



Published in final edited form as:

J Am Stat Assoc. 2009 June 1; 104(486): 803–815. doi:10.1198/jasa.2009.0130.

A Class of Transformed Mean Residual Life Models With Censored Survival Data

Liuquan Sun and

Professor, Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, P.R. China (E-mail: slq@amt.ac.cn)

Zhigang Zhang

Assistant Attending Biostatistician, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 (E-mail: zhangz@mskcc.org)

Abstract

The mean residual life function is an attractive alternative to the survival function or the hazard function of a survival time in practice. It provides the remaining life expectancy of a subject surviving up to time t . In this study, we propose a class of transformed mean residual life models for fitting survival data under right censoring. To estimate the model parameters, we make use of the inverse probability of censoring weighting approach and develop a system of estimating equations. Efficiency and robustness of the estimators are also studied. Both asymptotic and finite sample properties of the proposed estimators are established and the approach is applied to two real-life datasets collected from clinical trials.

Keywords

Censored survival data; Estimating equation approach; Inverse probability of censoring weighting; Mean residual life; Semiparametric efficiency; Transformed models

1. INTRODUCTION

For a nonnegative survival time T with finite expectation, the mean residual life function (MRLF) at time $t \geq 0$ is defined as

$$m(t) = E(T - t | T > t). \tag{1}$$

In order for the quantity $m(t)$ to make sense, at least practically, we shall insist that it is only defined when the condition $T > t$ has positive probability. Throughout this article, we assume T is absolutely continuous with a density $f(t)$ with respect to the Lebesgue measure. Denoting the corresponding survival and hazard functions as $F(t)$ and $\lambda(t)$, respectively, the following four relationships can be easily verified:

$$\begin{aligned} m(t) &= \frac{\int_t^\infty F(u) du}{F(t)}, \\ F(t) &= \frac{m(0)}{m(t)} \exp\left\{-\int_0^t \frac{du}{m(u)}\right\}, \\ f(t) &= F(t) \left[\frac{m'(t)+1}{m(t)}\right], \end{aligned}$$

and

$$\lambda(t) = \left[\frac{\dot{m}(t)+1}{m(t)} \right],$$

provided that all denominators are nonzero, where $\dot{m}(t)$ denotes the first derivative of $m(t)$.

An important property of MRLF is that $m(t) + t$ is nondecreasing, which is obvious because $m(t) + t = E(T|T > t)$. This is also reflected by the fact that $\dot{m}(\cdot) + 1 \geq 0$, as $f(t)$ and $\lambda(t)$ are always nonnegative. For other properties of MRLF that do not concern us here, we refer interested readers to Balkema and de Haan (1974), Hollander and Proschan (1975), Kotz and Shanbhag (1980), Arnold and Zahedi (1988), and the references within.

MRLF is of interest in many fields, such as reliability research, survival analysis and actuarial study, and so forth. For example, it is sometimes more informative to tell a prostate cancer patient how long he can survive or live without disease recurrence, in expectation, given his current situation (which of course includes the fact that he has “survived” or lived without the disease so far). As another example, a customer may be interested in knowing how much longer his or her computer can be used, given that the computer has worked normally for, say, t years. Later in this article, two real data examples are introduced and analyzed, which will further illustrate the usage of MRLF.

In the following, we will briefly review the literature on MRLF. Early studies on MRLF focused more on its probabilistic behaviors. See, for example, the book by Cox (1961) for a general introduction, and Swartz (1973) and Kotz and Shanbhag (1980) for some characterization and representation theorems. Balkema and de Haan (1974) in particular studied the limiting behavior of $m(t)$ when $t \rightarrow \infty$. More recent works include those of Hall and Wellner (1984) and Oakes and Dasu (1990). The concept of MRLF has also been generalized to multivariate distributions. See Arnold and Zahedi (1988) and a recent work by Bassan, Kochar, and Spizzichino (2002).

For statistical inference, many testing procedures have been established. Hollander and Proschan (1975) developed tests for testing exponentiality (under which the MRLF is constant) versus decreasing MRLF alternatives and exponentiality versus “new better than used” in expectation alternatives. In their work, the authors also discussed interesting distribution classes. Recently, El-Bassiouny and Alwasel (2003) proposed a more efficient test for the first scenario considered by Hollander and Proschan (1975). Dauxois (2003), on the other hand, considered testing exponentiality versus nonexponentiality when right-censored data are present.

Estimation of MRLF in homogeneous cases has not been addressed as extensively as the probabilistic properties or testing procedures. Of course, when no censoring is present, this is trivial, because one can easily compute the empirical MRLF. Actually, Bickel et al. (1993, p. 197) showed that it is the optimal (i.e., most efficient) estimator using semiparametric efficiency theory. For bivariate MRLF under order restriction, Rojo and Ghebremichael (2006) developed a projection estimator, again, when no censoring is present. But the literature on censored data seems limited. This issue is discussed briefly in Section 7.

When data are not homogeneous, the estimation problem becomes even more challenging. To our best knowledge, Oakes and Dasu (1990) proposed the first two-sample proportional mean residual life model

$$m_1(t) = \psi m_0(t), \quad (2)$$

where $\psi > 0$ represents the ratio of the two MRLFs. They gave a characterization theorem but did not provide a method for estimating ψ . Later, in Maguluri and Zhang (1994), model (2) was extended to incorporate covariates,

$$m(t|\mathbf{Z}) = m_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}), \quad (3)$$

where $m(t|\mathbf{Z})$ is the MRLF of the survival time T corresponding to a $p \times 1$ vector covariate \mathbf{Z} , $m_0(t)$ is some unknown baseline MRLF, and $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of regression parameters. They also proposed two estimators for $\boldsymbol{\beta}$ in model (3) with complete data (i.e., no censoring). Their approach was later modified to accommodate potential right censoring in Chen et al. (2005). Under the same situation and for the same purpose, Chen and Cheng (2005) developed another methodology by using generalized estimating equations. The latter approach does not rely on modeling the censoring mechanism, but requires estimating the baseline MRLF.

A new class of additive mean residual life model,

$$m(t|\mathbf{Z}) = m_0(t) + \boldsymbol{\beta}'\mathbf{Z}, \quad (4)$$

in which each component has a similar interpretation as their counterparts in model (3), was also proposed by Chen and Cheng (2006) and Chen (2007). In these articles, the authors discussed various estimation methodologies with or without right censoring.

In this article, we propose a more general family of regression models for the MRLF that take the form

$$m(t|\mathbf{Z}) = g\{m_0(t) + \boldsymbol{\beta}'\mathbf{Z}\}, \quad (5)$$

where $g: \mathcal{R} \mapsto \mathcal{R}^+$ is a prespecified “transformation” or “link” function, $m_0(t)$ is some unknown function (note that unless g is the identity link, $m_0(t)$ is not necessarily the baseline MRLF, whereas $g\{m_0(t)\}$ is), and $\boldsymbol{\beta}$ and \mathbf{Z} are defined similarly as before. Some important features of this model are outlined next.

First, we see that (5) defines a very rich family of models through the link function g , which allows the models to be useful in applications such as treatment assessment, resource planning, or short-term forecasting. In particular, it encompasses the Box-Cox transformation, in which g is given by $g(x) = [(x + 1)^\rho - 1]/\rho$, where $\rho = 0$ means that $g(x) = \log(x + 1)$. When $g(x) = \exp(x)$, (5) reduces to (3), and if $g(x) = x$ then (5) becomes (4). This means the proportional and additive mean residual life models are special cases of our model.

Second, we should be careful about the choice of an appropriate link function g . Practically, the choice may be based on prior data or the desiring interpretation of the regression parameters. Notice that this choice will also affect the interpretation of the function $m_0(t)$. Theoretically, however, there are some restrictions for choosing an appropriate g . In the following, we list the requirements: (A) g is twice continuously differentiable, (B) g is strictly increasing, and

(C) $g\{m_0(t) + \beta'Z\}$ is a proper MRLF for all possible values of Z (later we shall also assume that $t \in [0, \tau]$ for some finite τ and Z is bounded). Condition (A) is a result of the fact that we are dealing with absolutely continuous survival times. Condition (B) is required for technical arguments in later sections. Roughly speaking, it guarantees the uniqueness of the estimate of β_0 , true value of β . Condition (C) is obvious, because $g\{m_0(t) + \beta'Z\}$ is itself the MRLF of the survival time T given Z . More comments on this are given in Section 7.

Third, as pointed out by a referee, model (5) may prove useful in model selection. Notice that the Box-Cox transformation includes the additive model by letting $\rho = 1$. Thus, if we insist that g lies in the family of Box-Cox transformation but allow ρ to be estimated, this would provide a way in model selection between the additive mean residual life model and other models contained by the Box-Cox family. We will give a brief discussion in Section 7 on this research topic.

One important motivation behind model (5), in application, is certainly its flexibility in modeling covariate effects. We now look at a real data example. A recent study conducted at the Memorial Sloan-Kettering Cancer Center aims at investigating the potential effect of the radiation therapy dose level on the relapse time of the prostate cancer-specific antigen (PSA). Here, the relapse time, which is our survival time of interest, is defined from the end of the radiation therapy treatment to the time when the PSA level of the patient has been observed to cross a certain threshold value. Of course, this relapse time is sometimes censored by random withdrawal or other reasons. A recent study (Zeilefsky et al. 2008) has shown that the age of the patient may also play an important role in predicting the PSA relapse time. Therefore, we need to incorporate age as another covariate in the mean residual life regression model. However, it may not be reasonable to postulate that the effect of age on the MRLF of PSA relapse time is additive or multiplicative, because it would imply that a 5-year increase of age from 60 to 65 would have the same effect as a 5-year increase of age, say, from 80 to 85 (either additively or multiplicatively). As a matter of fact, it is more reasonable clinically to assume that the effect of age phases out gradually as it increases. In other words, we may take g to be the Box-Cox transformation with ρ as a fraction number like 0.5. More detailed discussion and data analysis results are reported in Section 6.

