZCW method yields comparable estimates as to the proposed method. The proposed method and the correctly specified parametric method yield negligible biases for all kinds of parameter estimates. Although the proposed estimate of β is less efficient than the parametric estimate, the efficiency loss is limited.

Tables 3 and 4 report the empirical biases and standard deviations where T follows the Weibull distribution for $\gamma_1 = 1$ and $\gamma_1 = 0.1$, respectively. In this

TABLE 1
Empirical biases and standard deviations where T follows log-normal distribution: $\gamma_1 = 1$

ξ	Parameter	Naive		Proposed			Parametric		ZCW	
		Bias	SD	Bias	SD	RE	Bias	SD	Bias	SD
$n = 2000$										
0.50	β_0	0.391	0.070	0.009	0.221	1.15	-0.017	0.192	0.020	0.194
0.50	β_1	-0.426	0.186	0.004	0.325	1.09	0.030	0.299	-0.112	0.288
0.50	β_2	0.238	0.111	0.024	0.185	1.12	0.015	0.165	0.439	0.169
0.50	α_0			-0.009	0.285	1.06	0.001	0.268	0.111	0.287
0.50	α_1			-0.005	0.454	1.04	0.001	0.435	-1.991	0.218
0.50	γ_0						-0.012	0.174		
0.50	γ_1			-0.006	0.196	1.07	0.002	0.184		
0.50	σ						-0.007	0.089		
0.20	β_0	0.314	0.075	0.015	0.138	1.16	-0.014	0.119	0.038	0.118
0.20	β_1	-0.382	0.135	-0.012	0.211	1.10	0.021	0.192	-0.128	0.177
0.20	β_2	0.246	0.111	0.010	0.143	1.06	-0.002	0.135	0.313	0.113
0.20	α_0			-0.012	0.190	1.17	0.017	0.162	0.082	0.160
0.20	α_1			-0.019	0.369	1.04	0.020	0.355	-1.372	0.136
0.20	γ_0						0.004	0.146		
0.20	γ_1			-0.002	0.187	1.07	0.001	0.175		
0.20	σ						-0.002	0.088		
$n = 200$										
0.50	β_0	0.404	0.226	-0.017	0.707	1.13	0.019	0.625	-0.182	0.913
0.50	β_1	-0.286	1.350	0.008	1.175	1.14	-0.033	1.035	0.203	1.613
0.50	β_2	0.249	0.357	0.046	0.726	1.20	0.017	0.605	0.655	0.774
0.50	α_0			0.011	0.859	1.08	-0.008	0.799	0.448	0.660
0.50	α_1			0.017	1.329	1.07	-0.008	1.246	-2.130	0.843
0.50	γ_0						-0.006	0.488		
0.50	γ_1			-0.011	0.664	1.03	-0.011	0.642		
0.50	σ						-0.012	0.369		
0.20	β_0	0.314	0.233	-0.017	0.544	1.35	0.020	0.403	-0.104	0.552
0.20	β_1	-0.332	0.445	0.021	0.725	1.05	-0.029	0.693	0.051	0.729
0.20	β_2	0.257	0.366	0.011	0.492	1.07	0.003	0.460	0.366	0.489
0.20	α_0			0.011	0.502	1.10	-0.041	0.456	0.302	0.462
0.20	α_1			0.014	1.204	1.08	-0.017	1.117	-1.383	0.524
0.20	γ_0						-0.008	0.420		
0.20	γ_1			-0.008	0.606	1.09	-0.008	0.555		
0.20	σ						-0.005	0.301		

case, we consider a smaller probability of $P(T < C \mid X = 1)$. We posit $\xi = 6$ and 3 so that $P(T < C \mid X = 1) = 0.04$ and 0.07. The conclusions were similar as those found in the log-normal model scenario.

In summary, the proposed method is robust to the distribution assumption for the seroconversion time. The ZCW's method, by ignoring the covariate effect in the distribution of seroconversion time, yields substantial biases for the parameters of interest when the survival time is covariate dependent. When the dependency is small, the ZCW's method may yield less biased or comparable

