Linear regression analysis of survival data with missing censoring indicators

Qihua Wang · Gregg E. Dinse

Abstract Linear regression analysis has been studied extensively in a random censorship setting, but typically all of the censoring indicators are assumed to be observed. In this paper, we develop synthetic data methods for estimating regression parameters in a linear model when some censoring indicators are missing. We define estimators based on regression calibration, imputation, and inverse probability weighting techniques, and we prove all three estimators are asymptotically normal. The finite-sample performance of each estimator is evaluated via simulation. We illustrate our methods by assessing the effects of sex and age on the time to non-ambulatory progression for patients in a brain cancer clinical trial.

Keywords Asymptotic normality · Censoring indicator · Imputation · Inverse probability weighting · Least squares · Missing at random · Regression calibration

1 Introduction

Linear regression is a popular statistical tool that has been used successfully in many areas. In survival analysis, a log-transformation of the response variable converts a conventional linear model to an accelerated failure time model, which is an appealing

Q. Wang
Department of Mathematics and Statistics, Yunnan University, Kunming 650091, China

Q. Wang
Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, China

G. E. Dinse (✉)
Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA
e-mail: dinse@niehs.nih.gov

alternative to the Cox (1972) proportional hazards model because of its direct interpretation (*cf.* Reid 1994). Linear regression has been studied extensively for the analysis of randomly right censored data, and various procedures have been developed; see, for example, Miller (1976), Buckley and James (1979), Koul et al. (1981), Leurgans (1987), Ritov (1990), Tsiatis (1990), Lai and Ying (1991); Ying (1993), Jin et al. (2003), and Li and Wang (2003). All of these methods assume that every censoring indicator is observed.

In practice, some censoring indicators may be unobserved, and we develop methods for this situation. For example, we analyzed data from a clinical trial involving patients with glioblastoma, a form of brain cancer. One measure of post-treatment quality of life is the length of time until a patient declines to a state in which mobility is lost and the cancer has returned. We evaluated data on adult glioblastoma patients over 50 years of age. Progression status was always known, but ambulatory status was unknown in some cases. By the end of the study, some patients were non-ambulatory and had progressed, some had progressed but remained ambulatory, some had progressed but their ambulatory status was unrecorded, and some had not progressed. Our analysis focused on the time to non-ambulatory progression, and these four types of events contributed uncensored observations, censored observations, observations with a missing censoring indicator, and censored observations, respectively. In particular, we were interested in how sex and age affected the time to non-ambulatory progression. Our methods have applications in many other areas, including analyses with missing death certificates in epidemiological studies, unperformed necropsies in animal experiments, and unidentified component failures in quality control settings.

The methods we propose are formulated in the context of a censored survival problem with missing censoring indicators, but they can be used to analyze competing risks data on event-specific failure times when some event indicators are unknown. In practice, the usual competing risks analysis focuses on a particular event type of interest and treats the times to failure from other events as censored observations on the event time of interest. Thus, one can view an unknown failure type as a missing censoring indicator and use our methods.

If censoring indicators are missing, standard estimators for censored data are not directly applicable. One solution is to ignore cases with missing data and apply conventional analyses to the complete cases only. However, the efficiency of this complete-case approach decreases as the degree of missingness increases. Also, the complete-case estimator is consistent only when censoring indicators are missing completely at random (MCAR). Alternatively, some analyses directly incorporate survival data with missing censoring indicators. For example, Dinse (1982); Lo (1991); McKeague and Subramanian (1998), and Subramanian (2004, 2006) proposed survival estimators. Within the regression context, Goetghebeur and Ryan (1990, 1995); Dewanji (1992); Lu and Tsiatis (2001), and Tsiatis et al. (2002) developed methods under a proportional hazards model; Zhou and Sun (2003) and Lu and Liang (2008) worked with an additive hazards model; and Gao and Tsiatis (2005) employed a linear transformation competing risks model. Linear regression techniques, however, have received little attention in situations where censoring indicators are missing, and thus we now focus on that particular problem.

Suppose there are n subjects and their observations follow the linear regression model

$$Y_i = X_i^\top \beta + \epsilon_i,$$

where Y_i is a response variable, X_i is a vector of d covariates, β is a d-vector of regression coefficients, and ϵ_i is a random error for subject $i (i = 1, \ldots, n)$. Given $\{X_1, \ldots, X_n\}$, the errors $\{\epsilon_1, \ldots, \epsilon_n\}$ are independent and identically distributed (*i.i.d.*) with mean zero. Let $\{C_1, \ldots, C_n\}$ be *i.i.d.* censoring variables, where C_i can randomly censor Y_i on the right. The observed response is $\widetilde{Y}_i = \min(Y_i, C_i)$, the censoring indicator is $\delta_i = I(Y_i \leq C_i)$, and the missingness indicator ξ_i is 1 if δ_i is observed and 0 otherwise. Although calling δ_i a censoring indicator might seem backwards, as $\delta_i = 1$ indicates an uncensored observation rather than a censored observation, we follow the convention that refers to δ_i as a censoring indicator.

We observe $\{\widetilde{Y}, X, \delta\}$ for subjects with complete data ($\xi = 1$) and $\{\widetilde{Y}, X\}$ for subjects with a missing censoring indicator ($\xi = 0$). Typically the missingness mechanism is assumed to produce data that are missing at random (MAR), which implies $pr(\xi = 1|\widetilde{Y}, X, \delta) = pr(\xi = 1|\widetilde{Y}, X)$. Alternatively, we assume $pr(\xi = 1|\widetilde{Y}, X, \delta) = pr(\xi = 1|\widetilde{Y})$, which is more restrictive than the MAR assumption but less restrictive than the common MCAR assumption that $pr(\xi = 1|\widetilde{Y}, X, \delta) = pr(\xi = 1)$. We made an assumption stricter than MAR to avoid postulating a parametric model for $pr(\xi = 1|\widetilde{Y}, X)$ and to sidestep problems encountered when applying a nonparametric method to sparse data, such as can arise in high-dimensional situations. Under the weaker MAR assumption, we expect to obtain similar results for a reasonable model in the parametric case or with sufficient data in the nonparametric case, but we leave such analyses for future work.

In this paper, we propose three estimators for β. Our methods extend an approach taken by Koul et al. (1981) when all censoring indicators are observed. They used the formula

$$Y_i^* = \frac{\delta_i \widetilde{Y}_i}{1 - G_n(\widetilde{Y}_i)} \tag{1}$$

to transform observed data $(\widetilde{Y}_i, \delta_i)$ into synthetic data Y_i^*, where $G_n(c)$ is a product-limit type estimator of the censoring distribution $G(c) = pr(C \leq c)$. Koul *et al.* estimated β by applying a modified least squares procedure to (Y_i^*, X_i), but their estimator does not handle missing censoring indicators. Also, they deflate the synthetic response Y_i^* to 0 when the true response Y_i exceeds C_i, and they inflate Y_i^* beyond Y_i when Y_i is actually observed, both of which are counterintuitive. Alternatively, our estimators ease this dilemma by replacing δ_i in (1) with a weighted average of δ_i and its estimated conditional mean $\hat{m}_i$, say $p_i \delta_i + (1 - p_i)\hat{m}_i$, which allows statistical inference about β when some censoring indicators are missing.

The article is organized as follows. Section 2 develops a synthetic data approach based on regression calibration, which sets $p_i \equiv 0$. Sections 3 and 4 define imputation and inverse probability weighted estimators, which set $p_i = \xi_i$ and $p_i = \xi_i/\hat{E}(\xi_i|\widetilde{Y}_i)$, respectively. As δ_i is unknown if $\xi_i = 0$, each of the three formulas for p_i is zero in

that situation. All three least squares estimators of β are shown to be asymptotically normal. Section 5 reports the results of a simulation study conducted to evaluate the finite sample behavior of the three proposed estimators, and Sect. 6 illustrates their use by applying them to the glioblastoma data. Finally, Sect. 7 discusses our estimators and the underlying assumptions, and proofs of the main results are given in the Appendix.

2 Least squares regression calibration estimator

If we define $Z_i^\top = (\widetilde{Y}_i, X_i^\top)$ and $m(z_i) = E(\delta_i | Z_i = z_i)$, then one natural generalization of Y_i^* replaces δ_i by its conditional expectation to obtain $Y_{i,R} = m(Z_i)\widetilde{Y}_i / \{1 - G(\widetilde{Y}_i)\}$. Averaging over δ_i and invoking the usual conditional expectation arguments yield:

$$E(Y_{i,R}|X_i) = \cdots = E\left[E\left\{ \frac{\widetilde{Y}_i \delta_i}{1 - G(\widetilde{Y}_i)} \middle| \widetilde{Y}_i, X_i \right\} \middle| X_i \right] = \cdots = E(Y_i|X_i) = X_i^\top \beta.$$

For known $m(\cdot)$ and $G(\cdot)$, this result suggests the following least squares estimator of β:

$$\widetilde{\beta}_R = \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1} \sum_{i=1}^n X_i Y_{i,R}.$$

In practice, however, $m(\cdot)$ and $G(\cdot)$ are usually unknown, and hence $\widetilde{\beta}_R$ is not defined. One simple solution is to calculate $\widetilde{\beta}_R$ using estimates of $m(\cdot)$ and $G(\cdot)$.