The rest of this article is organized as follows. In the next section, we will first discuss the situation in which the censoring time is independent of T and Z , and a general inference procedure based on estimating functions is proposed. The procedure can be easily implemented numerically and the asymptotic properties of the proposed estimators of regression parameters are established. Section 3 extends the method to the situation in which the censoring time may depend on Z through the proportional hazards model. Issues regarding efficiency and double robustness are discussed in Section 4. Section 5 reports some results from simulation studies conducted for evaluating the proposed methods. In Section 6, we apply the methodology to two real-life datasets arising from clinical trials, and some concluding remarks are given in Section 7. Regularity conditions and theoretical proofs are collected in the Appendix and a supplemental technical report, respectively.

2. INFERENCE UNDER INDEPENDENT CENSORING

Let C be the potential censoring time, and assume that the support of C is longer than that of the survival time T to ensure that the MRLF is estimable. In the technical arguments, this assumption also saves us from lengthy technical discussion of the tail behavior of the limiting distributions. In this section, we consider the situation when C is independent of both T and Z . Let $\{T_i, C_i, Z_i; i = 1, \dots, n\}$ be independent replicates of $\{T, C, Z\}$ and suppose that we observe $\{X_i, \Delta_i, Z_i; i = 1, \dots, n\}$, where $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. Here, $I(\cdot)$ is the indicator function.

Let $G(t)$ be the survival function of C and $\hat{G}(t)$ the Kaplan-Meier estimator of $G(t)$ based on $\{X_i, 1 - \Delta_i; i = 1, \dots, n\}$. Define

$$M_i(t) = \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)} [(X_i - t) - g\{m_0(t) + \beta'_0 \mathbf{Z}_i\}], \quad i=1, \dots, n.$$

Under model (5), $M_i(t)$ are zero-mean stochastic processes. Thus, for a given β , a reasonable estimator for $m_0(t)$ is the solution to

$$\sum_{i=1}^n \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)} [(X_i - t) - g\{m_0(t) + \beta' \mathbf{Z}_i\}] = 0, \quad 0 \leq t \leq \tau, \quad (6)$$

where $0 < \tau = \inf \{t: Pr(T \geq t) = 0\} < \infty$. Note that, without censoring, (6) reduces to

$\sum_{i=1}^n I(T_i > t) [(T_i - t) - g\{m_0(t) + \beta' \mathbf{Z}_i\}] = 0$, which is similar to the efficient nonparametric estimator of the MRLF discussed in Bickel et al. (1993). The inverse probability of censoring weighting (IPCW) technique has been extensively discussed in the literature. See, for example, Robins and Rotnitzky (1992); Robins, Rotnitzky, and Zhao (1994); and van der Laan and Robins (2003).

Denote this estimator by $\hat{m}_{a0}(t; \beta)$. To estimate β_0 , using the generalized estimating equation approach (Liang and Zeger, 1986) and again the IPCW technique, we propose the following class of estimating equations for β_0 :

$$\mathbf{U}_a(\beta) = \sum_{i=1}^n \int_0^\tau \frac{\Delta_i I(X_i > t) \mathbf{Z}_i}{\hat{G}(X_i)} [(X_i - t) - g\{\hat{m}_{a0}(t; \beta) + \beta' \mathbf{Z}_i\}] dH(t) = \mathbf{0}, \quad (7)$$

where $H(t)$ is an increasing weight function on $[0, \tau]$. Let $\hat{\beta}_a$ denote the solution to $\mathbf{U}_a(\beta) = \mathbf{0}$ and $\hat{m}_{a0}(t) := \hat{m}_{a0}(t; \hat{\beta}_a)$ the corresponding estimator of $m_0(t)$. Define

$$N_i^c(t) = I(X_i \leq t, \Delta_i = 0), \quad \hat{\pi}(t) = n^{-1} \sum_{i=1}^n I(X_i \geq t), \quad \hat{\Lambda}^c(t) = n^{-1} \sum_{i=1}^n \int_0^t dN_i^c(u) / \hat{\pi}(u),$$

$$\begin{aligned} \hat{M}_i^c(t) &= N_i^c(t) - \int_0^t I(X_i \geq u) d\hat{\Lambda}^c(u), \\ \hat{M}_i(t) &= \frac{\Delta_i I(X_i > t)}{\hat{G}(X_i)} [(X_i - t) - g\{\hat{m}_{a0}(t) + \hat{\beta}_a' \mathbf{Z}_i\}], \\ \hat{\mathbf{Q}}_a(t) &= n^{-1} \sum_{i=1}^n I(X_i \geq t) \int_0^\tau \hat{M}_i(u) \{\mathbf{Z}_i - \bar{\mathbf{Z}}_a(u; \hat{\beta}_a)\} dH(u) \end{aligned}$$

and

$$\bar{\mathbf{Z}}_a(t; \beta) = \frac{\sum_{i=1}^n \Delta_i \hat{G}(X_i)^{-1} I(X_i > t) g\{\hat{m}_{a0}(t; \beta) + \beta' \mathbf{Z}_i\} \mathbf{Z}_i}{\sum_{i=1}^n \Delta_i \hat{G}(X_i)^{-1} I(X_i > t) g\{\hat{m}_{a0}(t; \beta) + \beta' \mathbf{Z}_i\}},$$

where $\dot{g}(x) = dg(x)/dx$. We summarize the asymptotic properties of $\hat{\beta}_a$ and $\hat{m}_{a0}(t)$ in the following theorem.

Theorem 1

Under the regularity conditions (C1) through (C4) stated in the Appendix, we have the following:

- i. $\hat{\beta}_a$ and $\hat{m}_{a0}(t)$ always exist, and are unique and consistent.
- ii. $n^{1/2}(\hat{\beta}_a - \beta_0)$ is asymptotically normal with mean zero and a variance–covariance matrix that can be consistently estimated by $\hat{\mathbf{A}}^{-1}\Sigma_a\hat{\mathbf{A}}^{-1}$, where $\Sigma_a = n^{-1} \sum_{i=1}^n \widehat{\xi}_i^{\otimes 2}$,

$$\begin{aligned} \widehat{\xi}_i &= \int_0^\tau \widehat{M}_i(t) \{ \mathbf{Z}_i - \bar{\mathbf{Z}}_a(t; \widehat{\beta}_a) \} dH(t) + \int_0^\tau \frac{\widehat{Q}_a(t)}{\widehat{\pi}(t)} d\widehat{M}_i^c(t), \\ \widehat{\mathbf{A}} &= n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\Delta_i I(X_i \geq t)}{\widehat{G}(X_i)} \{ \mathbf{Z}_i - \bar{\mathbf{Z}}_a(t; \widehat{\beta}_a) \}^{\otimes 2} \dot{g} \{ \widehat{m}_{a0}(t) \\ &\quad + \widehat{\beta}_a' \mathbf{Z}_i \} dH(t) \end{aligned}$$

and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}'$ for a column vector \mathbf{v} .

- iii. $n^{1/2}\{\widehat{m}_{a0}(t) - m_0(t)\}$ ($0 \leq t \leq \tau$) converges weakly to a zero-mean Gaussian process with a covariance function that can be consistently estimated by

$$\widehat{\Gamma}_a(s, t) = n^{-1} \sum_{i=1}^n \widehat{\phi}_i(s) \widehat{\phi}_i(t) \text{ at } (s, t), \text{ where}$$

$$\begin{aligned} \widehat{\phi}_i(t) &= \frac{1}{\widehat{\Phi}_a(t)} \left[\widehat{M}_i(t) + \int_0^\tau \frac{\widehat{R}_a(t, u)}{\widehat{\pi}(u)} d\widehat{M}_i^c(u) \right] - \bar{\mathbf{Z}}_a(t; \widehat{\beta}_a)' \widehat{\mathbf{A}}^{-1} \widehat{\xi}_i, \\ \widehat{\Phi}_a(t) &= n^{-1} \sum_{i=1}^n \frac{\Delta_i I(X_i \geq t)}{\widehat{G}(X_i)} \dot{g} \{ \widehat{m}_{a0}(t) + \widehat{\beta}_a' \mathbf{Z}_i \}, \end{aligned}$$

$$\text{and } \widehat{R}_a(t, u) = n^{-1} \sum_{i=1}^n \widehat{M}_i(t) I(X_i \geq u).$$

The asymptotic normality for $\widehat{m}_{a0}(t)$, together with the consistent variance estimator $\widehat{\Gamma}_a$, enables us to construct pointwise confidence intervals for $m_0(t)$. To construct simultaneous confidence bands for $m_0(t)$ over a time interval of interest $[t_1, t_2]$, $0 < t_1 < t_2 \leq \tau$, we need to evaluate the distribution of the supremum of a related process over $[t_1, t_2]$. It is not possible to evaluate such distributions analytically because the limiting process of $n^{1/2}\{\widehat{m}_{a0}(t) - m_0(t)\}$ does not have an independent increments structure. To handle this problem, we can use a resampling scheme

to approximate the distribution of $n^{1/2}\{\widehat{m}_{a0}(t) - m_0(t)\}$. Define $\widehat{W}_a(t) = n^{-1/2} \sum_{i=1}^n \widehat{\phi}_i(t) \Omega_i$, where $(\Omega_1, \dots, \Omega_n)$ are independent, standard normal variables that are independent of the data $\{X_i, \Delta_i, \mathbf{Z}_i; i = 1, \dots, n\}$. According to the arguments of Lin et al. (2000), the distribution of the process $n^{1/2}\{\widehat{m}_{a0}(t) - m_0(t)\}$ can be approximated by that of the zero-mean Gaussian process $\widehat{W}_a(t)$. To approximate the distributions of $n^{1/2}\{\widehat{m}_{a0}(t) - m_0(t)\}$, we need to obtain a large number of realizations from $\widehat{W}_a(t)$ by repeatedly generating the normal random sample $(\Omega_1, \dots, \Omega_n)$ while fixing the data $\{X_i, \Delta_i, \mathbf{Z}_i; i = 1, \dots, n\}$ at their observed values. Using this simulation method, we may determine an approximate $1 - \alpha$ simultaneous confidence band for $m_0(t)$ over the time interval of interest $[t_1, t_2]$.