ξ	Parameter	Naive		Proposed			Parametric		ZCW	
		Bias	SD	Bias	SD	RE	Bias	SD	Bias	SD
$n = 2000$										
0.50	β_0	0.396	0.068	0.030	0.345	1.01	0.020	0.342	-0.033	0.616
0.50	β_1	-0.463	0.152	-0.019	0.546	1.01	-0.013	0.541	0.066	0.776
0.50	β_2	0.173	0.107	0.029	0.347	1.01	0.029	0.342	0.071	0.294
0.50	α_0			-0.013	0.461	1.01	-0.009	0.458	-0.038	0.483
0.50	α_1			-0.009	0.748	0.92	-0.003	0.815	-0.225	0.690
0.50	γ_0						-0.031	0.334		
0.50	γ_1			-0.011	0.272	0.76	0.015	0.358		
0.50	σ						-0.019	0.185		
0.20	β_0	0.310	0.072	-0.015	0.297	1.01	-0.011	0.293	0.025	0.348
0.20	β_1	-0.364	0.120	0.005	0.452	1.19	0.005	0.381	-0.014	0.465
0.20	β_2	0.146	0.107	-0.002	0.266	1.00	0.010	0.265	0.047	0.225
0.20	α_0			0.008	0.331	1.01	0.003	0.328	-0.060	0.268
0.20	α_1			0.010	0.575	1.05	0.031	0.548	-0.167	0.336
0.20	γ_0						-0.033	0.299		
0.20	γ_1			-0.012	0.263	0.77	0.017	0.341		
0.20	σ						-0.016	0.177		
$n = 200$										
0.50	β_0	0.392	0.227	-0.038	1.157	1.09	-0.031	1.064	-0.043	1.532
0.50	β_1	-0.377	0.915	0.023	1.999	1.04	0.018	1.916	0.099	1.882
0.50	β_2	0.220	0.371	-0.038	1.283	1.08	0.035	1.190	0.061	1.493
0.50	α_0			-0.021	1.569	1.15	-0.011	1.369	-0.030	1.556
0.50	α_1			0.011	2.033	0.94	-0.007	2.159	-0.098	1.900
0.50	γ_0						-0.035	1.214		
0.50	γ_1			0.015	1.234	0.91	-0.015	1.359		
0.50	σ						-0.020	0.620		
0.20	β_0	0.314	0.229	-0.013	1.055	1.06	-0.008	0.991	-0.021	1.403
0.20	β_1	-0.340	0.416	0.009	1.475	1.05	0.008	1.402	0.017	1.598
0.20	β_2	0.176	0.361	-0.015	1.245	1.07	0.017	1.159	0.042	1.184
0.20	α_0			0.016	1.306	1.02	-0.012	1.286	0.029	1.141
0.20	α_1			0.033	1.673	1.02	0.031	1.647	-0.093	1.366
0.20	γ_0						-0.035	1.143		
0.20	γ_1			0.015	1.199	0.96	-0.011	1.252		
0.20	σ						-0.015	0.607		

estimate to the proposed method. The naive method using the current status indicator D_i as the covariate also yields substantial biases for the parameters of interest. The R code for the simulation is available upon the request from the authors.

7. Concluding remarks

We have proposed a semiparametric AFT model to address a misclassified co-variate in the current status data scheme. This method overcomes the limitation

TABLE 3
Empirical biases and standard deviations where T follows Weibull distribution: $\gamma_1 = 1$

ξ	Parameter	Naive		Proposed			Parametric		ZCW	
		Bias	SD	Bias	SD	RE	Bias	SD	Bias	SD
$n = 2000$										
6	β_0	0.459	0.066	0.018	0.247	1.06	0.017	0.233	-0.270	0.549
6	β_1	-0.468	0.401	0.008	0.598	1.06	-0.005	0.563	0.155	0.828
6	β_2	0.197	0.112	-0.009	0.173	1.08	0.002	0.160	0.564	0.414
6	α_0			-0.012	0.109	1.25	0.011	0.087	0.807	0.308
6	α_1			0.005	0.183	1.21	-0.002	0.151	-1.973	0.702
6	γ_0						0.076	0.547		
6	γ_1			-0.017	0.423	1.10	0.016	0.383		
6	a						0.027	0.202		
3	β_0	0.444	0.070	0.022	0.204	1.07	-0.014	0.190	-0.195	0.351
3	β_1	-0.488	0.284	-0.009	0.426	1.04	0.005	0.409	0.139	0.546
3	β_2	0.206	0.110	-0.007	0.155	1.08	0.006	0.143	0.516	0.252
3	α_0			-0.014	0.121	1.48	0.015	0.082	0.742	0.239
3	α_1			0.018	0.165	1.15	-0.013	0.144	-2.010	0.477
3	γ_0						0.015	0.302		
3	γ_1			-0.009	0.324	1.10	0.008	0.292		
3	a						0.011	0.145		
$n = 200$										
6	β_0	0.460	0.210	0.019	0.726	1.17	0.020	0.619	-0.132	0.962
6	β_1	1.951	5.898	0.011	1.853	1.02	-0.014	1.809	0.365	1.845
6	β_2	0.209	0.361	0.001	0.587	1.23	0.005	0.476	0.421	0.736
6	α_0			-0.011	0.344	1.09	-0.014	0.317	0.388	0.646
6	α_1			0.005	0.605	1.19	0.003	0.508	-1.183	1.309
6	γ_0						0.066	2.090		
6	γ_1			-0.022	1.357	1.06	0.018	1.286		
6	a						0.029	0.687		
3	β_0	0.442	0.219	0.023	0.763	1.26	0.013	0.606	-0.195	1.142
3	β_1	0.316	3.479	0.018	1.508	1.11	-0.013	1.363	0.404	2.090
3	β_2	0.212	0.354	0.002	0.509	1.04	0.005	0.488	0.583	0.963
3	α_0			-0.006	0.399	1.35	-0.019	0.295	0.579	0.680
3	α_1			0.023	0.503	1.06	0.015	0.476	-1.711	1.387
3	γ_0						0.011	1.116		
3	γ_1			-0.013	1.162	1.14	0.007	1.017		
3	a						0.013	0.458		

of the method of [43] by accommodating the dependence of covariates on the distribution of the seroconversion time. The proposed method is robust to the distribution assumption of the suroconversion time by assuming a semiparametric AFT model. Both theory and simulation studies show that the proposed method yields consistent estimates of all parameters.