For example, consider a parametric model for $m(z)$ of the form $m_0(z, \theta)$, where $m_0(\cdot, \cdot)$ is a known continuous function and θ is a vector of unknown parameters. The parameter vector θ can be estimated by $\widehat{\theta}_n$, the maximizer of the likelihood

$$\prod_{i=1}^n \{m_0(Z_i, \theta)\}^{\xi_i \delta_i} \{1 - m_0(Z_i, \theta)\}^{\xi_i(1-\delta_i)},$$

and thus $m(z)$ can be estimated by $\widehat{m}_n(z) = m_0(z, \widehat{\theta}_n)$. As for estimating $G(\cdot)$, we begin by defining $\pi(\widetilde{y}) = E(\xi | \widetilde{Y} = \widetilde{y})$, which can be estimated nonparametrically by

$$\widehat{\pi}_n(\widetilde{y}) = \sum_{i=1}^n \xi_i W\left(\frac{\widetilde{y} - \widetilde{Y}_i}{b_n} \right) \bigg/ \sum_{i=1}^n W\left(\frac{\widetilde{y} - \widetilde{Y}_i}{b_n} \right),$$

where $W(\cdot)$ is a kernel function and b_n is a bandwidth sequence. Next, we define a Horvitz and Thompson (1952) type inverse probability weighted estimator of $u(\widetilde{y}) = E(\delta | \widetilde{Y} = \widetilde{y})$:

$$\widehat{u}_n(\widetilde{y}) = \sum_{i=1}^{n} \delta_i \left(\frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)} \right) K \left(\frac{\widetilde{y} - \widetilde{Y}_i}{h_n} \right) \bigg/ \sum_{i=1}^{n} \left(\frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)} \right) K \left(\frac{\widetilde{y} - \widetilde{Y}_i}{h_n} \right),$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth sequence. Finally, similar to Dikta (1998) and Wang and Ng (2008), we define the following estimator of $G(\widetilde{y})$:

$$\widehat{G}_n(\widetilde{y}) = 1 - \prod_{i:\widetilde{Y}_i < \widetilde{y}} \left(\frac{n - R_i}{n - R_i + 1} \right)^{1 - \widehat{u}_n(\widetilde{Y}_i)},$$

where R_i denotes the rank of $\widetilde{Y}_i (i = 1, \ldots, n)$. Now, analogous to $\widetilde{\beta}_R$, we can define the following least squares regression calibration estimator of β:

$$\widehat{\beta}_R = \left(\sum_{i=1}^{n} X_i X_i^\top \right)^{-1} \sum_{i=1}^{n} X_i \widehat{Y}_{i,R},$$

where $\widehat{Y}_{i,R}$ is $Y_{i,R}$ with $m(\cdot)$ and $G(\cdot)$ replaced by $\widehat{m}_n(\cdot)$ and $\widehat{G}_n(\cdot)$, respectively.

Define $\sigma_{r,s} = E \left[\frac{\pi(\widetilde{Y})}{m_0(Z,\theta)\{1-m_0(Z,\theta)\}} \frac{\partial m_0(Z,\theta)}{\partial \theta_r} \frac{\partial m_0(Z,\theta)}{\partial \theta_s} \right]$ for $1 \leq r, s \leq d + 1$ and assume:

(C.1) $m_0(z, \theta)$ has bounded first-order derivatives on θ.
(C.2) $\sigma_{r,s} < \infty$ for $1 \leq r, s \leq d + 1$ and matrix $A(\theta) = (\sigma_{r,s})$ is positive definite.
(C.3) $E \left\{ \|X\|^2 \widetilde{Y}^2 / \bar{H}^2(\widetilde{Y}) \right\} < \infty$, where $H(t) = pr(\widetilde{Y} \leq t)$ and $\bar{H}(t) = 1 - H(t)$.
(C.4) $E \left\{ \|X\|^2 \right\} < \infty$.
(C.5) $\pi(\cdot)$ has a bounded first-order derivative.
(C.6) $W(\cdot)$ is of order 1 with bounded support.
(C.7) $nb_n \longrightarrow \infty$ and $nb_n^2 \longrightarrow 0$.

The following theorem addresses the asymptotic normality of $\widehat{\beta}_R$.

Theorem 2.1 *Under assumptions (C.1), (C.2), (C.3) and (C.4), we have*

$$n^{1/2}(\widehat{\beta}_R - \beta) \xrightarrow{\mathcal{L}} N(0, V_R),$$

where $V_R = \Sigma^{-1} \Omega_R \Sigma^{-1}$, $\Sigma = E(XX^\top)$, and Ω_R is defined in (A.15) of the Appendix.

The structure of Ω_R is complex; thus, estimating V_R by replacing the unknowns in Ω_R with appropriate component estimates is difficult to implement in practice. Alternatively, we estimate V_R by $n\widehat{V}_J$, where $\widehat{V}_J$ is a jackknife estimator of the asymptotic variance of $\widehat{\beta}_R$. In particular, we use a jackknife variance estimator of the form proposed by Peddada and Patwardhan (1992), which was motivated by a minimum norm quadratic unbiased estimator (MINQUE) and which in our case reduces to the simple form:

$$\widehat{V}_J = \left(\sum_{i=1}^{n} X_i X_i^\top \right)^{-1} \left[\sum_{i=1}^{n} \left(\widehat{Y}_{i,R} - X_i^\top \widehat{\beta}_R \right)^2 X_i X_i^\top \right] \left(\sum_{i=1}^{n} X_i X_i^\top \right)^{-1}. \quad (2)$$

Theorem 2.2 *Under assumptions (C.1), (C.2), (C.3) and (C.4), we have*

$$n\widehat{V}_J \xrightarrow{p} V_R.$$

Proofs of Theorem 2.1 and Theorem 2.2 are given in the Appendix.

3 Least squares imputation estimator

Imputation is a popular method for dealing with missing data, and this section describes such an approach. When some censoring indicators are missing, some synthetic data in (1) are not defined. However, in the expression for Y_i^*, we can replace $G_n(\cdot)$ by $\widehat{G}_n(\cdot)$ and impute δ_i with $\widehat{m}_n(Z_i)$ if δ_i is missing ($\xi_i = 0$), which yields the imputed synthetic data

$$\widehat{Y}_{i,I} = \frac{\{\xi_i \delta_i + (1 - \xi_i)\widehat{m}_n(Z_i)\}\, \widetilde{Y}_i}{1 - \widehat{G}_n(\widetilde{Y}_i)} \qquad i = 1, \cdots, n.$$

Analogous to Sect. 2, we define a least squares imputation estimator for β:

$$\widehat{\beta}_I = \left(\sum_{i=1}^{n} X_i X_i^\top\right)^{-1} \sum_{i=1}^{n} X_i \widehat{Y}_{i,I}.$$

Under the assumed missingness mechanism, $\widehat{\beta}_I$ also can be motivated by the relationship

$$E(Y_I|X) = E\left(\left.\frac{[E(\xi|\widetilde{Y})\delta + \{1 - E(\xi|\widetilde{Y})\}m(Z)]\widetilde{Y}}{1 - G(\widetilde{Y})}\right|X\right) = \cdots = E(Y_R|X) = X^\top\beta,$$

where $Y_I = \{\xi\delta + (1 - \xi)m(Z)\}\widetilde{Y}/\{1 - G(\widetilde{Y})\}$ and $Y_R = m(Z)\widetilde{Y}/\{1 - G(\widetilde{Y})\}$.

The following theorem addresses the asymptotic normality of $\widehat{\beta}_I$.

Theorem 3.1 *Under assumptions (C.1), (C.2), (C.3) and (C.4), we have*

$$n^{1/2}\left(\widehat{\beta}_I - \beta\right) \xrightarrow{\mathcal{L}} N(0, V_I),$$

where $V_I = \Sigma^{-1}\Omega_I\Sigma^{-1}$, $\Omega_I = \Omega_R + \Omega_I^*$, and Ω_I^* is defined in (A.20) of the Appendix.

Similar to Sect. 2, we estimate V_I by the jackknife methods described earlier. Clearly $\widehat{\beta}_I$ has larger asymptotic variance than $\widehat{\beta}_R$, and hence $\widehat{\beta}_I$ is asymptotically less efficient than $\widehat{\beta}_R$. Intuitively, one reason for this may be that the imputed synthetic observation $\widehat{Y}_{i,I}$ is zero if Y_i is known to be censored ($\xi_i = 1$, $\delta_i = 0$). We see that the imputation approach creates synthetic data with greater variability than the regression calibration method by noting

$$var(Y_I|X) = E\left(\frac{[E(\xi|\widetilde{Y})\delta + \{1 - E(\xi|\widetilde{Y})\}m^2(Z)]\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\bigg|X\right) - E^2(Y_I|X)$$

$$= E\left(\frac{[\pi(\widetilde{Y})E(\delta|Z) + \{1 - \pi(\widetilde{Y})\}m^2(Z)]\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\bigg|X\right) - E^2(Y_R|X)$$

$$= E\left(\frac{\pi(\widetilde{Y})m(Z)\{1 - m(Z)\}\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\bigg|X\right) + var(Y_R|X) \geq var(Y_R|X).$$

4 Least squares inverse probability weighted estimator

Inverse probability weighting is another popular approach for dealing with missing data. If we define $Y_W = [\{\xi/\pi(\widetilde{Y})\}\delta + \{1 - \xi/\pi(\widetilde{Y})\}m(Z)]\widetilde{Y}/\{1 - G(\widetilde{Y})\}$, then under the assumed missingness mechanism, conditional expectations similar to those in Sects. 2 and 3 give

$$E(Y_W|X) = E\left(\frac{[\{E(\xi)/\pi(\widetilde{Y})\}\delta + \{1 - E(\xi)/\pi(\widetilde{Y})\}m(Z)]\widetilde{Y}}{1 - G(\widetilde{Y})}\bigg|X\right)$$

$$= \cdots = E(Y_R|X) = X^\top\beta.$$

In practice, $m(\cdot)$, $\pi(\cdot)$ and $G(\cdot)$ are usually unknown, but we can use estimates of these quantities to calculate inverse probability weighted synthetic data:

$$\widehat{Y}_{i,W} = \frac{\left[\left\{\frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)}\right\}\delta_i + \left\{1 - \frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)}\right\}\widehat{m}_n(Z_i)\right]\widetilde{Y}_i}{1 - \widehat{G}_n(\widetilde{Y}_i)} \qquad i = 1, \cdots, n.$$

Note that $\widehat{Y}_{i,W}$ is very similar to $\widehat{Y}_{i,I}$, except the former weights each ξ_i inversely proportional to the estimated conditional mean of ξ_i (*i.e.*, the estimated conditional probability that δ_i is known, given $\widetilde{Y}_i$). Applying least squares techniques to these synthetic data produces the inverse probability weighted estimator

$$\widehat{\beta}_W = \left(\sum_{i=1}^n X_i X_i^\top\right)^{-1} \sum_{i=1}^n X_i \widehat{Y}_{i,W}.$$

The following theorem addresses the asymptotic normality of $\widehat{\beta}_W$.