3. INFERENCE UNDER DEPENDENT CENSORING

We now consider the situation in which the censoring time C may depend on the covariate \mathbf{Z} . However, we assume that given \mathbf{Z} , T is independent of C . This actually corresponds to the scenario called “coarsening at random” in the literature of missing data. It has been pointed out (Robins and Ritov 1997) that for such problems it is wise to postulate a model for the censoring mechanism. Of course, if \mathbf{Z} takes only finitely many values and the sample size n is

large enough, we may be able to calculate the Kaplan-Meier estimates of the censoring distribution for each value of \mathbf{Z} . However, this is not feasible when \mathbf{Z} (or at least one component of \mathbf{Z}) is continuous or the sample size is moderate. To this end, we assume that the hazard function of C given \mathbf{Z} has the form

$$\lambda_c(t|\mathbf{Z})=\lambda_0(t)\exp\{\boldsymbol{\gamma}'\mathbf{Z}\}, \quad (8)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\gamma}$ is a vector of unknown regression parameters. Similar as before, we denote the true value of $\boldsymbol{\gamma}$ as $\boldsymbol{\gamma}_0$. We stress here that this model may not always adequately capture the censoring mechanism. As a matter of fact, the purpose of this modeling is merely to provide a platform for inference about $\boldsymbol{\beta}_0$ and $m_0(t)$. Later, we shall consider efficiency and double robustness of our approach (Section 4), in which we will derive more efficient and robust estimators, and show that even if this Cox model is incorrect, we may still be able to obtain consistent and asymptotically normal estimators of $\boldsymbol{\beta}_0$ and $m_0(t)$.

Note that the model from Section 2 is the model here with $\boldsymbol{\gamma}_0 = \mathbf{0}$. Of course, $\boldsymbol{\gamma}_0$, as well as $\lambda_0(t)$, needs to be estimated. A natural estimator of $\boldsymbol{\gamma}_0$ under model (8) is given by the convenient maximum partial likelihood estimator defined as the solution to

$$\mathbf{U}_r(\boldsymbol{\gamma})=\sum_{i=1}^n \int_0^{\tau}\{\mathbf{Z}_i-\bar{\mathbf{Z}}_r(t;\boldsymbol{\gamma})\}dN_i^c(t)=\mathbf{0}, \quad (9)$$

where $N_i^c(t)=I(X_i \leq t, \Delta_i=0)$, $\mathbf{Z}_r(t;\boldsymbol{\gamma})=\mathbf{S}^{(1)}(t;\boldsymbol{\gamma})/S^{(0)}(t;\boldsymbol{\gamma})$, and

$\mathbf{S}^{(k)}(t;\boldsymbol{\gamma})=\sum_{i=1}^n I(X_i \geq t)\mathbf{Z}_i^{\otimes k} \exp\{\boldsymbol{\gamma}'\mathbf{Z}_i\}$ for $k=0, 1$. Let $\hat{\boldsymbol{\gamma}}$ denote the estimator given by $\mathbf{U}_r(\boldsymbol{\gamma}) = \mathbf{0}$, and $\hat{\Lambda}_0(t)$ be the Breslow estimator of $\Lambda_0(t)=\int_0^t \lambda_0(u)du$ —in other words,

$$\hat{\Lambda}_0(t)=\sum_{i=1}^n \int_0^t \frac{dN_i^c(u)}{\sum_{j=1}^n I(X_j \geq u)\exp\{\hat{\boldsymbol{\gamma}}'\mathbf{Z}_j\}}.$$

Also, let $G_i(t;\boldsymbol{\gamma}_0, \Lambda_0)$ be the censoring survival distribution of C_i given \mathbf{Z}_i under model (8)—that is, $G_i(t;\boldsymbol{\gamma}_0, \Lambda_0)=\exp\{-\Lambda_0(t)\exp(\boldsymbol{\gamma}_0'\mathbf{Z}_i)\}$. For inference, motivated by the method given in the previous section, we consider the following equation for estimating $m_0(t)$:

$$\sum_{i=1}^n \frac{\Delta_i I(X_i > t)}{G_i(X_i; \hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0)} [(X_i - t) - g\{m_0(t) + \boldsymbol{\beta}'\mathbf{Z}_i\}] = 0, \quad 0 \leq t \leq \tau. \quad (10)$$

Denote this estimator by $\hat{m}_{b0}(t; \boldsymbol{\beta})$ for a given $\boldsymbol{\beta}$. Again, using the generalized estimating equation method, we propose the following class of IPCW estimating equations for $\boldsymbol{\beta}_0$ when the censoring time C_i may depend on \mathbf{Z}_i via model (8):

$$\begin{aligned} & \mathbf{U}_b(\boldsymbol{\beta}) \\ &= \sum_{i=1}^n \int_0^\tau \frac{\Delta_i I(X_i > t) \mathbf{Z}_i}{G_i(X_i; \boldsymbol{\gamma}, \Lambda_0)} [(X_i - t) - g\{\widehat{m}_{b0}(t; \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{Z}_i\}] \\ & \quad \times dH(t) = \mathbf{0}. \end{aligned} \tag{11}$$

Let $\widehat{\boldsymbol{\beta}}_b$ denote the solution to $\mathbf{U}_b(\boldsymbol{\beta}) = \mathbf{0}$ and $\widehat{m}_{b0}(t) := \widehat{m}_{b0}(t; \widehat{\boldsymbol{\beta}}_b)$. Define

$$\begin{aligned} \widetilde{M}_i^c(t) &= N_i^c(t) - \int_0^t I(X_i \geq u) \exp\{\boldsymbol{\gamma}' \mathbf{Z}_i\} d\widehat{\Lambda}_0(u), \\ \widetilde{M}_i^*(t) &= \frac{\Delta_i I(X_i > t)}{G_i(X_i; \boldsymbol{\gamma}, \Lambda_0)} [(X_i - t) - g\{\widehat{m}_{b0}(t) + \widehat{\boldsymbol{\beta}}_b' \mathbf{Z}_i\}], \\ \widehat{\mathbf{Q}}_b(t) &= n^{-1} \sum_{i=1}^n I(X_i \geq t) \exp\{\boldsymbol{\gamma}' \mathbf{Z}_i\} \\ & \quad \times \int_0^t \widetilde{M}_i^*(u) \{\mathbf{Z}_i - \overline{\mathbf{Z}}_b(u; \widehat{\boldsymbol{\beta}}_b)\} dH(u), \\ \widehat{\mathbf{P}} &= n^{-1} \sum_{i=1}^n \widehat{\Lambda}_0(X_i) \exp\{\boldsymbol{\gamma}' \mathbf{Z}_i\} \\ & \quad \times \int_0^\tau \widetilde{M}_i^*(u) \{\mathbf{Z}_i - \overline{\mathbf{Z}}_b(u; \widehat{\boldsymbol{\beta}}_b)\} dH(u) \mathbf{Z}_i', \\ \overline{\mathbf{Z}}_b(t; \boldsymbol{\beta}) &= \sum_{i=1}^n \frac{\Delta_i G_i(X_i; \boldsymbol{\gamma}, \Lambda_0)^{-1} I(X_i > t) g\{\widehat{m}_{b0}(t; \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{Z}_i\} \mathbf{Z}_i}{\sum_{i=1}^n \Delta_i G_i(X_i; \boldsymbol{\gamma}, \Lambda_0)^{-1} I(X_i > t) g\{\widehat{m}_{b0}(t; \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{Z}_i\}} \end{aligned}$$

We summarize the asymptotic properties of $\widehat{\boldsymbol{\beta}}_b$ and $\widehat{m}_{b0}(t)$ in the following theorem.