The question being studied can also fit into a general framework of missing data while missing of a covariate is at random, and the distribution of the observed values depends on the study time and other covariates. Though in

TABLE 4
Empirical biases and standard deviations where T follows Weibull distribution: $\gamma_1 = 0.1$

ξ	Parameter	Naive		Proposed			Parametric		ZCW	
		Bias	SD	Bias	SD	RE	Bias	SD	Bias	SD
$n = 2000$										
6	β_0	0.456	0.067	0.025	0.442	1.04	0.018	0.426	0.036	0.613
6	β_1	-0.511	0.297	-0.027	1.075	1.08	-0.024	0.997	0.035	1.853
6	β_2	0.191	0.110	0.016	0.348	1.01	0.014	0.344	0.018	0.516
6	α_0			-0.036	0.188	1.09	0.033	0.172	0.055	0.383
6	α_1			-0.042	0.384	1.26	0.050	0.305	-0.070	1.246
6	γ_0						0.018	1.150		
6	γ_1			0.003	0.477	1.15	-0.005	0.414		
6	a						0.030	0.338		
3	β_0	0.436	0.065	0.015	0.429	1.03	0.012	0.416	-0.028	0.595
3	β_1	-0.510	0.208	-0.014	0.994	1.09	-0.011	0.911	0.027	1.713
3	β_2	0.186	0.107	0.016	0.318	1.04	0.018	0.307	0.067	0.505
3	α_0			-0.018	0.168	1.03	-0.018	0.163	0.028	0.310
3	α_1			-0.042	0.312	1.15	0.043	0.272	-0.092	0.755
3	γ_0						0.019	0.777		
3	γ_1			-0.009	0.457	1.14	0.006	0.401		
3	a						0.019	0.333		
$n = 200$										
6	β_0	0.452	0.202	0.024	1.540	1.18	0.017	1.306	-0.049	1.705
6	β_1	1.984	5.928	0.032	3.956	1.16	0.026	3.393	1.076	5.154
6	β_2	0.221	0.354	0.024	1.361	1.01	0.018	1.350	0.361	2.008
6	α_0			-0.037	0.619	1.12	-0.035	0.555	0.261	0.980
6	α_1			0.071	1.224	1.15	0.034	1.068	-0.370	2.490
6	γ_0						0.025	3.171		
6	γ_1			0.007	1.511	1.17	0.005	1.287		
6	a						0.042	0.969		
3	β_0	0.435	0.214	0.018	1.639	1.14	0.017	1.441	-0.271	1.759
3	β_1	0.169	3.231	0.016	3.898	1.11	0.019	3.521	0.896	4.058
3	β_2	0.222	0.356	0.018	1.293	1.13	0.016	1.144	0.180	2.252
3	α_0			-0.064	0.591	1.17	-0.029	0.505	0.062	0.789
3	α_1			0.096	0.999	1.04	0.054	0.965	0.240	3.978
3	γ_0						-0.021	2.458		
3	γ_1			0.016	1.435	1.07	0.011	1.339		
3	a						0.027	1.137		

this paper, we considered a binary outcome using a logistic regression model, this method can be applied to different type of outcomes using generalized linear models. The estimation procedure introduced in Section 4 can be easily adapted to estimate parameters in the generalized linear model.

Appendix

This Appendix contains a detailed proof of Theorems 5.1 to 5.3 in Section 5.

Proof of Theorem 5.1

Let $S = (Y, D, Z, C)$. Recall the density for the observed data is $g(s \mid \theta, F) = g(y, d \mid z, c, \theta, F)h(z, c)$. Denote $\hat{g}_n(\cdot) = g(\cdot \mid \hat{\theta}_n, \hat{F}_n)$, and $g_0(\cdot) = g(\cdot \mid \theta_0, F_0)$. Let $P_n(\cdot)$ be the empirical distribution function of (Y_i, D_i, Z_i, C_i) $(i = 1, ..., n)$ and P be that of (Y, D, Z, C). Let Θ be the space for θ, $\mathcal{F}$ be the collection of all distribution functions F of the true response time T,

$$
\begin{aligned}
M(\theta, F) &= \int \log \frac{g(s \mid \theta, F)}{g(s \mid \theta_0, F_0)} dP(s) \\
&= \int g(y, d, z, c \mid \theta_0, F_0) \log \frac{g(y, d \mid z, c, \theta, F)}{g(y, d \mid z, c, \theta_0, F_0)} ds
\end{aligned}
$$

and

$$
M_n(\theta, F) = \int \log \frac{g(s \mid \theta, F)}{g(s \mid \theta_0, F_0)} dP_n(s) = \sum_{i=1}^{n} \frac{1}{n} \log \frac{g(Y_i, D_i \mid Z_i, C_i, \theta, F)}{g(Y_i, D_i \mid Z_i, C_i, \theta_0, F_0)}.
$$

Then

$$
(\theta_0, F_0) = \arg \max_{(\theta, F) \in (\Theta, \mathcal{F})} M(\theta, F), \quad \text{and} \quad (\hat{\theta}_n, \hat{F}_n) = \arg \max_{(\theta, F) \in (\Theta, \mathcal{F})} M_n(\theta, F).
$$

Let $a(s \mid \theta, F) = \log \big(g(y, d \mid z, c, \theta, F)/g(y, d \mid z, c, \theta_0, F_0)\big)$, $\mathcal{A} = \{a(s \mid \theta, F) : \theta \in \Theta, F \in \mathcal{F}\}$, $\mathcal{G} = \{g(y, d \mid z, c, \theta, F) : \theta \in \Theta, F \in \mathcal{F}\}$. For any function q, let $\|q\|_{L_1(P)} = \int |q(s)|P(ds)$ be its $L_1(P)$ norm. Let $N_{[\,]}(\epsilon, \mathcal{G}, L_1(P))$ be the minimum number of ϵ-brackets to cover $\mathcal{G}$ under norm $\|\cdot\|_{L_1(P)}$, and similarly for $N_{[\,]}(\epsilon, \mathcal{G}_0, L_1(P))$. We need to evaluate $N_{[\,]}(\epsilon, \mathcal{A}, L_1(P))$.