Theorem 4.1 *Under assumptions (C.1) through (C.7), we have*

$$n^{1/2}\left(\widehat{\beta}_W - \beta\right) \xrightarrow{\mathcal{L}} N(0, V_W),$$

where $V_W = \Sigma^{-1}\Omega_W\Sigma^{-1}$, $\Omega_W = \Omega_R + \Omega_W^*$, *and* Ω_W^* *is defined in (A.26) of the Appendix.*

As described earlier, V_W can be estimated by jackknife methods. We see $\widehat{\beta}_W$ has a larger asymptotic variance than $\widehat{\beta}_I$, and hence $\widehat{\beta}_R$, by noting $\alpha^\top(\Omega_W^* - \Omega_I^*)\alpha > 0$

for any d-vector α. One explanation is that if Y_i is known to be censored ($\xi_i = 1$, $\delta_i = 0$), then inverse probability weighting leads to synthetic data of the form $\widehat{Y}_{i,W} = \{1 - 1/\widehat{\pi}_n(\widetilde{Y}_i)\}\widehat{m}_n(Z_i)\widetilde{Y}_i/\{1 - \widehat{G}_n(\widetilde{Y}_i)\}$, the sign of which is the opposite of $\widetilde{Y}_i$, and hence Y_i. Thus, this method creates synthetic data that are more variable than those based on either Y_R or Y_I, which can be seen by noting

$$
\begin{aligned}
var(Y_W|X) &= E\left[\frac{\left\{\frac{E(\xi|\widetilde{Y})\delta}{\pi^2(\widetilde{Y})} + \frac{2E(\xi|\widetilde{Y})\delta m(Z)}{\pi(\widetilde{Y})} - \frac{2E(\xi|\widetilde{Y})\delta m(Z)}{\pi^2(\widetilde{Y})}\right\}\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\right. \\
&\quad \left.+ \frac{\left\{m^2(Z) - \frac{2E(\xi|\widetilde{Y})m^2(Z)}{\pi(\widetilde{Y})} + \frac{E(\xi|\widetilde{Y})m^2(Z)}{\pi^2(\widetilde{Y})}\right\}\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\right|X\right] - E^2(Y_W|X) \\
&= E\left[\frac{\left\{\frac{E(\delta|Z)}{\pi(\widetilde{Y})} + 2E(\delta|Z)m(Z) - \frac{2E(\delta|Z)m(Z)}{\pi(\widetilde{Y})}\right\}\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\right. \\
&\quad \left.+ \frac{\left\{m^2(Z) - 2m^2(Z) + \frac{m^2(Z)}{\pi(\widetilde{Y})}\right\}\widetilde{Y}^2}{\{1 - G(\widetilde{Y})\}^2}\right|X\right] - E^2(Y_R|X) \\
&= E\left[\frac{m(Z)\{1 - m(Z)\}\widetilde{Y}^2}{\pi(\widetilde{Y})\{1 - G(\widetilde{Y})\}^2}\right|X\right] + var(Y_R|X) \geq var(Y_I|X) \geq var(Y_R|X).
\end{aligned}
$$

One advantage of this approach is that if either $m(\cdot)$ or $\pi(\cdot)$ is modeled correctly, the estimator $\hat{\beta}_W$ will be consistent, which is the so-called "double-robustness" property (see, *e.g.*, Scharfstein et al. 1999; Lunceford and Davidian 2004; and Wang et al. 2004). Note that we estimate $\pi(\cdot)$ nonparametrically, and thus $\widehat{\beta}_W$ enjoys this robustness even if the model for $m(\cdot)$ is not specified correctly. This is an appealing property of $\widehat{\beta}_W$, but must be weighed against its larger asymptotic variance.

5 Simulation study

The finite-sample properties of the proposed estimators were evaluated via Monte Carlo simulation, with emphasis on the effects of sample size, percentage of observations censored, and percentage of censoring indicators missing. Response data were generated from the linear model: $Y = \alpha + \beta X + \epsilon$, with $\alpha = 3$ and $\beta = 0.5$. The random variables ϵ, X, and C were independently generated from N(0,1), U(0,2), and N(μ, 4) distributions, respectively. For each subject, the observed response was $\widetilde{Y} = min(Y, C)$ and the censoring indicator was $\delta = I(Y \leq C)$. Conditional on $\widetilde{Y} = \tilde{y}$, the probability that the censoring indicator was missing, $1 - \pi(\tilde{y})$, was determined by the logistic model: $log\{\pi(\tilde{y})/[1 - \pi(\tilde{y})]\} = \theta_1 + \theta_2\tilde{y}$.

Censoring rates of 20 and 40% were obtained by setting $\mu = 5.395$ and 4.067, respectively. Values for θ_1 and θ_2 were selected by specifying the proportion of subjects with a missing censoring indicator at $\tilde{y} = 0$ and by specifying the average proportion (over all $\tilde{y}$) with a missing censoring indicator. We simulated data so that the

average missingness rate was 20 or 40% and so that the missingness rate increased or decreased with $\tilde{y}$. We chose θ_1 so that $1 - \pi(0) = \{1 + exp(\theta_1)\}^{-1}$ was 0.15 or 0.25 when the average missingness rate was 20%, and so that $1 - \pi(0)$ was 0.30 or 0.50 when the average missingness rate was 40%. We also examined situations in which none of the censoring indicators were missing, both for 20 and 40% censoring, as well as the case where there was no censoring at all.

We generated 10,000 Monte Carlo random samples of size $n = 50, 100, 200$ and 400 under each combination of censoring and missingness. Every data set was analyzed under the model: $Y = \alpha + \beta X + \epsilon$. For each configuration, we averaged over the 10,000 data sets to estimate the mean squared error (MSE), bias, and standard error (SE) associated with the slope estimators $\widehat{\beta}_R, \widehat{\beta}_I$, and $\widehat{\beta}_W$. The results for the intercept estimators $\widehat{\alpha}_R, \widehat{\alpha}_I$, and $\widehat{\alpha}_W$ were qualitatively the same, so they are not presented. When none of the censoring indicators were missing, we also calculated the unbounded Koul et al. (1981) estimator:

$$\widehat{\beta}_K = \left(\sum_{i=1}^{n} X_i X_i^{\top} \right)^{-1} \sum_{i=1}^{n} X_i Y_i^{*},$$

where Y_i^{*} is defined in (1). We view $\widehat{\beta}_K$ as a reference for comparisons in our simulations.

We used the uniform kernel function $W(u) = \frac{1}{2}$ if $|u| \leq 1$ and $W(u) = 0$ otherwise, and the biweight kernel function $K(u) = \frac{15}{16}(1 - 2u^2 + u^4)$ if $|u| \leq 1$ and $K(u) = 0$ otherwise. The bandwidths were $h_n = b_n = n^{-1/3} max(\tilde{y})$. We estimated $m(z) = pr(\delta = 1|z)$ under the logistic model: $log\{m(z)/[1 - m(z)]\} = \gamma_1 + \gamma_2 \tilde{y} + \gamma_3 x$. When the data on δ are completely (or quasi-completely) separated, the maximum likelihood estimate of $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ does not exist (Albert and Anderson 1984; Santner and Duffy 1986). We excluded such data sets and continued simulating until we obtained 10,000 samples. The proportion of samples excluded did not exceed 0.3%, except when the average missingness rate was 40%, the censoring rate was 20%, and n was 50, in which case less than 2.3% of the samples were excluded.

The MSE, bias, and SE for $\widehat{\beta}_K, \widehat{\beta}_R, \widehat{\beta}_I$, and $\widehat{\beta}_W$ are presented in Tables 1, 2, and 3, respectively. The average jackknife estimate of SE is also shown in Table 3. Results are given only for situations where the average missingness rate increased with $\tilde{y}$, as the results for the decreasing cases were virtually identical. The Koul et al. (1981) estimator ($\widehat{\beta}_K$) is included as a benchmark when no censoring indicators are missing. The first row of each table corresponds to the special case of no censoring, where all of the estimators are identical because each synthetic response ($Y_i^{*}, \widehat{Y}_{i,R}, \widehat{Y}_{i,I}, \widehat{Y}_{i,W}$) reduces to the observed response Y_i.

Table 1 shows that when no censoring indicators were missing, the MSEs for $\widehat{\beta}_R$ were less than or equal to those for $\widehat{\beta}_K$, whereas the MSEs for $\widehat{\beta}_I$ and $\widehat{\beta}_W$ were slightly larger, at least for the 40% censoring case. For any given configuration, with or without missing censoring indicators, the MSE for $\widehat{\beta}_R$ never exceeded the MSE for $\widehat{\beta}_I$ or $\widehat{\beta}_W$. As expected, the MSEs for $\widehat{\beta}_R, \widehat{\beta}_I$, and $\widehat{\beta}_W$ decreased (i.e., improved) as sample sizes increased, as censoring rates decreased, and as average missingness rates decreased (see Table 1).

Table 1 Mean squared errors for the regression calibration ($\hat{\beta}_R$), imputation ($\hat{\beta}_I$), and inverse probability weighted ($\hat{\beta}_W$) estimators of β by censoring rate (CR), missingness rate (MR), and sample size (n) based on 10,000 simulated data sets with $\alpha = 3.0$ and $\beta = 0.5$

CR (%)	MR (%)	Estimator	Mean squared error by sample size			
			n = 50	n = 100	n = 200	n = 400
0	0	all	0.06	0.03	0.02	0.01
20	0	$\hat{\beta}_K$	0.37	0.19	0.10	0.05
		$\hat{\beta}_R$	0.35	0.18	0.09	0.05
		$\hat{\beta}_I$	0.39	0.19	0.10	0.05
		$\hat{\beta}_W$	0.39	0.19	0.10	0.05
20	20	$\hat{\beta}_R$	0.40	0.20	0.10	0.05
		$\hat{\beta}_I$	0.43	0.21	0.11	0.05
		$\hat{\beta}_W$	0.45	0.22	0.11	0.06
20	40	$\hat{\beta}_R$	0.49	0.23	0.11	0.06
		$\hat{\beta}_I$	0.52	0.24	0.12	0.06
		$\hat{\beta}_W$	0.60	0.28	0.14	0.07
40	0	$\hat{\beta}_K$	0.85	0.47	0.24	0.12
		$\hat{\beta}_R$	0.85	0.45	0.24	0.12
		$\hat{\beta}_I$	0.96	0.51	0.26	0.13
		$\hat{\beta}_W$	0.96	0.51	0.26	0.13
40	20	$\hat{\beta}_R$	0.95	0.50	0.26	0.13
		$\hat{\beta}_I$	1.05	0.55	0.27	0.14
		$\hat{\beta}_W$	1.14	0.60	0.30	0.15
40	40	$\hat{\beta}_R$	1.15	0.58	0.30	0.15
		$\hat{\beta}_I$	1.23	0.62	0.31	0.15
		$\hat{\beta}_W$	1.50	0.79	0.38	0.19

The Koul *et al.* (1981) estimator ($\hat{\beta}_K$) also was included when none of the censoring indicators were missing

The estimators also were evaluated with respect to bias and SE, the squares of which contribute equally to MSE. Table 2 gives the biases, all of which are negative when censoring is present; thus, each approach tended to underestimate the slope parameter. One interesting result is that applying our methods with missing censoring indicators usually produced less bias than applying the standard estimator of Koul et al. (1981) with complete censoring information. In fact, with a 40% censoring rate, the bias for each of our estimators (with missingness rates of either 20 or 40%) was always less than for $\hat{\beta}_K$ with no missing censoring indicators. Among the proposed estimators, the bias was consistently smallest for the inverse probability weighted estimator, $\hat{\beta}_W$. Also, biases decreased as sample sizes increased and as censoring rates decreased, but were not affected as much by missingness rates.