Theorem 2

Under the regularity conditions (C1) through (C3) and (C5) stated in the Appendix, we have

- i. $\widehat{\boldsymbol{\beta}}_b$ and $\widehat{m}_{b0}(t)$ always exist, and are unique and consistent.
- ii. $n^{1/2}(\widehat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)$ is asymptotically normal with mean zero and a variance–covariance matrix that can be consistently estimated by $\widehat{\mathbf{B}}^{-1} \widehat{\boldsymbol{\Sigma}}_b \widehat{\mathbf{B}}^{-1}$, where $\widehat{\boldsymbol{\Sigma}}_b = n^{-1} \sum_{i=1}^n \widehat{\boldsymbol{\eta}}_i^{\otimes 2}$,

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_i &= \int_0^\tau \widetilde{M}_i^*(t) \{\mathbf{Z}_i - \overline{\mathbf{Z}}_b(t; \widehat{\boldsymbol{\beta}}_b)\} dH(t) + \int_0^\tau \frac{\widehat{\mathbf{Q}}_b(t)}{S^{(0)}(t; \boldsymbol{\gamma})} d\widetilde{M}_i^c(t) \\ & \quad + \left[\widehat{\mathbf{P}} - \int_0^\tau \widehat{\mathbf{Q}}_b(t) \overline{\mathbf{Z}}_b(t; \boldsymbol{\gamma})' d\widehat{\Lambda}_0(t) \right] \widehat{\mathbf{D}}^{-1} \\ & \quad \times \int_0^\tau \{\mathbf{Z}_i - \overline{\mathbf{Z}}_b(t; \boldsymbol{\gamma})\} d\widetilde{M}_i^c(t), \\ \widehat{\mathbf{B}} &= n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\Delta_i I(X_i > t)}{G_i(X_i; \boldsymbol{\gamma}, \Lambda_0)} \{\mathbf{Z}_i - \overline{\mathbf{Z}}_b(t; \widehat{\boldsymbol{\beta}}_b)\}^{\otimes 2} \dot{g}\{\widehat{m}_{b0}(t) \\ & \quad + \widehat{\boldsymbol{\beta}}_b' \mathbf{Z}_i\} dH(t), \end{aligned}$$

and $\widehat{\mathbf{D}} = -n^{-1} \partial \mathbf{U}_r(\widehat{\boldsymbol{\gamma}}) / \partial \boldsymbol{\gamma}'$.

- iii. $n^{1/2}\{\widehat{m}_{b0}(t) - m_0(t)\}$ ($0 \leq t \leq \tau$) converges weakly to a zero-mean Gaussian process with a covariance function at (s, t) that can be consistently estimated by

$$\widehat{\Gamma}_b(s, t) = n^{-1} \sum_{i=1}^n \widehat{\psi}_i(s) \widehat{\psi}_i(t), \text{ where}$$

$$\begin{aligned}
 \widehat{\psi}_i(t) &= \frac{1}{\widehat{\Phi}_b(t)} \left[\widehat{M}_i^*(t) + \int_0^t \frac{\widehat{R}_b(t,u)}{S^{(0)}(u;\widehat{\gamma})} d\widetilde{M}_i^c(u) \right] \\
 -\widehat{\mathbf{Z}}_b(t;\widehat{\beta}_b)' \widehat{\mathbf{B}}^{-1} \widehat{\eta}_i + \widehat{\psi}(t) D^{-1} \int_0^t \{\mathbf{Z}_i - \widehat{\mathbf{Z}}_r(u;\widehat{\gamma})\} d\widetilde{M}_i^c(u), \\
 \widehat{\Phi}_b(t) &= n^{-1} \sum_{i=1}^n \frac{\Delta I(X_i \geq t)}{G_i(X_i; \widehat{\gamma}, \Lambda_0)} \dot{g} \{ \widehat{m}_{b0}(t) + \widehat{\beta}_b' \mathbf{Z}_i \}, \\
 \widehat{\psi}(t) &= n^{-1} \sum_{i=1}^n \widehat{M}_i^*(t) \widehat{\Lambda}_0(X_i) \exp\{\widehat{\gamma}' \mathbf{Z}_i\} \\
 &\quad \times \mathbf{Z}_i' - \int_0^t \widehat{R}_b(t,u) \widehat{\mathbf{Z}}_r(u;\widehat{\gamma})' d\widehat{\Lambda}_0(u), \\
 \widehat{R}_b(t,u) &= n^{-1} \sum_{i=1}^n \widehat{M}_i^*(t) \exp\{\widehat{\gamma}' \mathbf{Z}_i\} I(X_i \geq u).
 \end{aligned}$$

To construct a simultaneous confidence band for $m_0(t)$ over $[t_1, t_2]$, we may again use the resampling method mentioned in Section 2. Namely, we approximate the distribution of the process $n^{1/2}\{\widehat{m}_{b0}(t) - m_0(t)\}$ by that of the zero-mean Gaussian process $\widehat{W}_b(t)$, where

$$\widehat{W}_b(t) = n^{1/2} \sum_{i=1}^n \widehat{\psi}_i(t) \Omega_i \text{ and } (\Omega_1, \dots, \Omega_n) \text{ are independent, standard normal variables that are independent of the observed data.}$$

4. EFFICIENCY AND DOUBLE ROBUSTNESS CONSIDERATIONS

Because estimating functions (7) and (11) are given in a somewhat ad hoc fashion using the IPCW method, a question naturally arises regarding how efficient and robust the resulting inference procedures are. Here we provide insights into these problems by using the semiparametric efficiency theory developed by Bickel et al. (1993); Robins, Rotnitzky, and Zhao (1994); and van der Laan and Robins (2003). For simplicity, we will only study the efficiency and double robustness in the case of independent censoring. A discussion on the case of dependent censoring, although similar to those given later conceptually, is much messier algebraically and is thus omitted.

4.1 Efficiency Considerations

The fundamental idea of semiparametric efficiency is conveyed through the so-called “tangent spaces” and orthogonal projections in Hilbert spaces. Normally, one makes enough regularity assumptions in order for the problem to be meaningful practically and solvable theoretically (Newey, 1990). For instance, it is usually insisted that we deal only with regular and asymptotically linear estimators. This excludes superefficient estimators and allows us to identify influence functions. Notice that this is exactly the situation we have here for $\widehat{\beta}_a$ (see Theorem 1 and its proof). Under these assumptions, the tangent space is taken to be the closed linear span of the scores of the parametric submodels of the semiparametric model of interest. One may then orthogonally project an arbitrary influence function onto this tangent space and obtain the efficient influence function. Alternatively, if the parametric part and the nonparametric part of the semiparametric model can be explicitly separated, it might be possible to express the tangent space by the direct sum of two subspaces: one corresponding to the parameter of interest and the other corresponding to the nuisance parameter. The latter is sometimes referred to as the “nuisance tangent space.” One may then take the orthogonal projection of the score of the parametric part onto the nuisance tangent space and obtain the efficient score function.

These approaches, although theoretically elegant, may not be computationally feasible. The difficulty usually lies in two aspects: (1) finding the tangent space (or the nuisance tangent space) and (2) calculating the orthogonal projection. The first difficulty is mostly the result of the complicated structure of the semiparametric model, and thus the scores. The second difficulty may be the result of several reasons. For example, the orthogonal projection might be a conditional expectation that requires extra modeling and awkward numerical integration.

When data are coarsened or censored and the coarsening or censoring mechanism needs to be modeled, these difficulties multiply. For our problem, unfortunately, both difficulties exist. Therefore, we take the following approach. First, because β_a has already been shown to be regular and asymptotically linear, we will take advantage of its estimating function $U_a(\beta)$ and the influence function. We will rewrite it using a martingale integral expression for later algebraic convenience. Next, we will add an extra term to this estimating function and, hence, improve the efficiency. The intuitive idea behind this construction is that the estimator $\hat{\beta}_a$ is based on an estimating function that only uses complete cases. Thus, it is reasonable to expect that we can extract more information from the “incomplete” (i.e., censored) cases. We will show that this is indeed the case by proving that the augmented IPCW (AIPCW) estimator is always more efficient than the original, simple IPCW estimator. Additional issues like information bound and double robustness will also be discussed at the end of this section.

First, notice that we have already derived the influence function for $\hat{\beta}_a$ in Theorem 1 and its proof. Namely,

$$\begin{aligned} n^{1/2}(\widehat{\beta}_a - \beta_0) &= \mathbf{A}^{-1}n^{-1/2}U_a(\beta_0) + o_p(1) \\ &= \mathbf{A}^{-1}n^{-1/2} \sum_{i=1}^n \xi_i + o_p(1). \end{aligned}$$

In other words, the influence function is $\mathbf{A}^{-1}\xi_i$. For a definition of \mathbf{A} , see (C4) in the Appendix. In the proof of Theorem 1, we also show that \mathbf{A} is actually the limit of $-n^{-1}\partial U_a(\beta_0)/\partial \beta'$. For the ease of later computations, we now give ξ_i (and hence the influence function) another expression. Define

$$\begin{aligned} \mathbf{D}_i &= \int_0^\tau I(T_i > t) [(T_i - t) - g\{m_0(t) + \beta_0' \mathbf{Z}_i\}] \{\mathbf{Z}_i - \bar{\mathbf{z}}_a(t)\} \\ &\quad \times d\tilde{H}(t), \quad i=1, \dots, n, \end{aligned}$$

and \mathbf{D} accordingly, where $\bar{\mathbf{z}}_a(t)$ and $\tilde{H}(t)$ are the limits of $\mathbf{Z}_a(t; \beta_0)$ and $H(t)$. Let $\Lambda^c(t) = -\log(G(t))$ be the cumulative hazard function of the censoring time, and

$$M_i^c(t) = N_i^c(t) - \int_0^t I(X_i \geq u) d\Lambda^c(u), \quad i=1, \dots, n.$$