Note that with (C1) and (C2), there is $c_0 > 0$ such that for all (y, d, z, c) and all (θ, F),

$$
\begin{aligned}
&g(y, d \mid z, c, \theta, F) \geq \\
&\Big(\frac{\exp(\beta_0 + \beta_2^T z)}{1 + \exp(\beta_0 + \beta_2^T z)}\Big)^y \Big(\frac{1}{1 + \exp(\beta_0 + \beta_2^T z)}\Big)^{1-y} \frac{1}{1 + \exp(\alpha_0 + \alpha_1^T z)} \geq c_0
\end{aligned}
$$

or

$$
\sup_{(y, d, z, c), (\theta, F) \in (\Theta, \mathcal{F})} \frac{1}{g(y, d \mid z, c, \theta, F)} \leq \frac{1}{c_0} := C < \infty.
$$

So for any $a(\cdot \mid \theta_1, F_1), a(\cdot \mid \theta_2, F_2) \in \mathcal{A}$, for some $(\bar{\theta}, \bar{F})$ between (θ_1, F_1) and (θ_2, F_2),

$$
\begin{aligned}
&\big|a(s \mid \theta_2, F_2) - a(s \mid \theta_1, F_1)\big| \\
&= \big|\log g(y, d \mid z, c, \theta_2, F_2) - \log g(y, d \mid z, c, \theta_1, F_1)\big| \\
&= \frac{1}{g(y, d \mid z, c, \bar{\theta}, \bar{F})}\big|g(y, d \mid z, c, \theta_2, F_2) - g(y, d \mid z, c, \theta_1, F_1)\big| \\
&\leq C\big|g(y, d \mid z, c, \theta_2, F_2) - g(y, d \mid z, c, \theta_1, F_1)\big|.
\end{aligned}
$$

Consequently,

$$N_{[\ ]}(\epsilon, \mathcal{A}, L_1(P)) \le N_{[\ ]}(\frac{\epsilon}{C}, \mathcal{G}, L_1(P)).$$

So below, we only need to evaluate $N_{[\ ]}(\epsilon, \mathcal{G}, L_1(P))$.

Note that $g(y, d \mid z, c, \theta, F)$ is of the form (see the expression before (6))

$$g(y, d \mid z, c, \theta, F) = g_1(y, z \mid \beta)g_2(y, z \mid \beta)g_3(z \mid \alpha)$$

$$+g_4(y, z \mid \beta)g_5(y, z \mid \beta)g_6(d, z, c \mid \gamma, F)g_7(d, z, c \mid \gamma, F)g_8(z \mid \alpha),$$

with $g_1(1, z \mid \beta) = 1 - g_2(0, z \mid \beta)$, $g_3(z \mid \alpha) = 1 - g_8(z \mid \alpha)$ and $g_4(1, z \mid \beta) = 1 - g_5(0, z \mid \beta)$. To unify notation, we write $g_1(y, z \mid \beta), ..., g_8(z \mid \alpha)$ as $g_1(s \mid \theta, F), ..., g_8(s \mid \theta, F)$. Then $0 \le g_j(\cdot \mid \cdot) \le 1$ $(j = 1, ..., 8)$, and for (θ_1, F_1), $(\theta_2, F_2) \in (\Theta, \mathcal{F})$,

$$\big|g(y, d \mid z, c, \theta_1, F_1) - g(y, d \mid z, c, \theta_2, F_2)\big| \le \sum_{j=1}^{8} \big|g_j(s \mid \theta_1, F_1) - g_j(s \mid \theta_2, F_2)\big|.$$

Let $\mathcal{G}_j = \{g_j(s \mid \theta, F) : \theta \in \Theta, F \in \mathcal{F}\}$, $(j = 1, ..., 8)$. Then

$$N_{[\ ]}(\epsilon, \mathcal{G}, L_1(P)) \le \prod_{j=1}^{8} N_{[\ ]}(\frac{\epsilon}{8}, \mathcal{G}_j, L_1(P)).$$

Below we evaluate $N_{[\ ]}(\epsilon, \mathcal{G}_j, L_1(P))$ $(j = 1, ..., 8)$. Note that $g_2(1, z \mid \beta) \equiv 1$. Write $g_2(0, z \mid \beta) = (1 + \exp(\beta^T \tilde{z})^{-1}$ with $\tilde{z} = (1, z^T)^T$. For $\theta_1, \theta_2 \in \Theta$, there is a $\bar{\beta}$ lies between β_1 and β_2 such that

$$\big|g_2(0, z \mid \beta_2) - g_2(0, z|\beta_1)\big| = \frac{\exp(\bar{\beta}^T \tilde{z})\big|\tilde{z}^T(\beta_2 - \beta_1)\big|}{[1 + \exp(\bar{\beta}^T \tilde{z})]^2} \le |\tilde{z}^T(\beta_2 - \beta_1)| \le C\|\beta_2 - \beta_1\|,$$

for some $0 < C < \infty$. The last step above is by (C2). Thus, with $d = dim(\theta)$,

$$N_{[\ ]}(\frac{\epsilon}{8}, \mathcal{G}_2, L_1(P)) \le N_{[\ ]}(\frac{\epsilon}{8C}, \Theta, \|\cdot\|) = O(\frac{1}{\epsilon^d}) < \infty.$$

Since $g_1(1, z \mid \beta) = 1 - g_2(0, z \mid \beta)$, $N_{[\ ]}(\epsilon/8, \mathcal{G}_1, L_1(P)) = N_{[\ ]}(\epsilon/8, \mathcal{G}_2, L_1(P)) = O(1/\epsilon^d) < \infty$. By the same way, $N_{[\ ]}(\epsilon/8, \mathcal{G}_j, L_1(P)) = O(1/\epsilon^d) < \infty$, $(j = 3, 4, 5, 8)$.