Table 3 gives the SEs and their jackknife estimates. The SE patterns mimic those for the MSEs in Table 1 because bias and SE contribute equally to MSE and the SEs are

Table 2 Biases for the regression calibration ($\hat{\beta}_R$), imputation ($\hat{\beta}_I$), and inverse probability weighted ($\hat{\beta}_W$) estimators of β by censoring rate (CR), missingness rate (MR), and sample size (n) based on 10,000 simulated data sets with $\alpha = 3.0$ and $\beta = 0.5$

			Bias by sample size			
CR (%)	MR (%)	Estimator	n = 50	n = 100	n = 200	n = 400
0	0	all	0.007	−0.001	0.001	−0.001
20	0	$\hat{\beta}_K$	−0.087	−0.056	−0.039	−0.026
		$\hat{\beta}_R$	−0.085	−0.059	−0.045	−0.035
		$\hat{\beta}_I$	−0.056	−0.034	−0.024	−0.016
		$\hat{\beta}_W$	−0.056	−0.034	−0.024	−0.016
20	20	$\hat{\beta}_R$	−0.086	−0.060	−0.046	−0.035
		$\hat{\beta}_I$	−0.061	−0.039	−0.030	−0.019
		$\hat{\beta}_W$	−0.054	−0.034	−0.026	−0.016
20	40	$\hat{\beta}_R$	−0.081	−0.062	−0.046	−0.034
		$\hat{\beta}_I$	−0.060	−0.046	−0.033	−0.022
		$\hat{\beta}_W$	−0.045	−0.036	−0.025	−0.015
40	0	$\hat{\beta}_K$	−0.189	−0.122	−0.089	−0.061
		$\hat{\beta}_R$	−0.151	−0.091	−0.067	−0.050
		$\hat{\beta}_I$	−0.117	−0.066	−0.048	−0.033
		$\hat{\beta}_W$	−0.117	−0.066	−0.048	−0.033
40	20	$\hat{\beta}_R$	−0.151	−0.091	−0.068	−0.046
		$\hat{\beta}_I$	−0.124	−0.073	−0.053	−0.033
		$\hat{\beta}_W$	−0.117	−0.070	−0.051	−0.031
40	40	$\hat{\beta}_R$	−0.151	−0.092	−0.066	−0.040
		$\hat{\beta}_I$	−0.129	−0.080	−0.055	−0.030
		$\hat{\beta}_W$	−0.116	−0.078	−0.049	−0.025

The Koul et al. (1981) estimator ($\hat{\beta}_K$) also was included when none of the censoring indicators were missing

larger (in absolute value) than the biases. As with the MSEs, $\widehat{\beta}_R$ had the smallest SEs and $\widehat{\beta}_W$ had the largest. The SEs decreased as n increased, as censoring decreased, and as missingness decreased. The jackknife estimates were always good for $\widehat{\beta}_W$, but became too small for $\widehat{\beta}_I$, and even smaller for $\widehat{\beta}_R$, as n decreased, as censoring increased, and as missingness increased.

In order to assess robustness, we analyzed the same simulated data with a "poor" model choice for $m(z)$. Rather than using $log\{m(z)/[1 - m(z)]\} = \gamma_1 + \gamma_2\tilde{y} + \gamma_3 x$ to model $m(z)$ as a logistic function of $\tilde{y}$ and x, we set $\gamma_2 = \gamma_3 = 0$ and modeled $m(z)$ as a constant, which gives $\hat{m}(z) \equiv \hat{m} = \sum_{i=1}^{n} \xi_i \delta_i / \sum_{i=1}^{n} \xi_i$. Table 4 displays the biases and SEs obtained for this poor model choice. The results for no missing censoring indicators ($MR = 0\%$) are the same as in Tables 2 and 3, except for $\hat{\beta}_R$, as only $\hat{\beta}_R$ depends on $\hat{m}(z)$ in this situation. The SEs always decrease as sample sizes increase, as does the bias of $\hat{\beta}_W$, but the biases of $\hat{\beta}_R$ and $\hat{\beta}_I$ actually increase with the

Table 3 Standard errors (and mean jackknife estimates) for the regression calibration ($\hat{\beta}_R$), imputation ($\hat{\beta}_I$), and inverse probability weighted ($\hat{\beta}_W$) estimators of β by censoring rate (CR), missingness rate (MR), and sample size (n) based on 10,000 simulated data sets with $\alpha = 3.0$ and $\beta = 0.5$

CR (%)	MR (%)	Estimator	Standard error (and Jackknife estimate) by sample size			
			n = 50	n = 100	n = 200	n = 400
0	0	all	0.25 (0.24)	0.17 (0.17)	0.12 (0.12)	0.09 (0.09)
20	0	$\hat{\beta}_K$	0.60 (0.59)	0.43 (0.43)	0.31 (0.32)	0.22 (0.23)
		$\hat{\beta}_R$	0.59 (0.45)	0.42 (0.34)	0.30 (0.25)	0.21 (0.18)
		$\hat{\beta}_I$	0.62 (0.61)	0.44 (0.44)	0.31 (0.32)	0.22 (0.23)
		$\hat{\beta}_W$	0.62 (0.61)	0.44 (0.44)	0.31 (0.32)	0.22 (0.23)
20	20	$\hat{\beta}_R$	0.63 (0.45)	0.44 (0.34)	0.31 (0.25)	0.22 (0.18)
		$\hat{\beta}_I$	0.65 (0.58)	0.46 (0.42)	0.32 (0.31)	0.23 (0.22)
		$\hat{\beta}_W$	0.67 (0.65)	0.47 (0.47)	0.33 (0.34)	0.24 (0.25)
20	40	$\hat{\beta}_R$	0.70 (0.46)	0.48 (0.34)	0.33 (0.25)	0.24 (0.18)
		$\hat{\beta}_I$	0.72 (0.55)	0.49 (0.41)	0.34 (0.30)	0.25 (0.21)
		$\hat{\beta}_W$	0.77 (0.70)	0.53 (0.52)	0.37 (0.38)	0.26 (0.27)
40	0	$\hat{\beta}_K$	0.90 (0.88)	0.67 (0.67)	0.48 (0.49)	0.35 (0.36)
		$\hat{\beta}_R$	0.91 (0.76)	0.67 (0.58)	0.48 (0.43)	0.34 (0.31)
		$\hat{\beta}_I$	0.97 (0.94)	0.71 (0.70)	0.51 (0.51)	0.36 (0.37)
		$\hat{\beta}_W$	0.97 (0.94)	0.71 (0.70)	0.51 (0.51)	0.36 (0.37)
40	20	$\hat{\beta}_R$	0.97 (0.76)	0.70 (0.58)	0.50 (0.43)	0.36 (0.32)
		$\hat{\beta}_I$	1.02 (0.91)	0.74 (0.68)	0.52 (0.50)	0.37 (0.36)
		$\hat{\beta}_W$	1.06 (1.01)	0.77 (0.75)	0.54 (0.55)	0.39 (0.40)
40	40	$\hat{\beta}_R$	1.06 (0.77)	0.76 (0.59)	0.54 (0.44)	0.39 (0.32)
		$\hat{\beta}_I$	1.10 (0.88)	0.78 (0.66)	0.56 (0.49)	0.39 (0.35)
		$\hat{\beta}_W$	1.22 (1.11)	0.89 (0.84)	0.62 (0.61)	0.44 (0.45)

The Koul et al. (1981) estimator ($\hat{\beta}_K$) also was included when none of the censoring indicators were missing

sample size in many cases. In fact, the bias of $\hat{\beta}_R$ is now positive and increases with n in every situation, whereas it was always negative and approached 0 as n increased when using the better model for $m(z)$. The bias of $\hat{\beta}_I$ increases with the sample size, which makes the negative biases smaller in absolute value, but once the biases become positive, they grow in absolute value as n increases. In contrast, the bias of $\hat{\beta}_W$ for the poor model choice is nearly identical to the bias of $\hat{\beta}_W$ when using the better model. As predicted theoretically, the double-robustness property of $\hat{\beta}_W$ apparently keeps its performance from degrading if a poor choice is made when modeling $m(z)$.