Then $M_i^c(t)$ are martingales with respect to the σ -filtration

$$\sigma \{I(X_i \geq u), I(X_i \leq u, \Delta_i=0), 0 \leq u \leq t; I(T_i \leq v), 0 \leq v \leq \tau; \mathbf{Z}_i, i=1, \dots, n\}.$$

Similar settings can be found in Zhao and Tsiatis (1999) and Bang and Tsiatis (2000, 2002). Consider the martingale representation of the Kaplan-Meier estimator (Fleming and Harrington, 1991, p. 97)

$$\frac{\widehat{G}(t) - G(t)}{G(t)} = - \int_0^t \frac{\widehat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{n\widehat{\pi}(u)} \tag{12}$$

and the result from Robins and Rotnitzky (1992) that

$$\frac{\Delta_i}{G(X_i)} = 1 - \int_0^\tau \frac{dM_i^c(t)}{G(t)}. \quad (13)$$

It then follows from (12), (13), and some algebraic manipulation that

$$\begin{aligned} \mathbf{U}_a(\beta_0) &= \sum_{i=1}^n \frac{\Delta_i \mathbf{D}_i}{G(X_i)} - \sum_{i=1}^n \Delta_i \mathbf{D}_i \frac{\widehat{G}(X_i) - G(X_i)}{G(X_i) \widehat{G}(X_i)} + o_p(n^{1/2}) \\ &= \sum_{i=1}^n \left[\mathbf{D}_i - \int_0^\tau \{\mathbf{D}_i - \mathbf{K}(t; \mathbf{D})\} \frac{dM_i^c(t)}{G(t)} \right] + o_p(n^{1/2}), \end{aligned}$$

where $K(t; W) = E\{WI(T \geq t)\}/F(t)$ for any random variable or functional W , and $F(t) = P\{T \geq t\}$ is the (marginal) survival function of T . This means the i th influence function of the regular and asymptotically linear estimator $\hat{\beta}_a$ is

$$\mathbf{A}^{-1} \left[\mathbf{D}_i - \int_0^\tau \{\mathbf{D}_i - \mathbf{K}(t; \mathbf{D})\} \frac{dM_i^c(t)}{G(t)} \right]. \quad (14)$$

The alternative expression (14) for $\mathbf{U}_a(\beta_0)$ (inside the bracket) has a nice interpretation. The first term, \mathbf{D}_i , tells us the variability of our estimating equation if there were no censoring. The second term represents additional variability when the censoring is present and its mechanism needs to be estimated. Note that the covariance of these two terms is zero because \mathbf{D}_i and $M_i^c(t)$ are uncorrelated.

We now seek an AIPCW estimator that is more efficient than $\hat{\beta}_a$. Put another way, it should have a smaller asymptotic variance than (14). Here, by convention, we say that a variance matrix \mathbf{A} is smaller than a variance matrix \mathbf{B} if $\mathbf{B} - \mathbf{A}$ is nonnegative definite. According to the semiparametric efficiency theory for missing data (Robins, Rotnitzky, and Zhao 1994; van der Laan and Robins, 2003), influence functions of the regression parameter may be expressed by the sum of two (correlated) parts with all parameters taken to be their true values. The first part is the projection of an IPCW-based complete case influence function onto the orthogonal space of the nuisance tangent subspace induced by estimating the censoring mechanism (\hat{G} , in our case; see theorem 1.3 of van der Laan and Robins (2003)). The second part, an augmented term, is a similar projection of a function based on observed data with a conditional expectation that is zero given the (unobserved) complete data. It is this augmented term that results in greater efficiency and/or provides double robustness when chosen properly. Motivated by the previous idea, we consider the following class of (the i th) functions

$$\begin{aligned} &\mathbf{D}_i - \int_0^\tau \{\mathbf{D}_i - \mathbf{K}(t; \mathbf{D})\} \frac{dM_i^c(t)}{G(t)} \\ &+ \int_0^\tau \{\mathbf{e}(t; \mathbf{Z}_i) - \mathbf{K}(t; \mathbf{e})\} \frac{dM_i^c(t)}{G(t)}, \end{aligned} \quad (15)$$

where $\mathbf{e}(t; \mathbf{Z}_i)$ is an arbitrary p -dimensional vector of functions involving \mathbf{Z}_i . The fact that this augmented term (third part of (15)) is indeed such a projection as mentioned earlier and has a conditional expectation zero given the complete data follows from Tsiatis (2006, theorem 9.2

on p. 208 and discussion on p. 217). Also, we note that this term takes into account censored cases and hence includes more “information” than the simple IPCW device. Now, for any chosen $\mathbf{e}(t; \mathbf{Z}_i)$, we can improve efficiency over the simple IPCW estimator by multiplying the last term of (15) by a constant matrix $\Upsilon = \mathbf{V}_1 \mathbf{V}_2^{-1}$, where

$$\mathbf{V}_1 = E \left[\int_0^\tau \mathbf{D}_i \{ \mathbf{e}(t; \mathbf{Z}_i) - \mathbf{K}(t; \mathbf{e}) \}' I(T_i \geq t) \frac{d\Lambda^c(t)}{G(t)} \right],$$

and

$$\mathbf{V}_2 = E \left[\int_0^\tau \{ \mathbf{e}(t; \mathbf{Z}_i) - \mathbf{K}(t; \mathbf{e}) \}^{\otimes 2} I(X_i \geq t) \frac{d\Lambda^c(t)}{G(t)^2} \right].$$

Note that \mathbf{V}_1 is actually the covariance of the second and the third terms in (15) whereas \mathbf{V}_2 is the variance of the third term using the stochastic integration theory. With this modification, the “improved” function (15) now becomes

$$\begin{aligned} & \mathbf{D}_i - \int_0^\tau \{ \mathbf{D}_i - \mathbf{K}(t; \mathbf{D}) \} \frac{dM_i^c(t)}{G(t)} \\ & + \Upsilon \int_0^\tau \{ \mathbf{e}(t; \mathbf{Z}_i) - \mathbf{K}(t; \mathbf{e}) \} \frac{dM_i^c(t)}{G(t)}, \end{aligned} \tag{16}$$

with a variance that is

$$\begin{aligned} & \text{var}\{ \mathbf{D}_i \} + E \left[\int_0^\tau \{ \mathbf{D}_i - \mathbf{K}(t; \mathbf{D}) \}^{\otimes 2} I(X_i \geq t) \frac{d\Lambda^c(t)}{G(t)^2} \right] \\ & - \mathbf{V}_1 \mathbf{V}_2^{-1} \mathbf{V}_1', \end{aligned} \tag{17}$$

which is always smaller than that of the term in the bracket of (14). This implies that if we can construct an estimating equation based on (16), then the resulting estimator should have a smaller variance than $\hat{\beta}_a$ as a result of (17) and the fact that the augmented term in (16) is independent of β . (This condition can actually be weakened, because the augmented term has a zero conditional expectation, as noted later.) To illustrate this point further, we may construct an AIPCW estimating equation in the following way. Define

$$\tilde{\mathbf{D}}_i(\beta) = \int_0^\tau I(T_i > t) \mathbf{Z}_i [(T_i - t) - g\{\widehat{m}_{a0}(t; \beta) + \beta' \mathbf{Z}_i\}] dH(t).$$

A class of AIPCW estimating equations according to (16) is given by

$$\tilde{\mathbf{U}}_a(\beta) = \sum_{i=1}^n \left\{ \frac{\Delta_i \tilde{\mathbf{D}}_i(\beta)}{\widehat{G}(X_i)} + \widehat{\Upsilon}(\widehat{\beta}_a) \int_0^\tau \left\{ \mathbf{e}(t; \mathbf{Z}_i) - \widehat{\mathbf{K}}(t; \mathbf{e}) \right\} \frac{dN_i^c(t)}{\widehat{G}(t)} \right\}, \tag{18}$$

where $\widehat{\mathbf{K}}(t; \mathbf{e}) = \sum_{i=1}^n \mathbf{e}(t; \mathbf{Z}_i) I(X_i \geq t) / \sum_{i=1}^n I(X_i \geq t)$, $\widehat{\Upsilon}(\widehat{\beta}_a) = \widehat{\mathbf{V}}_1(\widehat{\beta}_a) \widehat{\mathbf{V}}_2^{-1}$,

$$\widehat{\mathbf{V}}_1(\widehat{\boldsymbol{\beta}}_a) = n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\Delta_i \widehat{\mathbf{D}}_i(\widehat{\boldsymbol{\beta}}_a)}{\widehat{G}(X_i)} \{ \mathbf{e}(t; \mathbf{Z}_i) - \widehat{\mathbf{K}}(t; \mathbf{e}) \}' I(X_i \geq t) \times \frac{d\widehat{\Lambda}^c(t)}{\widehat{G}(t)},$$

and

$$\widehat{\mathbf{V}}_2 = n^{-1} \sum_{i=1}^n \int_0^\tau \{ \mathbf{e}(t; \mathbf{Z}_i) - \widehat{\mathbf{K}}(t; \mathbf{e}) \}^{\otimes 2} I(X_i \geq t) \frac{d\widehat{\Lambda}^c(t)}{\widehat{G}(t)^2}.$$