Also, $g_6(s \mid \theta, F) = g_6(d, z, c \mid \gamma, F)$, $g_6(0, z, c \mid \gamma, F) \equiv 1$ and $g_6(1, z, c \mid \gamma, F) = F(\log c - \gamma^T z) := F(x)$ is a bounded monotone function from $R \mapsto [0, 1]$. Similarly, $g_7(s \mid \theta, F) = 1 - F(x)$ is a bounded monotone function from $R \mapsto [0, 1]$. Thus by Theorem 2.7.5 in [36], for some generic constant $0 < C < \infty$,

$$N_{[\ ]}(\frac{\epsilon}{8}, \mathcal{G}_j, L_1(P)) \le \exp\{\frac{C}{\epsilon}\}, \quad (j = 6, 7).$$

Now we get, for some generic constant $0 < C < \infty$,

$$N_{[\ ]}(\epsilon, \mathcal{A}, L_1(P)) \le N_{[\ ]}(\frac{\epsilon}{C}, \mathcal{G}, L_1(P)) \le \prod_{j=1}^{8} N_{[\ ]}(\frac{\epsilon}{8}, \mathcal{G}_j, L_1(P))$$

$$\leq \quad \frac{C}{\epsilon^{6d}} \exp\{\frac{2C}{\epsilon}\} < \infty.$$

Thus, by Theorem 2.4.1 in [36], $\mathcal{A}$ is a Glivenko-Cantelli class with respect to P, i.e,

$$\sup_{(\theta \in \Theta, F \in \mathcal{F})} |M_n(\theta, F) - M(\theta, F)| \to 0, \quad a.s.$$

Let $d(\theta, F; \theta_0, F_0) = \|\theta - \theta_0\| + \int |F(x) - F_0(x)| dF_0(x)$. Since the model is identifiable, (θ_0, F_0) is the unique maximizer of $M(\theta, F)$, so for every $\delta > 0$,

$$\sup_{d(\theta, F; \theta_0, F_0) > \delta} M(\theta, F) < M(\theta_0, F_0).$$

Also, $(\hat{\theta}_n, \hat{F}_n)$ is the MLE implies $P_n M(\hat{\theta}_n, \hat{F}_n) \geq P_n M(\theta_0, F_0)$. So by Theorem 5.8 in [35],

$$d(\hat{\theta}_n, \hat{F}_n; \theta_0, F_0) \overset{a.s.}{\to} 0.$$

This gives the desired result.

Proof of Theorem 5.2

Denote $\mathbf{V}^0 = (Y, X, D, Z, C)$ for the original data, and $\mathbf{V} = (Y, D, Z, C)$ for the observed incomplete data. The likelihood for $\mathbf{V}^0$ is

$$Q(\mathbf{v}^0) = Q(\mathbf{v}^0 \mid \theta, F) = \pi^y(x, z \mid \theta) \big[1 - \pi(x, z \mid \theta)\big]^{1-y} p^x(z \mid \theta) \big[1 - p(z \mid \theta)\big]^{1-x}$$

$$\times F(\log c - \gamma^T z)^d \big[1 - F(\log c - \gamma^T z)\big]^{1-d} h(z, c),$$

where $\pi(x, z \mid \theta)$ and $p(z \mid \theta)$ are given in (3) and (5). Denote the density-mass for the observed $\mathbf{V}$ given in (6) by

$$P(\mathbf{v}) = P(\mathbf{v} \mid \theta, F) = P(y, d, z, c \mid \theta, F) = \sum_{x=0}^{1} Q(y, x, d, z, c).$$

We identify the tangent space as $\dot{Q} = L_2^0(Q)$. Let $p(Z \mid \theta)$ as given in (5). The score operator $\mathbf{i}$ for model (6) is given by, $\forall a \in L_2^0(Q)$,

$$(\mathbf{i}a)(\mathbf{v}) = (\mathbf{i}a)(y, d, z, c) = E\Big(a(\mathbf{V}^0)\Big|\mathbf{v}\Big)$$

$$= a(y, 0, d, z, c)[1 - p(z \mid \theta)] + a(y, 1, d, z, c)p(z \mid \theta).$$

Its adjoint $\mathbf{i}^*$ satisfies $< b, \mathbf{i}a >_{L_2(P)} = < \mathbf{i}^*b, a >_{L_2(Q)}, \forall (a, b) \in (L_2^0(Q), L_2^0(P))$. Since

$$< b, \mathbf{i}a >_{L_2(P)} = \sum_y \sum_d \int \int b(y, d, z, c) \big[a(y, 0, d, z, c)[1 - p(z \mid \theta)]$$

$$+ a(y, 1, d, z, c)p(z \mid \theta)\big] P(y, d, z, c) dz dc$$

$$= \sum_y \sum_x \sum_d \int \int \left[b(y,d,z,c) \frac{[1-p(z\mid\theta)]^{1-x}p^x(z\mid\theta)P(y,d,z,c)}{Q(y,x,d,z,c)} \right]$$
$$a(y,x,d,z,c)Q(y,x,d,z,c)dzdx,$$

as $\mathbf{i}^* b \in L_2^0(Q)$, from the above we get,

$$(\mathbf{i}^* b)(\mathbf{v}^0) = e(\mathbf{v}^0) - E_Q(e(\mathbf{V}^0)),$$
$$e(\mathbf{v}^0) = b(y,d,z,c)\frac{[1-p(z\mid\theta)]^{1-x}p^x(z\mid\theta)P(y,d,z,c)}{Q(y,x,d,z,c)}.$$