6 Example: analysis of glioblastoma data

We illustrate our methods with the glioblastoma data introduced earlier. These data are from a brain cancer clinical trial. As a measure of post-treatment quality of life, we focus on the number of days until a patient declines to a state in which he has lost

Table 4 Biases (and standard errors) for the regression calibration ($\hat{\beta}_R$), imputation ($\hat{\beta}_I$), and inverse probability weighted ($\hat{\beta}_W$) estimators of β for a poor choice of $m(z)$ [i.e., $m(z) \equiv m$] by censoring rate (CR), missingness rate (MR), and sample size (n) based on 10,000 simulated data sets with $\alpha = 3.0$ and $\beta = 0.5$

			Bias (and standard error) by sample size			
CR (%)	MR (%)	Estimator	n = 50	n = 100	n = 200	n = 400
0	0	all	0.007 (0.25)	−0.001 (0.17)	0.001 (0.12)	−0.001 (0.09)
20	0	$\hat{\beta}_K$	−0.087 (0.60)	−0.056 (0.43)	−0.039 (0.31)	−0.026 (0.22)
		$\hat{\beta}_R$	0.136 (0.42)	0.148 (0.30)	0.163 (0.22)	0.167 (0.16)
		$\hat{\beta}_I$	−0.056 (0.62)	−0.034 (0.44)	−0.024 (0.31)	−0.016 (0.22)
		$\hat{\beta}_W$	−0.056 (0.62)	−0.034 (0.44)	−0.024 (0.31)	−0.016 (0.22)
20	20	$\hat{\beta}_R$	0.137 (0.43)	0.149 (0.31)	0.163 (0.22)	0.166 (0.16)
		$\hat{\beta}_I$	−0.016 (0.59)	0.003 (0.42)	0.012 (0.29)	0.020 (0.21)
		$\hat{\beta}_W$	−0.054 (0.67)	−0.033 (0.47)	−0.026 (0.33)	−0.016 (0.24)
20	40	$\hat{\beta}_R$	0.137 (0.44)	0.151 (0.32)	0.164 (0.23)	0.167 (0.17)
		$\hat{\beta}_I$	0.027 (0.56)	0.040 (0.40)	0.049 (0.28)	0.056 (0.20)
		$\hat{\beta}_W$	−0.046 (0.76)	−0.033 (0.53)	−0.026 (0.37)	−0.016 (0.26)
40	0	$\hat{\beta}_K$	−0.189 (0.90)	−0.122 (0.67)	−0.089 (0.48)	−0.061 (0.35)
		$\hat{\beta}_R$	0.111 (0.53)	0.144 (0.40)	0.159 (0.29)	0.167 (0.21)
		$\hat{\beta}_I$	−0.117 (0.97)	−0.066 (0.71)	−0.048 (0.51)	−0.033 (0.36)
		$\hat{\beta}_W$	−0.117 (0.97)	−0.066 (0.71)	−0.048 (0.51)	−0.033 (0.36)
40	20	$\hat{\beta}_R$	0.109 (0.54)	0.141 (0.40)	0.156 (0.29)	0.164 (0.21)
		$\hat{\beta}_I$	−0.078 (0.90)	−0.031 (0.66)	−0.014 (0.47)	0.003 (0.33)
		$\hat{\beta}_W$	−0.122 (1.06)	−0.071 (0.77)	−0.053 (0.54)	−0.032 (0.39)
40	40	$\hat{\beta}_R$	0.106 (0.55)	0.137 (0.42)	0.152 (0.30)	0.160 (0.22)
		$\hat{\beta}_I$	−0.031 (0.84)	0.004 (0.60)	0.019 (0.43)	0.037 (0.30)
		$\hat{\beta}_W$	−0.119 (1.22)	−0.079 (0.87)	−0.058 (0.61)	−0.032 (0.43)

The Koul et al. (1981) estimator ($\hat{\beta}_K$) also was included when none of the censoring indicators were missing

his mobility and his cancer has returned. Thus, the endpoint of interest is the time (in days) to non-ambulatory progression, and we want to relate that outcome to the patient's sex and age at study entry.

We have data on 276 glioblastoma patients who were over 50 years old when they entered the trial. By the end of the study, 59 of these patients progressed and were no longer ambulatory, 14 progressed and were still ambulatory, 166 did not progress, and 37 progressed but had an unknown ambulatory status. Therefore, with respect to the time to non-ambulatory progression, we have 59 uncensored observations ($\xi = 1$, $\delta = 1$), 180 censored observations ($\xi = 1$, $\delta = 0$), and 37 observations with a missing censoring indicator ($\xi = 0$). We have data on a covariate vector $X = (X_1, X_2, X_3)^\top$, where X_1 is fixed at 1, X_2 indicates the patient's sex (0 for female, 1 for male), and X_3 is the patient's age (in years) at study entry. Our data set includes information on 109 women and 167 men, with ages ranging from 51 to 74.

Table 5 Estimated regression coefficients (and jackknife estimates of their standard errors) under two models for the glioblastoma data from the brain cancer clinical trial (n = 276)

Model terms		Type of estimator		
		Regression calibration	Imputation	Inverse probability weighted
		Model 1: Parallel lines (sex-specific intercepts, common age slope)		
Intercept	$\hat{\beta}_1$	18.45 (3.89)	22.64 (8.85)	23.23 (10.21)
Sex	$\hat{\beta}_2$	−1.51 (0.97)	−1.23 (1.28)	−1.20 (1.50)
Age	$\hat{\beta}_3$	−0.224 (0.059)	−0.310 (0.142)	−0.323 (0.163)
		Model 2: Individual lines (separate intercept and age slope per sex)		
Intercept	$\hat{\beta}_1$	25.64 (7.55)	25.94 (10.15)	25.81 (11.73)
Sex	$\hat{\beta}_2$	−11.97 (8.73)	−6.27 (16.62)	−5.23 (19.19)
Age	$\hat{\beta}_3$	−0.342 (0.117)	−0.364 (0.156)	−0.366 (0.180)
Sex-age interaction	$\hat{\beta}_4$	0.172 (0.134)	0.083 (0.258)	0.067 (0.297)

We applied our methods to the glioblastoma data to estimate the effects of sex and age on the time to non-ambulatory progression. As a first step, we fitted the linear regression model: $E(Y|X = x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, where Y is the log of the time to non-ambulatory progression. This initial model specifies parallel lines, with a separate intercept for each sex and a common slope in age. To help assess the appropriateness of the model, we added a sex-age interaction term ($\beta_4 x_2 x_3$), which allows each line to have its own sex-specific intercept and slope. Estimated regression coefficients and their standard errors, obtained under both models via all three proposed methods, are shown in Table 5. As a further means of assessing model suitability, we also fitted a separate quadratic model in age for each sex, as well as a model with quadratic-age curves that shared common age terms, but none of the quadratic terms were significant, so these results are not presented.

Based on 2-sided Wald tests, none of the analyses provided evidence of a sex effect, but to varying degrees they all suggested a detrimental age effect, with time to non-ambulatory progression decreasing with age. The expanded model gave no hint of a sex-age interaction, regardless of which method was used, so we focus on the simpler model with only main effects for sex and age. Relative to the typical 0.05 level, the statistical significance of the age effect was marginal ($p = 0.047$) in the inverse probability weighted analysis, slightly stronger ($p = 0.029$) in the imputation analysis, and fairly high ($p < 0.001$) in the regression calibration analysis. Though excluding 13% of the data seems inefficient, if we ignore the 37 observations with a missing censoring indicator, the Koul et al. (1981) analysis gives similar results, with no evidence of a sex effect and a significant age effect ($p = 0.022$).

Thus, time to non-ambulatory progression does not differ significantly between men and women, but it appears to decrease (i.e., quality of life worsens) with age. Although evidence of this age effect is weak for the inverse probability weighted analysis, it is consistent across all three of the proposed analyses, as well as the Koul et al. (1981) analysis. The relative sizes of the estimated standard errors in Table 5 agree with our simulation results, where the SE for $\hat{\beta}_R$ was smallest (and its jackknife

estimates were too small), followed by the SE for $\hat{\beta}_I$, and finally by the SE for $\hat{\beta}_W$ (with good jackknife estimates). This suggests that the test based on $\hat{\beta}_W$ might be the least powerful, which further strengthens the conclusion that the observed age effect is real, as even the least powerful test was statistically significant.

In this example, expressing the failure time of interest as a function of two component times reveals a potential source of bias. That is, the time to non-ambulatory progression can be viewed as the maximum of the time to loss of mobility and the time to cancer progression. We treat patients who have progressed but are still ambulatory at the end of the study as contributing censored failure times, but this censoring is likely informative and may introduce a bias, as ambulatory patients who have already progressed should be at greater risk of failing than ambulatory patients who have not yet progressed. Similarly, patients in remission who have already lost their mobility should be at greater risk of failing than patients in remission who are still ambulatory. This problem is not unique to our approach, however, and affects a wide class of existing survival techniques, including something as simple as using a Kaplan–Meier curve to estimate the distribution of time to non-ambulatory progression. Future efforts should be directed at developing methods that properly account for this type of informative censoring, such as a multivariate procedure, although such developments are beyond the scope of this article.

7 Discussion

We propose three estimators for the regression coefficients in a linear least squares analysis of censored survival data when some censoring indicators are missing. Our methods generalize the synthetic data methods of Koul et al. (1981), which do not allow for missing censoring indicators. The proposed estimators modify the synthetic responses of Koul *et al.* so that each missing censoring indicator (δ) is replaced by its estimated conditional expectation. Our estimators differ from each other only with respect to how they handle the known censoring indicators. The regression calibration estimator replaces each known δ by its estimated mean, the imputation estimator uses the known δ, and the inverse probability weighted estimator substitutes a weighted average of the known δ and its estimated mean.

Similar to Dikta (1998), our regression calibration method can be used even when all of the censoring indicators are observed. In that situation, the Koul et al. (1981) estimator ($\widehat{\beta}_K$) transforms the observations ($\widetilde{Y}_i, \delta_i$) into synthetic responses (Y_i^*) and then applies a modified least squares procedure to (Y_i^*, X_i). Their approach is motivated by the fact that $E[\delta_i \widetilde{Y}_i \{1 - G(\widetilde{Y}_i)\}^{-1} | X_i] = X_i^\top \beta$, but $\widehat{\beta}_K$ can perform poorly with small sample sizes. Under the MAR assumption, we have

$$var\left\{ \frac{\delta_i \widetilde{Y}_i}{1 - G(\widetilde{Y}_i)} \bigg| X_i \right\} \geq var\left\{ \frac{m(Z_i)\widetilde{Y}_i}{1 - G(\widetilde{Y}_i)} \bigg| X_i \right\},$$

which implies that $\widehat{\beta}_K$ might be improved by replacing δ_i in Y_i^* with a regression estimator of $m(Z_i)$ based on $\{(\widetilde{Y}_i, \delta_i) : i = 1, \ldots, n\}$. Our simulation results support this suggestion, but a detailed discussion is beyond the scope of this paper.