Let $\widetilde{\boldsymbol{\beta}}_a$ be the improved estimator—in other words, the solution to $\widetilde{\mathbf{U}}_a(\boldsymbol{\beta}) = \mathbf{0}$. Note that $\widehat{\pi}(t) = \widetilde{F}(t)\widehat{G}(t)$ where $\widetilde{F}(t)$ is the Kaplan-Meier estimator of $F(t)$. Then the variance of $n^{1/2}(\widetilde{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_0)$ can be consistently estimated by $\widetilde{\mathbf{A}}_a^{-1} \sum_a (\widetilde{\mathbf{A}}_a)^{\prime}$, where $\widetilde{\mathbf{A}}_a = -n^{-1} \partial \widetilde{\mathbf{U}}_a(\widetilde{\boldsymbol{\beta}}_a) / \partial \boldsymbol{\beta}' = -n^{-1} \partial \mathbf{U}_a(\widetilde{\boldsymbol{\beta}}_a) / \partial \boldsymbol{\beta}'$ (hence also converges to \mathbf{A}),

$$\begin{aligned} \widetilde{\Sigma}_a &= n^{-1} \sum_{i=1}^n \frac{\Delta_i \widetilde{\mathbf{D}}_i(\widetilde{\boldsymbol{\beta}}_a)^{\otimes 2}}{\widetilde{G}(X_i)} + n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \widetilde{\mathbf{K}}(t; \mathbf{D}) \right. \\ &\quad \left. - \widetilde{\mathbf{K}}(t; \mathbf{D})^{\otimes 2} \right\} \frac{dN_i^c(t)}{\widetilde{G}(t)^2} - \widehat{\mathbf{V}}_1(\widehat{\boldsymbol{\beta}}_a) \widehat{\mathbf{V}}_2^{-1} \widehat{\mathbf{V}}_1(\widehat{\boldsymbol{\beta}}_a)' \end{aligned}$$

and $\widetilde{\mathbf{K}}(t; \mathbf{W}) = \{ n \widetilde{F}(t) \}^{-1} \sum_{i=1}^n \Delta_i \mathbf{W} I(X_i \geq t) / \widetilde{G}(X_i)$ for any random vector or matrix \mathbf{W} .

This class of improved estimators is generated through the functional of $\mathbf{e}(t; \mathbf{Z}_i)$. In practice, we can obtain an improved estimator by choosing any form of $\mathbf{e}(t; \mathbf{Z}_i)$ that has a positive variance. Taking $\mathbf{e}(t; \mathbf{Z}_i) = \mathbf{0}$, we get the simple IPCW estimator $\widehat{\boldsymbol{\beta}}_a$. Within this class of all regular and asymptotically linear estimators, the most efficient estimator is obtained by finding the functional $\mathbf{e}(t; \mathbf{Z}_i)$ that minimizes the variance of (15). According to Robins, Rotnitzky, and Zhao (1994, proposition 2.3 and (41)) and Tsiatis (2006, theorems 10.1 and 10.4, and section 10.4), the minimum variance is attained when $\mathbf{e}(t; \mathbf{Z}_i) = E[\mathbf{D}_i | T_i \geq t, \mathbf{Z}_i]$. This is because right censoring can be regarded as a special case of monotone missing data patterns. As a matter of fact, this optimal estimator also provides double protection against model misspecifications, as we shall discuss later. Computational issues such as how to estimate $E[\mathbf{D}_i | T_i \geq t, \mathbf{Z}_i]$ will be addressed at the end of this section.

Remark—The augmented estimating function $\widetilde{\mathbf{U}}_a(\boldsymbol{\beta})$ in (18) may also be defined by using $\widehat{\boldsymbol{\pi}}(\boldsymbol{\beta})$ instead of $\widehat{\boldsymbol{\pi}}(\widehat{\boldsymbol{\beta}}_a)$. In other words, we may let the augmented term involve $\boldsymbol{\beta}$ as well. As shown by Tsiatis (2006, theorem 10.3) and others, this will not affect the asymptotic variance theoretically. The reason behind this remarkable property is that the augmented term has a conditional expectation zero, as we discussed before (15). However, it is more computationally expensive to implement this new augmented estimating function.

4.2 Double Robustness Study

In a censoring data problem, an estimator is double robust (DR) if it remains consistent when either the model for the censoring mechanism or a model for the complete data distribution is correctly specified (see Robins and Rotnitzky (2001) for an existence theorem). For our

problem, we wish to provide such “double protection” because the censoring distribution might be wrongly modeled. In general, double robustness may result in certain efficiency lost; however, the AIPCW estimator does not have this shortcoming. As argued in a series of studies (e.g., Scharfstein, Rotnitzky, and Robins 1999; van der Laan and Robins 2003; Rotnitzky and Robins 2005), the optimal AIPCW estimator provides both efficiency gain and double robustness.

In the following, we consider the DR property of the optimal estimator derived in the last subsection. The major mathematical tools can be found in van der Laan and Robins (2003, chaps. 1 and 2) and Tsiatis (2006, section 10.3). Recall we have concluded that the efficient influence function is obtained if we take $\mathbf{e}(t; \mathbf{Z}_i) = E[\mathbf{D}_i | T_i \geq t, \mathbf{Z}_i]$ in (15). It can be easily verified that in this case (15) simplifies to

$$\frac{\Delta_i \mathbf{D}_i}{G(X_i)} + \int_0^\tau E[\mathbf{D}_i | T_i \geq t, \mathbf{Z}_i] \frac{dM_i^c(t)}{G(t)}. \quad (19)$$

For notational convenience we will now write \mathbf{D}_i as $\mathbf{D}_i(T_i, \mathbf{Z}_i)$ to emphasize that \mathbf{D}_i depends on T_i and \mathbf{Z}_i , but we suppress the subscript i for the rest of this section so there is no confusion. Define

$$\begin{aligned} \mathbf{Q}_F(t, \mathbf{Z}) &:= -E[\mathbf{D}(T, \mathbf{Z}) | T \geq t, \mathbf{Z}] \\ &= \frac{1}{F(t|\mathbf{Z})} \int_t^\tau \mathbf{D}(u, \mathbf{Z}) dF(u|\mathbf{Z}), \end{aligned}$$

where $F(t|\mathbf{Z})$ is the survival function of T given \mathbf{Z} . Then (19) becomes

$$\mathbf{Y}_{F,G}(X, \Delta, \mathbf{Z}) := \frac{\Delta \mathbf{D}(X, \mathbf{Z})}{G(X)} - \int_0^\tau \mathbf{Q}_F(t, \mathbf{Z}) \frac{dM^c(t)}{G(t)}. \quad (20)$$

Equation (20) implies that the efficient influence function of the estimator depends on two distribution functions, $F(\cdot|\mathbf{Z})$ and $G(\cdot)$, which, in practice, will both need to be modeled (in contrast, the simple IPCW estimator only requires modeling $G(\cdot)$, which under independent censoring assumption is the nonparametric Kaplan-Meier estimator). As in Rubin and van der Laan (2006), where $\mathbf{Y}_{F,G}(X, \Delta, \mathbf{Z})$ possesses the following double robustness property:

$$E\{\mathbf{Y}_{F_1, G_1}(X, \Delta, \mathbf{Z}) | \mathbf{Z}\} = \mathbf{0} \text{ if } F_1 = F \text{ or } G_1 = G. \quad (21)$$

This means that we only need to model one of them correctly. That is, we get two chances of obtaining a consistent estimator. This double robustness property implies that, in practice, a minor misspecification of the model for $G(\cdot|\mathbf{Z})$ can be corrected by doing a good job in modeling $F(\cdot|\mathbf{Z})$, and vice versa.

Writing $\tilde{\mathbf{D}}_i(\beta)$ as $\tilde{\mathbf{D}}_i(\beta; T_i, \mathbf{Z}_i)$ to emphasize that $\tilde{\mathbf{D}}_i$ also depends on T_i and \mathbf{Z}_i , the DR estimator $\hat{\beta}_{DR}$, which is also the efficient AIPCW estimator, can be given as the solution to $\mathbf{U}_{DR}(\beta) = \mathbf{0}$, where

$$\begin{aligned}
 \mathbf{U}_{DR}(\boldsymbol{\beta}) &= \sum_{i=1}^n \left[\frac{\Delta_i \tilde{\mathbf{D}}_i(\boldsymbol{\beta}; T_i, \mathbf{Z}_i)}{\tilde{G}(X_i)} - \int_0^\tau \widehat{\mathbf{Q}}_F(t, \mathbf{Z}_i) \frac{d\tilde{M}_i^c(t)}{\tilde{G}(t)} \right], \\
 \widehat{\mathbf{Q}}_F(t, \mathbf{Z}_i) &= \frac{1}{\widehat{F}(t|\mathbf{Z}_i)} \int_t^\tau \tilde{\mathbf{D}}_i(\boldsymbol{\beta}; u, \mathbf{Z}_i) d\widehat{F}(u|\mathbf{Z}_i),
 \end{aligned}
 \tag{22}$$

and $\widehat{F}(t|\mathbf{Z})$ is an estimator of $F(t|\mathbf{Z})$. This may seem like circular reasoning, because if we had methods to estimate $F(t|\mathbf{Z})$, we would not need to develop this whole theory in the first place. The rationale here is that we can propose a simple working model for estimating $F(t|\mathbf{Z})$. For example, we may use a parametric model or a Cox model even if they are not correct. An interesting fact is that we can even use model (5) for this purpose. That is, we try to obtain an estimator of $F(t|\mathbf{Z})$ under model (5) (note that this estimator cannot be based on the censoring distribution assumption), plug it back to (22), and construct more efficient and robust estimating equations. This way, the new estimator will be consistent and asymptotically normal if either model (5) (or whatever models we may wish to posit as a working model for approximating $F(t|\mathbf{Z})$) or the censoring mechanism assumption is correct.