Now we have, $\forall a \in L_2^0(Q)$,

$$(\mathbf{i}^* \mathbf{i} a)(\mathbf{v}^0) = \mathbf{i}^*[(\mathbf{i} a)(\mathbf{v})] = \mathbf{i}^* \left[a(y,0,d,z,c)[1-p(z\mid\theta)] + a(y,1,d,z,c)p(z\mid\theta) \right]$$

$$= \sum_s a(y,s,d,z,c)[1-p(z\mid\theta)]^{1-s}p^s(z\mid\theta)\frac{[1-p(z\mid\theta)]^{1-x}p^x(z\mid\theta)P(y,d,z,c)}{Q(y,x,d,z,c)}$$

$$- \sum_y \sum_x \sum_d \sum_s \int \int a(y,s,d,z,c)[1-p(z\mid\theta)]^{1-s}p^s(z\mid\theta)$$

$$\times \frac{[1-p(z\mid\theta)]^{1-x}p^x(z\mid\theta)P(y,d,z,c)}{Q(y,x,d,z,c)}Q(y,x,d,z,c)dzdc$$

$$:= \sum_s a(y,s,d,z,c)g(s,y,x,d,z,c) - \int \sum_s a(y,s,d,z,c)g(s,\mathbf{u})Q(d\mathbf{u}).$$

Thus solving $\mathbf{i}^* \mathbf{i} a = a_1$ in $a \in L_2^0(Q)$ for any given $a_1 \in L_2^0(Q)$ amounts to solve the above equation (set it to a_1) in a. It is easy to see that the array $\{g(s,y,x,d,z,c) : s,y,x,d = 0,1\}$ is not of full rank, so the above equation cannot be solved, i.e., $\mathbf{i}^* \mathbf{i}$ is not boundedly invertible. Yuan, Xu and Zheng [42] used this method for several incomplete data models, more details can be found there.

Proof of Theorem 5.3

Let $\ell(\theta, F \mid s)$ be the log-likelihood based on data $s = (y,z,d,c)$, and $\mathbf{D}_n$ be the set of all the observed data, and $M_n(\theta, F)$ and $M(\theta, F)$ as defined in the proof of Theorem 5.1.

Since $(\hat\theta_n, \hat F_n) \in (\Theta, \mathcal{F})$ is the MLE of (θ_0, F_0), it is the M-estimator based on the log-likelihood function $\ell(\theta, F \mid \mathbf{D}_n)$ on the constrained parameter space $(\Theta, \mathcal{F})$, and so $M_n(\hat\theta_n, \hat F_n) \geq M_n(\theta, F) \geq M_n(\theta, F) - O_p(r_n^{-2})$, for any $(\theta, F) \in (\Theta, \mathcal{F})$ and any positive sequence $r_n \to \infty$.

Let $d(\hat\theta_n - \theta_0, \hat F_n - F_0)$ as in the proof of Theorem 5.1. Since conditions of Theorem 5.1 is satisfied, $\|\hat\theta_n - \theta_0\| + \|\hat F_n - F_0\| \xrightarrow{P} 0$.

Note that $(\theta_0, F_0) = \arg\sup_{(\theta,F)\in(\Theta,\mathcal{F})} M(\theta, F)$. Denote $f(\theta, F) := \ell(\theta, F \mid s) - \ell(\theta_0, F_0 \mid s)$ be the log-likelihood ratio, and $f^{(1)}(\theta, F)$ and $f^{(2)}(\theta, F)$ be the first and second order partial derivatives of $f(\theta, F)$ with respect to $(\theta, F(\cdot))$, here the partial derivative with respect to F is in the Gâteaux sense. Denote $\ell^{(1)}(\theta, F \mid s)$ and $\ell^{(2)}(\theta, F \mid s)$ similarly. Note that $f(\theta_0, F_0) = 0$, $E_{(\theta_0,F_0)}f^{(1)}(\theta_0, F_0) = E_{(\theta_0,F_0)}\ell^{(1)}(\theta_0, F_0 \mid s) = 0$, and $E_{(\theta_0,F_0)}f^{(2)}(\theta_0, F_0) = E_{(\theta_0,F_0)}\ell^{(2)}(\theta_0, F_0 \mid s)$ is negative definite, so using Taylor expansion, $M(\theta, F) - M(\theta_0, F_0) = (1/2)E_{(\theta_0,F_0)}[(\theta - \theta_0, F - F_0)'f^{(2)}(\bar{\theta}, \bar{F})(\theta - \theta_0, F - F_0)] < 0$ is of order $O\big(d^2(\theta - \theta_0, F - F_0)\big)$, in small neighborhood of (θ_0, F_0), where $(\bar{\theta}, \bar{F})$ is an intermediate value between (θ, F) and (θ_0, F_0). So for any $0 < \delta_n \to 0$ and any δ and some $\eta < \infty$ with $\delta_n < \delta \le \eta$, for some $0 < C < \infty$,

$$\sup_{\delta/2 < d(\theta - \theta_0, F - F_0) \le \delta,\ (\theta,F)\in(\Theta,\mathcal{F})} M(\theta, F) - \mathbb{M}(\theta_0, F_0) \le -C\delta^2.$$

Next we show, with E^* for outer expectation,

$$E^* \sup_{\delta/2 < d(\theta - \theta_0, F - F_0) \le \delta,\ (\theta,F)\in(\Theta,\mathcal{F})} \sqrt{n}\Big|(M_n - M)(\theta, F) - (M_n - M)(\theta_0, F_0)\Big| \le \phi_n(\delta),$$

$$(A.0)$$

for some decreasing function $\phi_n(\cdot)$.