Along these same lines, we show in Sects. 3 and 4 that $\widehat{\beta}_I$ and $\widehat{\beta}_W$ have larger asymptotic variances than $\widehat{\beta}_R$, and thus our imputation and inverse probability weighted estimators are asymptotically less efficient than our regression calibration estimator. The simulation studies confirm these results (see Tables 3 and 4). This suggests that, at least with respect to this measure of performance, the regression calibration method for handling missing censoring indicators is better than the other two methods, even though the other two methods can lead to asymptotically efficient estimators in some situations (see, e.g., Robins et al. 1994; Hahn 1998; Hirano et al. 2003; and Wang et al. 2004).

On the other hand, there are several good reasons to favor $\widehat{\beta}_W$ over $\widehat{\beta}_I$ or $\widehat{\beta}_R$. The inverse probability weighted estimator is less biased than the other estimators (see Tables 2 and 4) and its jackknife estimates of SE were the most accurate (see Table 3). Also, the inverse probability weighted estimator enjoys the double-robustness property, which makes it less susceptible to problems caused by incorrectly specified models (see Table 4).

How should one balance the advantages and disadvantages of each approach? In general, we recommend the inverse probability weighted estimator. Not only is the inverse probability weighted estimator more robust to poor model choices for $m(z)$, but relative to the other estimators, it is less biased and its jackknife variance estimates are more accurate. Although the regression calibration estimator has the smallest variance, its relatively large bias can produce misleading results in practice. In fact, our simulation showed that if a poor model was selected for $m(z)$, the bias of $\widehat{\beta}_R$ could actually increase with the sample size. Currently, the variance estimator for $\widehat{\beta}_R$ does not adequately account for uncertainty in estimating $m(z)$ and $G(z)$. Also, the variance estimates are functions of the parameter estimates and $\widehat{\beta}_R$ has the largest bias; thus, even though the true variance of $\widehat{\beta}_R$ is smallest, the variance of $\widehat{\beta}_R$ also can be severely underestimated (see Table 3), making results appear too significant. Future efforts should focus on accounting for uncertainty in estimating $m(z)$ and $G(z)$, but until then, we cannot recommend using the regression calibration method in its present form.

Clearly, our estimators depend on the choice of bandwidths. However, each $\hat{\beta}$ is a global functional and thus the rate $n^{1/2}$ asymptotic normality of the proposed estimators suggests that a proper choice of b_n in assumption (C.7) does not depend on the first order term of the mean squared error. Rather, it may depend only on second (or higher) order terms. This implies that the selection of b_n may be fairly flexible as long as the bandwidths satisfy (C.7).

We assume the censoring variables $(C_1, \ldots, C_n)$ are $i.i.d.$, which is the same assumption made by Koul et al. (1981). This form of censoring is more restrictive than the standard noninformative censoring assumed in most survival analyses, as the latter does not require the censoring distribution $G(c)$ to be estimated. Our methods, however, as well as those developed by Koul *et al.*, rely on synthetic data that are functions of an estimate of $G(c)$. One avenue of future research might involve extending our methods by expressing $G(c)$ as a function of the covariates. Generally, this requires either a parametric model, the choice of which must be justified, or else a nonparametric approach, which could suffer in high-dimensional situations if the data

were too sparse. Alternatively, we might view $G(c)$ as the marginal distribution of C, which "averages" over the covariates.

Appendix: Proofs of theorems

Define $Grad[m_0(z, \theta)] = \left(\frac{\partial m_0(z,\theta)}{\partial \theta_1}, \frac{\partial m_0(z,\theta)}{\partial \theta_2}, \ldots, \frac{\partial m_0(z,\theta)}{\partial \theta_d}, \frac{\partial m_0(z,\theta)}{\partial \theta_{d+1}} \right)^\top$ and set $A(\theta) = (\sigma_{r,s})$ for $1 \leq r, s \leq d+1$, where $\sigma_{r,s} = E\left[\frac{\pi(\widetilde{Y})}{m_0(Z,\theta)\{1-m_0(Z,\theta)\}} \frac{\partial m_0(Z,\theta)}{\partial \theta_r} \frac{\partial m_0(Z,\theta)}{\partial \theta_s} \right]$.
Furthermore, let $\alpha(z_1, z_2) = Grad[m_0(z_1, \theta)]^\top A^{-1}(\theta) Grad[m_0(z_2, \theta)]$ and assume the following conditions:

(C.1) $m_0(z, \theta)$ has bounded first-order derivatives on θ.
(C.2) $\sigma_{r,s} < \infty$ for $1 \leq r, s \leq d+1$ and the matrix $A(\theta) = (\sigma_{r,s})$ is positive definite.
(C.3) $E\left\{ \|X\|^2 \widetilde{Y}^2 / \bar{H}^2(\widetilde{Y}) \right\} < \infty$, where $H(t) = pr(\widetilde{Y} \leq t)$ and $\bar{H}(t) = 1 - H(t)$.
(C.4) $E\left\{ \|X\|^2 \right\} < \infty$.
(C.5) $\pi(\cdot)$ has a bounded first-order derivative.
(C.6) $W(\cdot)$ is of order 1 with bounded support.
(C.7) $nb_n \longrightarrow \infty$ and $nb_n^2 \longrightarrow 0$.

Proof of Theorem 2.1 For $\Sigma_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$ and $A_n = n^{-1/2} \sum_{i=1}^n X_i (\widehat{Y}_{i,R} - X_i^\top \beta)$, we can write

$$n^{1/2}(\widehat{\beta}_R - \beta) = \Sigma_n^{-1} A_n. \tag{A.1}$$

By adding and subtracting identical terms, we can rewrite A_n as follows:

$$A_n = n^{-1/2} \sum_{i=1}^n X_i (Y_{i,R} - X_i^\top \beta) + n^{-1/2} \sum_{i=1}^n X_i \frac{\widetilde{Y}_i \{m_0(Z_i, \widehat{\theta}_n) - m_0(Z_i, \theta)\}}{1 - G(\widetilde{Y}_i)}$$
$$+ n^{-1/2} \sum_{i=1}^n X_i \left\{ \frac{\widetilde{Y}_i m_0(Z_i, \theta)}{1 - G(\widetilde{Y}_i)} \right\} \frac{\widehat{G}_n(\widetilde{Y}_i) - G(\widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)} + o_p(1)$$
$$:= T_{n1} + T_{n2} + T_{n3} + o_p(1). \tag{A.2}$$

Define $\mu(z) = E\left\{ \frac{X\widetilde{Y}}{1-G(\widetilde{Y})} \alpha(Z, Z_j) \Big| Z_j = z \right\}$. Under conditions (C.1) and (C.2), we have

$$T_{n2} = n^{-1/2} \sum_{j=1}^n \frac{\mu(Z_j)\xi_j\{\delta_j - m_0(Z_j, \theta)\}}{m_0(Z_j, \theta)\{1 - m_0(Z_j, \theta)\}} + o_p(1). \tag{A.3}$$

Set $\widetilde{H}_0(t) = pr(\widetilde{Y}_i > t, \delta_i = 0)$ and define

$$
\psi(\widetilde{Y}_i, \delta_i, \xi_i; t) = \left[-\frac{\{\xi_i - \pi(\widetilde{Y}_i)\}\{\delta_i - u(\widetilde{Y}_i)\}}{\pi(\widetilde{Y}_i)\{1 - H(\widetilde{Y}_i)\}} I(\widetilde{Y}_i \le t) \right.
$$
$$
\left. + \int_0^{\widetilde{Y}_i \wedge t} \frac{d\widetilde{H}_0(s)}{\{1 - H(s)\}^2} + \frac{I(\widetilde{Y}_i \le t, \delta_i = 0)}{1 - H(\widetilde{Y}_i)} \right], \qquad (A.4)
$$

where $u(\cdot)$ is given in Sect. 2. Similar to Wang and Ng (2008), we have

$$
\widehat{G}_n(t) - G(t) = \frac{1 - G(t)}{n} \sum_{i=1}^n \psi(\widetilde{Y}_i, \delta_i, \xi_i; t) + o_p(n^{-\frac{1}{2}}).
$$

The third term in (A.2) can be rewritten as

$$
T_{n3} = n^{-3/2} \sum_{i=1}^n \sum_{j=1}^n X_i \frac{\widetilde{Y}_i m_0(Z_i, \theta) \psi(\widetilde{Y}_j, \delta_j, \xi_j; \widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)} + o_p(1). \qquad (A.5)
$$

Define

$$
h\left(X_i, \widetilde{Y}_i, \delta_i, \xi_i; X_j, \widetilde{Y}_j, \delta_j, \xi_j\right) = \frac{X_i \widetilde{Y}_i m_0(Z_i, \theta) \psi(\widetilde{Y}_j, \delta_j, \xi_j; \widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)}
$$
$$
+ \frac{X_j \widetilde{Y}_j m_0(Z_j, \theta) \psi(\widetilde{Y}_i, \delta_i, \xi_i; \widetilde{Y}_j)}{1 - G(\widetilde{Y}_j)}
$$

and set $U_n = n^{-3/2} \sum_{i<j} h(X_i, \widetilde{Y}_i, \delta_i, \xi_i; X_j, \widetilde{Y}_j, \delta_j, \xi_j)$, which leads to

$$
T_{n3} = U_n + o_p(1). \qquad (A.6)
$$

Combining (A.2), (A.3), and (A.6) gives

$$
A_n = T_{n1} + M_n + U_n + o_p(1), \qquad (A.7)
$$

where T_{n1} is defined in (A.2) and M_n is the main component of T_{n2} in (A.3).