5. SIMULATION STUDIES

In this section we report some results from simulation studies conducted to assess the performance of the estimation procedures developed in Sections 2, 3, and 4 with the focus on estimation of the scalar β_0 . For cases in which the censoring time C is independent of T and Z , we let $Z \sim \text{Bernoulli}\{0, 1\}$ with probability 0.5 and C follow a uniform distribution $[0, V]$, with V varying to yield desiring censoring percentages. For cases in which the censoring time depends on the covariate Z , C is generated from the Cox model (8) with $\gamma_0 = 1$ and $\lambda_0(t)$ being various constants for generating different censoring percentages. In both cases, care is taken to ensure the support of C is longer than the support of T for all Z . We considered several transformation functions g , but for the purpose of comparison we will only report in this section the results from $g(t) = \exp(t)$, which corresponds to the proportional mean residual life model (3) and $g(t) = t$, which corresponds to the additive mean residual life model (4). Data analysis results based on other choices of g can be found in Section 6. To generate T according to model (5), the true β_0 is chosen to be 0 or 0.5 and the baseline function $m_0(t)$ is taken from the Hall-Wellner family—in other words, $m_0(t) = g^{-1}\{(D_1 t + D_2)^+\}$, where $D_1 > -1$, $D_2 > 0$, and d^+ denotes $dI(d \geq 0)$ for any quantity d . In our simulation, we consider two scenarios. One is that $D_1 = -0.5$ and $D_2 = 0.5$, which gives a uniform distribution $[0, 1]$ when $Z = 0$. The other is $D_1 = -1/3$ and $D_2 = 1$, which corresponds to the survival function $F(t) = (1 - t/3)^2$, $0 \leq t \leq 3$ when $Z = 0$ and mimics the exponential distribution resulting from the right skewness. When $Z = 0.5$, one can use the formula

$$F(t|Z) = \frac{m_0(t|Z)}{m_0(t|Z)} \exp \left\{ - \int_0^t \frac{du}{m_0(u|Z)} \right\}$$

to compute the conditional survival function of T , where $m(t|Z)$ follows model (5) and depends on the choice of g .

Because it is difficult to evaluate the effect of the integrator H analytically, we compare simulation results from different H with the focus on the variance of $\hat{\beta}_a$ and $\hat{\beta}_b$. Notice that, theoretically, any H that is increasing and converges almost surely to a deterministic function—say, H —should give an asymptotically unbiased point estimate of β_0 . Thus, the asymptotic variance indicates the efficiency of the estimator. In our simulations, we compare three

different H functions: (1) $H_1(t) = \sum_{i=1}^n \Delta_i I(X_i \leq t)$, (2) H_2 is a step function with jumps 1 at

fixed time points (e.g., seven preselected time points), and (3) $H_3(t) = 0$ if $t < 0$ and 1 otherwise. Here, the first H takes into account all uncensored observations and assigns them equal weight. The second one reduces computation needs and may avoid tail instability. The last H only has a jump at $t = 0$ and may be used if one is only interested in knowing β_0 . All simulation studies are based on 1,000 replications.

Table 1 shows the results from independent censoring. The censoring percentages are around 10% or 30% and the sample sizes are 100 or 200. It can be seen that the point estimates of β_0 and their standard error estimates are very accurate. Besides, the 95% empirical coverage probabilities based on normal approximation are reasonable. Table 2 shows similar results for cases in which the censoring time depends on the covariate. For both cases, the baseline MRLF with $D_1 = -0.5$ and $D_2 = 0.5$ mentioned earlier is used. Q-Q plots for the estimates indicate that asymptotic normality is a good approximation (figures not shown). Therefore, we conclude that the limiting distributions for the proposed estimators work quite well with moderate sample sizes. We also considered other choices of the transformation function g and different simulation setups, and obtained similar results.

In Table 3 we compare the performance of IPCW and two AIPCW estimators using $H_1(t)$. Moreover, we program the quasi-partial score (QPS) estimators proposed in Chen and Cheng (2005,2006) for the proportional mean residual life model and the additive mean residual life model, respectively. To make our simulation results comparable with those of Chen and Cheng (2005,2006), we mimic their simulation setup by letting the censoring be an independent uniform random variable, covariate a Bernoulli random variable as before, and the baseline T either uniform or right skewed as mentioned earlier. Because the validity of the IPCW and QPS estimators have already been established earlier and in Chen and Cheng (2005,2006), respectively, we only report their relative efficiency with respect to the AIPCW estimator (AIPCW1, to be defined shortly). Here, by convention, the relative efficiency of an unbiased estimator $\hat{\theta}_1$ with respect to another unbiased estimator $\hat{\theta}_2$ for the true parameter θ_0 is defined as $\text{var}(\hat{\theta}_2)/\text{var}(\hat{\theta}_1)$, which in our simulations is approximated by their corresponding sample estimates. In Table 3, the two AIPCW estimators are based on (22). That is, they are the efficient and DR estimators. The difference between them is that one (denoted by AIPCW1 in Table 3) uses the Kaplan-Meier estimator for calculating $\hat{F}(t|Z)$, which is applicable when the covariate takes finitely many values. The other (denoted by AIPCW2 in Table 3) uses the estimate $\hat{\beta}_a$, $\hat{m}_{a0}(t)$, model (5), and the relationship between MRLF and the survival function described in Section 1 for calculating $\hat{F}(t|Z)$. That is,

$$\begin{aligned} \hat{F}(t|Z) &= \frac{\hat{m}(0|Z)}{\hat{m}(t|Z)} \exp \left\{ - \int_0^t \frac{du}{\hat{m}(u|Z)} \right\} \\ &= \frac{g(\hat{m}_{a0}(0) + \hat{\beta}_a Z)}{g(\hat{m}_{a0}(t) + \hat{\beta}_a Z)} \exp \left\{ - \int_0^t \frac{du}{g(\hat{m}_{a0}(u) + \hat{\beta}_a Z)} \right\}. \end{aligned}$$

Note that this would be the choice when the covariate is continuous or takes too many values compared with the sample size. It can be seen that both AIPCW estimators are quite accurate, with AIPCW1 slightly more efficient. Moreover, the AIPCW estimators are more efficient than the IPCW estimator as expected. The comparisons between our estimators and the QPS estimators indicate that our estimators have greater efficiency in all scenarios.

6. APPLICATIONS

In this section, we apply the proposed methodology to two real-life datasets obtained from clinical trials.

6.1 Lung Cancer Study

The purpose of this clinical trial was to investigate the impact of the systematic combination of chemotherapy with regular radiotherapy on survival and recurrence patterns in patients with surgically resected lung cancer. In total, 172 patients were randomized between November 1979 and May 1985 with 86 in the radiotherapy (RT) group, 78 in the radio-therapy plus chemotherapy (RT+CAP) group, and the remaining eight patients became ineligible for statistical analysis. Both time to death and disease-free survival time are of interest in this study. For the purpose of illustration, we consider only the disease-free survival time as the response (T), which is subject to right censoring if the patient was alive and had never had a tumor recurrence by the end of the study. In addition to the treatment factor, which is denoted by Z_1 , we also consider another important histology factor, squamous versus non-squamous/mixed cell type, and denote it by Z_2 . We define $Z_1 = 1$ if the patient was in the RT+CAP group and 0 otherwise, and $Z_2 = 1$ if the patient had the squamous cell type and 0 otherwise. For a detailed description and some preliminary statistical analysis results, see Lad, Rubinstein, and Sadeghi (1988) and Piantadosi (2005).

To illustrate our method, we first apply the Cox regression model by using the censoring time as the response and the two covariates mentioned earlier. It appears that the covariate effects are not significant (both p -values > 0.50). However, we caution here that it is certainly possible to apply other approaches to find better solutions for modeling the censoring mechanism. Besides, it could be the case that the covariate effects (on censoring) are nonlinear or the distributions are nonnormal so that the regression coefficients are zero but independence fails, as pointed out by the editor. However, for purposes of illustration and for the time being, we treat censoring as independent for simplicity. We then apply the proposed transformed mean residual life model and estimate the regression parameter β_0 (together with the baseline MRLF $g\{m_0(t)\}$) by using the estimating equations (6) and (7) with the integrator H_1 and three different choices of the transformation function g . It is seen (Table 4) that both covariates have a significant impact on the mean residual life. Namely, with adjuvant chemotherapy, the expected residual disease-free time is significantly longer. Also, patients in the nonsquamous/mixed cell-type group tend to have longer expected residual disease-free time.

Notice that the magnitudes of the parameter estimates are quite different for various choices of the link function g . This is because the parameters have different interpretations for different g . For instance, if $g(t) = t$ (i.e., model (5) reduces to the additive model (4)), then the parameter estimate means that the adjuvant chemotherapy increases the expected residual disease-free time by approximately 383 days when compared with radiotherapy alone. On the other hand, when $g(t) = \exp(t)$, which yields model (3), the parameter estimate should be interpreted as the ratio of the mean residual function between the two groups. In other words, the adjuvant chemotherapy doubles ($e^{0.726} \approx 2$) the mean residual disease-free time when compared with radiotherapy alone. Also using the proportional mean residual life model and their estimation approach, Chen et al. (2005) obtained similar results.