For this, let $\mathcal{A}$ as defined in the proof of Theorem 5.1, $N_{[\ ]}(\epsilon, \mathcal{A}, L_2(P))$ be the bracketing number of size ϵ, under the $L_2(P)$ norm, to cover the space $\mathcal{A}$. In the proof of Theorem 5.1, we showed that $N_{[\ ]}(\epsilon, \mathcal{A}, L_1(P)) = O\big(\epsilon^{-6d}\exp(C/\epsilon)\big)$. The same conclusion is true with $L_1(P)$ replaced by $L_2(P)$, as the evaluation there is true for $L_r(P)$ with all $r > 0$. Let C be some generic finite positive constant, and

$$J_{[\ ]}(\epsilon, (\Theta, \mathcal{F}), L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\ ]}(\epsilon, (\Theta, \mathcal{F}), L_2(P))}d\epsilon$$

$$= \int_{-\infty}^{\log \delta} e^t \sqrt{1 + Ce^{-t}}dt \le C \int_{-\infty}^{\log \delta} e^t e^{-t/2}dt = C\delta^{1/2}.$$

In the above, we used the fact that for small $\delta > 0$, $\log \delta < 0$, so $e^{-t} > 1$ and $1 + Ce^{-t} \le (1 + C)e^{-t}$ on $(-\infty, \log \delta)$.

Let $\mathcal{G} = \{f(\cdot \mid \theta, F) : (\theta, F) \in (\Theta, \mathcal{F}), \delta/2 < d(\theta - \theta_0, F - F_0) \le \delta\}$, $\mathbb{G}_n f = \sqrt{n}(P_n - P)f$, $\|\mathbb{G}_n\|_{\mathcal{G}} = \sup_{f\in\mathcal{G}} \mathbb{G}_n f$, and $\|f\| = \sup_s f(s \mid \theta, F)$. Note that since $\mathcal{G}$ is a subset of log-likelihood ratios, and $\ell(\theta_0, F_0 \mid s)$ is fixed, so $N_{[\ ]}(\epsilon, \mathcal{G}, L_2(P)) \le N_{[\ ]}(\epsilon, (\Theta, \mathcal{F}), L_2(P))$ and consequently, $J_{[\ ]}(\epsilon, \mathcal{G}, L_2(P)) \le J_{[\ ]}(\epsilon, (\Theta, \mathcal{F}), L_2(P)) \le C\delta^{1/2}$.

Also, it is easy to see that, with our specification of the likelihood, $Pf^2 < \delta^2$ and $\|f\|_\infty < C$ for all $f \in \mathcal{G}$, for some $0 < C < \infty$. Thus, by Lemma 3.4.2 in [36] (p.324),

$$E^*\|\mathbb{G}_n\|_{\mathcal{G}} \le J_{[\ ]}(\epsilon, \mathcal{G}, L_2(P))\Big(1 + \frac{J_{[\ ]}(\epsilon, \mathcal{G}, L_2(P))}{\delta^2\sqrt{n}}C\Big) \le C\delta^{1/2}(1 + \delta^{-3/2}n^{-1/2}),$$

which is (A.0) with $\phi_n(\delta) = C\delta^{1/2}(1 + \delta^{-3/2}n^{-1/2})$. Take $r_n = n^{1/3}$, then

$$r_n^2\phi_n\Big(\frac{1}{r_n}\Big) = Cn^{1/3} \leq \sqrt{n}.$$

Now, all conditions of Theorem 3.4.1 in [36] (p.322) are satisfied, so by this Theorem, $r_n d(\hat{\theta}_n - \theta_0, \hat{F}_n - F_0) = n^{1/3}d(\hat{\theta}_n - \theta_0, \hat{F}_n - F_0) = O_p(1)$. $\qquad\square$

References

[1] Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics*, 26, 641-647. MR0073895

[2] Bauer, K., Eriksson, B., Lassen, M., and Turpie, A. (2001). Fondaparinux compared to enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. *New England Journal of Medicine*, 345, 1298-1394.

[3] Balan, T. A., and Putter, H. (2017). frailtyEM: An R package for estimating semiparametric shared frailty models.

[4] Bordes, L., Chauveau, D., and Vandekerknove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Statist. Data Anal.* **51**, 5429-5443. MR2370882

[5] Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68**, 417-422. MR0626401

[6] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. Chapman and Hall, London. MR1630517

[7] Chen, B., Li, P., Qin, J., and Yu, T. (2016). Using a monotonic density ratio model to find the asymptotically optimal combination of multiple diagnostic tests. *J. Amer. Statist. Assoc.* **111**, 861-874. MR3538711

[8] Chen, J., Zhang, D., and Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biometrics* **3**, 347-360. MR2703312

[9] Cook, R. J., White, B. J. G., Yi, G. Y., Lee, K. A., and Warkentin, T. E. (2008). Analysis of a non-susceptible fraction with current status data. *Statistics in Medicine*, 27, 2715-2730. MR2440061

[10] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38. MR0501537

[11] Eriksson, B. I., Bauer, K. A., Lassen, M. R., Turpie, A. G. G., for the Steering Committee of the Pentasaccharide in Hip-Fracture Surgery Study. (2001). Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after hip-fracture surgery. *New England Journal of Medicine*, 345, 1298-1394.