By the central limit theorem, we have

$$
T_{n1} \longrightarrow N(0, \Omega_1), \qquad (A.8)
$$

and under the assumed missingness mechanism, we have

$$
M_n \xrightarrow{\mathcal{L}} N(0, \Omega_2), \qquad (A.9)
$$

where $\Omega_1 = E\{XX^\top(Y_R - X^\top\beta)^2\}$ and $\Omega_2 = E\left[\frac{\mu(\tilde{Y})\mu^\top(\tilde{Y})\pi(\tilde{Y})}{m_0(Z,\theta)\{1-m_0(Z,\theta)\}}\right]$. For any d-vector of constants, say α, let $h_\alpha(\cdot) = \alpha^\top h(\cdot)$, $U_{n\alpha} = \alpha^\top U_n$, and $g_\alpha(\cdot) = \alpha^\top g(\cdot)$, where

$$g(X_1, \tilde{Y}_1, \delta_1, \xi_1) = E\left\{\frac{X_2\tilde{Y}_2m_0(Z_2,\theta)\psi(\tilde{Y}_1, \delta_1, \xi_1; \tilde{Y}_2)}{1-G(\tilde{Y}_2)}\bigg| X_1, \tilde{Y}_1, \delta_1, \xi_1\right\}.$$

Under the assumed missingness mechanism, we have

$$E\{\psi(\tilde{Y}_i, \delta_i, \xi_i; \tilde{Y}_j)|\tilde{Y}_j\} = 0 \qquad \text{for } i \neq j,$$

from which it is straightforward to verify that $E\{h_\alpha(X_1, \tilde{Y}_1, \delta_1, \xi_1; X_2, \tilde{Y}_2, \delta_2, \xi_2)\} = 0$. As a result of the relationship

$$E\{\psi^2(\tilde{Y}_i, \delta_i, \xi_i; \tilde{Y}_j)|\tilde{Y}_j\} = -\{1-G(\tilde{Y}_j)\}^2 \int_0^{\tilde{Y}_j} \{\bar{H}(s)\}^{-2}\, d\tilde{H}_0(s)$$

$$\leq \frac{\{1-G(\tilde{Y}_j)\}^2}{\bar{H}^2(\tilde{Y}_j)} \quad \text{for } i \neq j$$

and condition (C.3), it follows that $E(h_\alpha^2) < \infty$. By the central limit theorem for U-statistics, we have

$$U_{n\alpha} \xrightarrow{\mathcal{L}} N(0, \alpha^\top\Omega_3\alpha),$$

where

$$\Omega_3 = E\{g(X_1, \tilde{Y}_1, \delta_1, \xi_1)g^\top(X_1, \tilde{Y}_1, \delta_1, \xi_1)\}$$

and $\alpha^\top\Omega_3\alpha = E(g_\alpha^2)$. This result implies

$$U_n \xrightarrow{\mathcal{L}} N(0, \Omega_3). \tag{A.10}$$

As for covariances, under the assumed missingness mechanism, we have

$$cov(T_{n1}, M_n) = E\left[X_i\left\{\frac{\tilde{Y}_im_0(Z_i)}{1-G(Z_i)} - X_i^\top\beta\right\}\frac{\mu^\top(\tilde{Y}_i)\pi(\tilde{Y}_i)\{\delta_i - m_0(\tilde{Y}_i)\}}{m_0(Z_i,\theta)\{1-m_0(Z_i,\theta)\}}\right] = 0 \tag{A.11}$$

Note that $E\{X_k(Y_{k,R} - X_k^\top\beta)h_{ij}^\top\} = 0$ for $k \neq i, k \neq j, i \neq j$ and $E\{X_i(Y_{i,R} - X_i^\top\beta)h_{ij}\} = E\{X_j(Y_{j,R} - X_j^\top\beta)h_{ij}\}$, where $h_{ij} = h(X_i, \tilde{Y}_i, \delta_i, \xi_i; X_j, \tilde{Y}_j, \delta_j, \xi_j)$.

Similarly, we have

$$
\begin{aligned}
cov&(T_{n1}, U_n) \\
&= \left(1 - \frac{1}{n}\right) E\left\{ E\left[X_1(Y_{1,R} - X_1^\top \beta)\psi(\widetilde{Y}_1, \delta_1, \xi_1; \widetilde{Y}_2)\Big|\widetilde{Y}_2\right] \frac{X_2^\top \widetilde{Y}_2 m_0(Z_2, \theta)}{1 - G(\widetilde{Y}_2)} \right\} \\
&\longrightarrow E\left\{ E\left[X_1(Y_{1,R} - X_1^\top \beta)\psi(\widetilde{Y}_1, \delta_1, \xi_1; \widetilde{Y}_2)\Big|\widetilde{Y}_2\right] \frac{X_2^\top \widetilde{Y}_2 m_0(Z_2, \theta)}{1 - G(\widetilde{Y}_2)} \right\} \\
&= E\left\{ X_1(Y_{1,R} - X_1^\top \beta) \frac{X_2^\top \widetilde{Y}_2 m_0(Z_2, \theta)}{1 - G(\widetilde{Y}_2)} \left[\int_0^{\widetilde{Y}_2 \wedge \widetilde{Y}_1} \frac{d\widetilde{H}_0(s)}{\{1 - H(s)\}^2} + \frac{I(\widetilde{Y}_1 \le \widetilde{Y}_2, \delta_1 = 0)}{1 - H(\widetilde{Y}_1)} \right] \right\} \\
&:= \Omega_{1,3}.
\end{aligned}
\tag{A.12}
$$

and

$$
\begin{aligned}
cov&(M_n, U_n) \\
&= -\left(1 - \frac{1}{n}\right) E\left[\frac{m_0(Z_2, \theta)\widetilde{Y}_2 X_2 \mu^\top(Y_1)}{\{1 - G(\widetilde{Y}_2)\}\{1 - H(\widetilde{Y}_1)\}} I(\widetilde{Y}_1 \le \widetilde{Y}_2) \right] \\
&\longrightarrow -E\left[\frac{m_0(Z_2, \theta)\widetilde{Y}_2 X_2 \mu^\top(\widetilde{Y}_1)}{\{1 - G(\widetilde{Y}_2)\}\{1 - H(\widetilde{Y}_1)\}} I(\widetilde{Y}_1 \le \widetilde{Y}_2) \right] := \Omega_{2,3}.
\end{aligned}
\tag{A.13}
$$

Together Eqs. (A.7)–(A.9) and (A.10)–(A.13) prove

$$
A_n \xrightarrow{\mathcal{L}} N(0, \Omega_R),
\tag{A.14}
$$

where

$$
\Omega_R = \Omega_1 + \Omega_2 + \Omega_3 + 2\Omega_{1,3} + 2\Omega_{2,3}.
\tag{A.15}
$$

By the law of large numbers and (C.4), it follows that $n^{-1} \sum_{i=1}^n X_i X_i^\top \xrightarrow{p} \Sigma$. This result, together with (A.1) and (A.14), prove Theorem 2.1.

Proof of Theorem 2.2 Asymptotic representations similar to (A.7) can be obtained for $\sum_{i=1}^n (\widehat{Y}_{i,R} - X_i^\top \widehat{\beta}_R)^2 X_i X_i^\top$ for $i = 1, \cdots, n$. Then, by applying these representations to (2) and using some algebra, Theorem 2.2 can be proved.

Proof of Theorem 3.1 By standard arguments, we have

$$
\begin{aligned}
n^{-1/2} &\sum_{i=1}^n X_i(\widehat{Y}_{i,I} - X_i^\top \beta) \\
&= n^{-1/2} \sum_{i=1}^n X_i \left[\frac{\{\delta_i \xi_i + (1 - \xi_i)m_0(Z_i, \theta)\}\widetilde{Y}_i}{1 - G(\widetilde{Y}_i)} - X_i^\top \beta \right]
\end{aligned}
$$

$$+ n^{-1/2} \sum_{i=1}^{n} X_i \frac{(1 - \xi_i)\{m_0(Z_i, \widehat{\theta}_n) - m_0(Z_i, \theta)\}\widetilde{Y}_i}{1 - G(\widetilde{Y}_i)}$$

$$+ n^{-1/2} \sum_{i=1}^{n} \frac{X_i \widetilde{Y}_i \{\delta_i \xi_i + (1 - \xi_i) m_0(Z_i, \theta)\}}{1 - G(\widetilde{Y}_i)} \left\{ \frac{\widehat{G}_n(\widetilde{Y}_i) - G(\widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)} \right\} + o_p(1)$$

$$= T_{n1} + T_{n2} + T_{n3} + T_{n4} + T_{n5} + T_{n6} + o_p(1), \tag{A.16}$$

where T_{n1}, T_{n2} and T_{n3} are defined in (A.2), and T_{n4}, T_{n5} and T_{n6} are defined as follows

$$T_{n4} = n^{-1/2} \sum_{i=1}^{n} X_i \frac{\widetilde{Y}_i \xi_i \{\delta_i - m_0(Z_i, \theta)\}}{1 - G(\widetilde{Y}_i)},$$

$$T_{n5} = -n^{-1/2} \sum_{i=1}^{n} X_i \frac{\widetilde{Y}_i \xi_i \{m_0(Z_i, \widehat{\theta}_n) - m_0(Z_i, \theta)\}}{1 - G(\widetilde{Y}_i)},$$

$$T_{n6} = n^{-1/2} \sum_{i=1}^{n} X_i \frac{\widetilde{Y}_i \xi_i \{\delta_i - m_0(Z_i, \theta)\}}{1 - G(\widetilde{Y}_i)} \left\{ \frac{\widehat{G}_n(\widetilde{Y}_i) - G(\widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)} \right\}.$$

By (A.2) and (A.7), we have

$$T_{n1} + T_{n2} + T_{n3} = T_n + M_n + U_n + o_p(1). \tag{A.17}$$

Let $\widetilde{\mu}(Z_j) = E \left\{ \frac{X \widetilde{Y} \pi(\widetilde{Y})}{1 - G(\widetilde{Y})} \alpha(Z, Z_j) \middle| Z_j \right\}$. Under conditions (C.1) and (C.2), using arguments similar to those leading to (A.3)–(A.5), we get

$$T_{n5} = -n^{-1/2} \sum_{j=1}^{n} \frac{\xi_j \{\delta_j - m(Z_j, \theta)\}}{m_0(Z_j, \theta)\{1 - m_0(Z_j, \theta)\}} \widetilde{\mu}(Z_j) + o_p(1). \tag{A.18}$$

Recalling the definition of T_{n4}, it follows from (A.18) that

$$T_{n4} + T_{n5} = n^{-1/2} \sum_{i=1}^{n} \left[\frac{X_i \widetilde{Y}_i}{1 - G(\widetilde{Y}_i)} - \frac{\widetilde{\mu}(Z_i)}{m_0(Z_i, \theta)\{1 - m_0(Z_i, \theta)\}} \right] \xi_i \{\delta_i - m_0(Z_i, \theta)\} + o_p(1)$$

$$:= T_{n45} + o_p(1).$$

By the central limit theorem, we have

$$T_{n45} \xrightarrow{\mathcal{L}} N(0, \Omega_I^*), \tag{A.19}$$

where

$$\Omega_I^* = E \left[L(X, \widetilde{Y}) L^\top(X, \widetilde{Y}) \pi(\widetilde{Y}) m_0(Z, \theta)\{1 - m_0(Z, \theta)\} \right] \tag{A.20}$$

and $L(X, \widetilde{Y}) = \frac{X\widetilde{Y}}{1-G(\widetilde{Y})} - \frac{\widetilde{\mu}(\widetilde{Y})}{m_0(Z,\theta)\{1-m_0(Z,\theta)\}}$. Similar to the way (A.5) was derived, we obtain

$$
\begin{aligned}
T_{n6} &= n^{-3/2} \sum_{i \neq j} \frac{X_i \widetilde{Y}_i \xi_i \{\delta_i - m_0(Z_i, \theta)\} \psi(\widetilde{Y}_j, \delta_j, \xi_j; \widetilde{Y}_i)}{1 - G(\widetilde{Y}_i)} + o_p(1) \\
&:= T_{n6,1} + o_p(1).
\end{aligned}
\tag{A.21}
$$

Note that

$$
\begin{aligned}
E\left[X_i \widetilde{Y}_i \xi_i \{\delta_i - m_0(Z_i, \theta)\} \middle| \widetilde{Y}_i \right] &= E\left[X_i \widetilde{Y}_i E(\xi_i | X_i, \widetilde{Y}_i, \delta_i) \{\delta_i - m_0(Z_i, \theta)\} \middle| \widetilde{Y}_i \right] \\
&= E\left[X_i \widetilde{Y}_i E(\xi_i | X_i, \widetilde{Y}_i) \{E(\delta_i | \widetilde{Y}_i, X_i) - E(\delta_i | \widetilde{Y}_i)\} \middle| \widetilde{Y}_i \right] = 0.
\end{aligned}
$$

From the fact that $E\{\psi(\widetilde{Y}_j, \delta_j, \xi_j; \widetilde{Y}_i) | \widetilde{Y}_i\} = 0$, it follows that

$$
E\|T_{n6,1}\|^2 \leq 2n^{-1} E\left\{ \|X_1\|^2 \widetilde{Y}_1^2 \psi^2(\widetilde{Y}_2, \delta_2, \xi_2; \widetilde{Y}_1) \right\}.
\tag{A.22}
$$

Equations (A.21) and (A.22) together prove that

$$
T_{n6} = o_p(1).
\tag{A.23}
$$

Under the assumed missingness mechanism, a little algebra shows $cov(T_{n1}, T_{n45}) = 0$, $cov(U_n, T_{n45}) = 0$ and $cov(M_n, T_{n45}) = 0$. This result, Eqs. (A.14), (A.16), (A.17), (A.19), and (A.23), and the fact that $\frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top \xrightarrow{p} \Sigma$ together prove Theorem 3.1.

Proof of Theorem 4.1 Similar to the derivation of (A.16), by (C.5), (C.6) and (C.7), we have

$$
\begin{aligned}
&n^{-1/2} \sum_{i=1}^{n} \left(\widehat{Y}_{i,W} - X_i^\top \beta \right) \\
&= n^{-1/2} X_i \left(\widehat{Y}_{i,R} - X_i^\top \beta \right) + n^{-1/2} \sum_{i=1}^{n} X_i \frac{\frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)} \{\delta_i - m_0(\widetilde{Y}_i, \theta)\} \widetilde{Y}_i}{1 - \widehat{G}_n(\widetilde{Y}_i)} \\
&\quad + n^{-1/2} \sum_{i=1}^{n} X_i \frac{\frac{\xi_i}{\widehat{\pi}_n(\widetilde{Y}_i)} \{m_0(Z_i, \theta) - m_0(Z_i, \widehat{\theta}_n)\} \widetilde{Y}_i}{1 - \widehat{G}_n(\widetilde{Y}_i)} \\
&= T_{n1} + M_n + U_n + S_n + R_n + o_p(1),
\end{aligned}
\tag{A.24}
$$

where T_{n1}, M_n and U_n are defined in (A.7), and S_n and R_n are given by

$$S_n = n^{-1/2} \sum_{i=1}^{n} X_i \frac{\frac{\xi_i}{\pi(\widetilde{Y}_i)}\{\delta_i - m_0(Z_i, \theta)\}\widetilde{Y}_i}{1 - G(\widetilde{Y}_i)}$$

$$R_n = n^{-1/2} \sum_{i=1}^{n} X_i \frac{\frac{\xi_i}{\pi(\widetilde{Y}_i)}\{m_0(Z_i, \theta) - m_0(Z_i, \widehat{\theta}_n)\}\widetilde{Y}_i}{1 - G(\widetilde{Y}_i)}.$$

Similar to the way (A.3) was derived, we can show that

$$R_n = n^{-1/2} \sum_{i=1}^{n} \frac{\xi_j\{\delta_j - m(X_j, \theta)\}}{m_0(Z_j, \theta)\{1 - m_0(Z_j, \theta)\}} \mu(\widetilde{Y}_j) + o_p(1).$$

Hence, it follows from the central limit theorem that

$$S_n + R_n = n^{-1/2} \sum_{i=1}^{n} \left[\frac{X_i \widetilde{Y}_i}{1 - G(\widetilde{Y}_i)} - \frac{\pi(\widetilde{Y}_i)\mu(\widetilde{Y}_i)}{m_0(Z_i, \theta)\{1 - m_0(Z_i, \theta)\}} \right]$$
$$\times \frac{\xi_i\{\delta_i - m_0(Z_i, \theta)\}}{\pi(\widetilde{Y}_i)} + o_p(1)$$
$$\xrightarrow{\mathcal{L}} N(0, \Omega_W^*), \tag{A.25}$$

where

$$\Omega_W^* = E\left[\widetilde{L}(X, \widetilde{Y})\widetilde{L}^\top(X, \widetilde{Y}) \frac{m_0(Z, \theta)\{1 - m_0(Z, \theta)\}}{\pi(\widetilde{Y})} \right] \tag{A.26}$$

and $\widetilde{L}(X, \widetilde{Y}) = \frac{X\widetilde{Y}}{1 - G(\widetilde{Y})} - \frac{\pi(\widetilde{Y})\mu(\widetilde{Y})}{m_0(Z, \theta)\{1 - m_0(Z, \theta)\}}$.

It can be shown that $cov(T_{n1} + M_n + U_n, S_n + R_n) = 0$. This result, together with Eqs. (A.7), (A.14), and (A.24), and (A.25), and the fact that $\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top \xrightarrow{P} \Sigma$, prove Theorem 4.1.

References

Albert A, Anderson JA (1984) On the existence of maximum likelihood estimates in logistic regression models. Biometrika 71:1–10

Buckley J, James I (1979) Linear regression with censored data. Biometrika 66:429–436

Cox DR (1972) Regression models and life tables (with discussion). J R Stat Soc B 34:187–220

Dewanji A (1992) A note on a test for competing risks with missing failure type. Biometrika 79: 855–857

Dikta G (1998) On semiparametric random censorship models. J Stat Plan Inference 66:253–279

Dinse GE (1982) Nonparametric estimation for partially-complete time and type of failure data. Biometrics 38:417–431

Gao G, Tsiatis AA (2005) Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. Biometrika 92:875–891

Goetghebeur EJ, Ryan L (1990) A modified logrank test for competing risks with missing failure type. Biometrika 77:207–211

Goetghebeur EJ, Ryan L (1995) Analysis of competing risks survival data when some failure types are missing. Biometrika 82:821–833

Hahn J (1998) On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 66:315–331

Hirano K, Imbens GW, Ridder G (2003) Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71:1161–1189

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663–685

Jin Z, Lin DY, Wei LJ, Ying Z (2003) Rank-based inference for the accelerated failure time model. Biometrika 90:341–353

Koul H, Susarla V, van Ryzin J (1981) Regression analysis with randomly right-censored data. Ann Stat 9:1276–1288

Lai TL, Ying Z (1991) Rank regression methods for left-truncated and right-censored data. Ann Stat 19:531–556

Leurgans S (1987) Linear models, random censoring and synthetic data. Biometrika 74:301–309

Li G, Wang QH (2003) Empirical likelihood regression analysis for right censored data. Stat Sinica 13:51–68

Lo S-H (1991) Estimating a survival function with incomplete cause-of-death data. J Multivar Anal 39:217–235

Lu W, Liang Y (2008) Analysis of competing risks data with missing cause of failure under additive hazards model. Stat Sinica 18:219–234

Lu K, Tsiatis AA (2001) Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. Biometrics 57:1191–1197

Lunceford JK, Davidian M (2004) Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med 23:2937–2960

McKeague IW, Subramanian S (1998) Product-limit estimators and Cox regression with missing censoring information. Scand J Stat 25:589–601

Miller RG (1976) Least squares regression with censored data. Biometrika 63:449–464

Peddada SD, Patwardhan G (1992) Jackknife variance estimators in linear models. Biometrika 79:654–657

Reid N (1994) A conversation with Sir David Cox. Stat Sci 9:439–455

Ritov Y (1990) Estimation in a linear regression model with censored data. Ann Stat 18:303–328

Robins JM, Rotnitzky A, Zhao LP (1994) Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 89:846–866

Santner TJ, Duffy DE (1986) A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. Biometrika 73:755–758

Scharfstein DO, Rotnitzky A, Robins JM (1999) Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). J Am Stat Assoc 94:1096–1146

Subramanian S (2004) Asymptotically efficient estimation of a survival function in the missing censoring indicator model. Nonparametr Stat 16:797–817

Subramanian S (2006) Survival analysis for the missing censoring indicator model using kernel density estimation techniques. Stat Methodol 3:125–136

Tsiatis AA (1990) Estimating regression parameters using linear rank tests for censored data. Ann Stat 18:354–372

Tsiatis AA, Davidian M, McNeney B (2002) Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. Biometrika 89:238–244

Wang QH, Ng K (2008) Asymptotically efficient product-limit estimators with censoring indicators missing at random. Stat Sinica 18:749–768

Wang QH, Linton O, Härdle W (2004) Semiparametric regression analysis with missing response at random. J Am Stat Assoc 99:334–345

Ying Z (1993) A large sample study of rank estimation for censored regression data. Ann Stat 21:76–99

Zhou X, Sun L (2003) Additive hazards regression with missing censoring information. Stat Sinica 13:1237–1257