[12] GROENEBOOM, P. AND WELLNER, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkháuser Verlag, Basel. MR1180321

[13] GROENEBOOM, P., MAATHUIS, M. H.,AND WELLNER, J. A. (2008a). Current status data with competing risks: consistency and rates of convergence of the MLE. *Ann. Statist.* **36**, 1031-1063. MR2418648

[14] GROENEBOOM, P., MAATHUIS, M. H., AND WELLNER, J. A. (2008b). Current status data with competing risks: limiting distribution of the MLE. *Ann. Statist.* **36**, 1064-1089. MR2418649

[15] GUSTAFSON, P. (2004). *Measurement error and misclassification in statistics and epidemiology*. Boca Raton, Florida: CRC Press. MR2005104

[16] HANLEY, J. A., AND PARNES, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39**, 129-139. MR0712744

[17] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*, 24 (2), 540-568. MR1394975

[18] HUANG, J., AND WELLNER, J. A. (1997). Interval censored survival data: a review of recent progress. In D. Lin and T. Fleming (eds.), *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pp. 123-169. New York: Springer-Verlag.

[19] JEWELL, N. P., AND SHIBOSKI, S. C. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics*, 46, 1131-1150.

[20] JIN, Z. (2016). Semiparametric accelerated failure time model for the analysis of right censored data. *Communications for Statistical Applications and Methods*, 23, 467-478.

[21] LAM, K. F., AND XUE, H. Q. (2005). A semiparametric regression curve model with current status data. *Biometrika*, 92, 573-586. MR2202647

[22] LASSEN, M. R., BAUER, K. A., ERIKSSON, B. I., TURPIE, A. G. G., FOR THE EUROPEAN PENTASACCHARIDE HIP ELECTIVE SURGERY STUDY STEERING COMMITTEE (EPHESIS). (2002). Postoperative fondaparinux versus preoperative enoxaparin for prevention of venous thromboembolism in elective hip-replacement surgery: a randomized double-blind comparison. *Lancet*, 359, 1715-1720.

[23] LAWLESS, J. F., KALBFLEISCH, J. D., AND WILD, C. J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society, Series B*, 61, 413-438. MR1680310

[24] LIN, D. Y., OAKES, D., AND YING, Z. (1998). Additive hazard regression with current status data. *Biometrika*, 85, 289-298. MR1649115

[25] MAATHUIS, M.H., AND HUDGENS, M.G. (2011). Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. *Biometrika* **98**, 325-340. MR2806431

[26] MUÑOZ, A. (1980). Nonparametric estimation from censored bivariate observations. Technical Report, Department of Statistics, Stanford University.

[27] PEPE, M. S., AND FLEMING, T. R. (1991). A nonparametric method for dealing with mismeasured covariate data. *Journal of the American Statis-*

tical Association, 86, 108-113. MR1137103

[28] R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/. ISBN: 3-900051-07-0.

[29] Rabinowitz, D., Tsiatis, A. A., Aragon, J. (1995). Regression with interval censored data. *Biometrika.* 82,501-514. MR1366277

[30] Rossini, A. J., and Tsiatis, A. A. (1996). A semiparametric proportion-alodds regression model for the analysis of current status data. *Journal of the American Statistical Association*, 91, 713-721. MR1395738

[31] Schafer, D. W., and Purdy, K. G. (1996). Likelihood analysis for errors-in-variables regression with replicate measurements. *Biometrika*, 93, 813-824. MR1440046

[32] Shiboski, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis*, 4, 29-50.

[33] Tian, L. and Cai, T. (2006). On Accelerated Failure Time Model for Current Status and Interval Censored Data. *Biometrika*, 93, 329-242. MR2278087

[34] Turpie, A. G. G., Bauer, K. A.,, Eriksson, B. I., and Lassen, M. R. (2002). Postoperative fondaparinux versus postoperative enoxaparin for prevention of venous thromboembolism after elective hip-replacement surgery: a randomized double-blind trial. *Lancet*, 359, 1721-1726.

[35] van der Vaart, A. (2002). *Semiparametric Statistics*, in Part III, *Lectures on Probability Theory and Statistics*, Springer. MR1915446

[36] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*, Springer. MR1385671

[37] Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871-1879.

[38] White, R. H., Romano, P. S., Zhou, H., Rodrigo, J., and Bargar, W. (1998). Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. *Journal of the American Medical Association*, 158, 1525-1531.

[39] Wu, C. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* 95-103. MR0684867

[40] Yan, Y. and Yi, G. Y. (2016). A Class of Functional Methods for Error-Contaminated Survival Data under Additive Hazards Models with Replicate Measurements. *Journal of the American Statistical Association*, 111, 684-695. MR3538697

[41] Yi, G. Y., and Cook, R. J. (1998). Errors in the measurement of covariates. *Encyclopedia of Biostatistics*, 3, 1741-1748.

[42] Yuan, A., Xu, J., Zheng, G. (2012). Root-n estimability of some missing data models, *Journal of Multivariate Analysis*, **106**, 147-166. MR2887685

[43] Zeng, L., Cook, R. J., and Warkentin, T. E. (2010). Regression analysis with a misclassified covariate from a current status observation scheme. *Biometrics*, 66, 415-425. MR2758821