An interesting phenomenon is that if we use the Cox proportional hazards model for this problem we will get a different interpretation on the effect of cell type (although for the effect of treatment, the interpretations are in the same direction). The Cox model indicates that both treatment and cell type are significant covariates (p -values are 0.001 and 0.003, respectively), but the coefficient for cell type is -0.570 , which means patients in the nonsquamous/mixed cell-type group have a higher risk. A closer examination of the data reveals the reason. The largest uncensored observation (recurrence time) in the squamous group is only 1,388 days, whereas the largest three uncensored observations in the nonsquamous/mixed group are 1,577, 2,375, and 2,566 days. In other words, the mean residual time for the nonsquamous/mixed group is greatly inflated by these extremely large values. One can also see this from the fact that the Cox proportional hazards model and most other survival regression models based on

the hazard function essentially measure the instantaneous risk for failure, which might be explained as a conditional probability, whereas the mean residual lifetime measures some conditional average life span and thus can be heavily affected by extreme values. As is well known, the Cox model parameter estimate is rank invariant, but that of the mean residual lifetime model is not.

Theoretically, it has been pointed out (Shaked and Shantikumar, 1993) that the ordering of the MRLF is weaker than the ordering of the hazard function. Namely, if two MRLFs $m_a(t)$ and $m_b(t)$ have the relationship that $m_a(t) \geq m_b(t)$ for all t , then their corresponding hazard functions, denoted by $\lambda_a(t)$ and $\lambda_b(t)$, respectively, need not necessarily follow that $\lambda_a(t) \leq \lambda_b(t)$ for all t . The inverse, however, is always true. This implies that modeling the MRLF can be an attractive alternative when the ordering assumption (for instance, the proportional hazards assumption) on the hazard function is violated. For the lung cancer study considered here, graphical partial residual diagnostics according to Grambsch and Therneau (1994) suggest that the proportional hazards assumption is questionable for the covariate cell type as a result of the quite obvious systematic pattern of the residuals.

6.2 Prostate Cancer Study

The interest of a recent study conducted at the Memorial Sloan-Kettering Cancer Center is to examine the effect of the radiation therapy dose level and patient age on the PSA relapse time (see the Introduction for the definition). Let Z_1 denote the age of the patient before the initial treatment, which ranges from 42 to 87 years continuously, and Z_2 denote dose, which takes two values: 1 = low dose (≤ 70.2 Gy) and 2 = high dose ($(75.6$ Gy, 81 Gy]) (these cutoffs are suggested by experienced radiotherapists). In total, 191 patients underwent radiotherapy. These patients were all classified as high risk based on their baseline PSA value, Gleason score, and T-stage. They also all had adjuvant hormonal therapy. In this analysis, we first apply a Cox model on the censoring time using age and dose level as covariates. It turns out that both are significant and thus we have the situation discussed in Section 3 (i.e., the censoring mechanism depends on covariates and needs to be modeled). Again, we emphasize here that a more comprehensive study may be conducted to fine-tune the modeling approaches for the censoring distribution. However, for the time being, we shall use the Cox model for such a purpose. Implementing (10) and (11) gives us coefficient estimates of age and dose level of 14.60 and 127.47, respectively. The standard errors for these two point estimates, based on Theorem 2, are 8.61 and 68.12. These results suggest that higher doses significantly elongate the residual PSA relapse time, whereas age has only a marginal effect. An interesting issue is that older patients tend to have longer relapse times, which is somewhat counterintuitive, but agrees with a previous study using the Cox model for assessing hazards (Zeilefsky et al. 2008). Note here that, as discussed in the Introduction, we used the Box-Cox transformation for the link function g with $\rho = 0.5$ because neither the additive nor the multiplicative modeling assumption sounds natural for this clinical problem. However, one can certainly enjoy other choices of g , and thus the flexibility of our approach, based on prior data information or physicians' assertions if they are available.

7. CONCLUDING REMARKS

In this article we proposed a class of transformation mean residual life models. This very general class of models encompasses some previously established models. For example, the proportional mean residual life model and the additive mean residual life model are special cases of our model. IPCW-based generalized estimating equation approaches are used to estimate the regression parameters of interest. We also considered efficiency and double robustness properties of our estimators and suggested improved (AIPCW) estimators. Both asymptotic theory and simulation studies show that the IPCW and the AIPCW estimators perform well.

Notice that, unlike some previously established methods, the proposed estimating equations here have intuitive meanings. They essentially yield the weighted least-square-type estimators, with the weight determined by the censoring probability. Therefore, although our approaches are proposed for regression purposes, they can be very easily modified for one-sample problems. Our approaches provide a nice platform for obtaining efficient estimators (without censoring, our approaches naturally reduce to that of Bickel et al. (1993)), and our efficiency and double robustness arguments also seem to be the first for the MRLF estimation problems under right censoring.

As we saw in both theory and applications, understanding and modeling the censoring mechanism is important for our approaches. Even though the double robustness property provides us a certain amount of luxury of misspecifying the model of censorship, we should make every effort to find a good approximation. In our numerical studies, when the censoring time may depend on \mathbf{Z} , the Cox proportional hazards model was used as the working model for the censoring time. Of course, we can also choose some other semiparametric regression models as the working model for censoring. For example, we may use the proportional mean residual life model or the additive mean residual life model and obtain the estimators of the censoring model parameters using the approach of Chen and Cheng (2005) or Chen and Cheng (2006). Then the estimator of the censoring survival distribution $G(t; \mathbf{Z})$ can be obtained using the inversion formula as before. Thus, the unknown parameter in model (5) can be estimated by using the procedures in Sections 3 and 4.

For the examples, various models with different choices of the link function g were considered. It would be helpful to provide a method for selecting or comparing different link functions. For a given link function, following Lin et al. (2000), we can use a residual-based procedure for checking the adequacy of the model. If, on the other hand, we restrict the function g to the Box-Cox transformation family but allow ρ to be estimated from the observed data, then it would provide a way in model selection between the additive mean residual life model and other alternatives. Let $g_\rho(x) = [(x + 1)^\rho - 1]/\rho$. For parameter estimation, similar to (7) and (11), we may use the following class of estimating equations for β_0 and ρ :

$$\mathbf{U}_a^*(\beta, \rho) = \sum_{i=1}^n \int_0^{\tau} \frac{\Delta_i I(X_i > t) \mathbf{Z}_i^*(t)}{\widehat{G}(X_i)} [(X_i - t) - g_\rho\{\widehat{m}_{a0}^*(t; \beta, \rho) + \beta' \mathbf{Z}_i\}] dH(t) = \mathbf{0}$$

under independent censoring and

$$\mathbf{U}_b^*(\beta, \rho) = \sum_{i=1}^n \int_0^{\tau} \frac{\Delta_i I(X_i > t) \mathbf{Z}_i^*(t)}{G_i(X_i; \widehat{\gamma}, \widehat{\lambda}_0)} [(X_i - t) - g_\rho\{\widehat{m}_{b0}^*(t; \beta, \rho) + \beta' \mathbf{Z}_i\}] dH(t) = \mathbf{0}$$

under dependent censoring, where $\mathbf{Z}_i^*(t) = (\mathbf{Z}'_i, W_i(t))'$, $W_i(t)$ is a known bounded function of t , \mathbf{Z}_i , and C_i , and $\widehat{m}_{a0}^*(t; \beta, \rho)$ and $\widehat{m}_{b0}^*(t; \beta, \rho)$ are the solutions to (6) and (10) by replacing g with g_ρ . The asymptotic properties of the estimators for β_0 and ρ can be obtained along the lines of this article, and will be reported elsewhere.

One important issue with the mean residual life model is the embedded model constraint. That is, $m(t) + t$ should be nondecreasing in t . Chen and Cheng (2005, 2006) discussed this constraint for the proportional mean residual life model (3) and the additive mean residual life model (4). For the class of transformed mean residual life model (5), however, $m(t; \mathbf{Z}) + t = g\{m_0(t) + \beta' \mathbf{Z}\} + t$ may not always satisfy this constraint for certain β unless $m_0(t)$ itself is nondecreasing.

The necessary condition for this constraint is that $g\{m_0(t)\} + t$ is nondecreasing. Further research is needed to provide a necessary and sufficient condition for this constraint under model (5) with a general transformation function g .

Another interesting issue is the effect of the integrator H , which essentially plays the role of a weight function. Our simulation results seem to suggest that estimating the baseline function $m_0(t)$ at all uncensored observations does not necessarily improve either the accuracy of the point estimate or its efficiency. Here we refer interested readers to Lin, Wei, and Ying (2001), where they also pointed out the difficulty of finding an optimal H analytically and thus compared different H by simulation in dealing with point processes.

Acknowledgments

The first author's research was supported in part by the National Natural Science Foundation of China Grants (nos. 10571169 and 10731010) and the National Basic Research Programme (no. 2007CB814902). The second author's research was supported in part by the National Cancer Institute Core Grant (CA008748). The authors thank the editor, Professor Stephen L. Portnoy, an associate editor, and two referees for their insightful comments and suggestions that greatly improved the article.

References

- Arnold BC, Zahedi H. On Multivariate Mean Remaining Life Functions. *Journal of Multivariate Analysis* 1988;25:1–9.
- Balkema AA, de Haan L. Residual Life Time at Great Age. *Annals of Probability* 1974;2:792–804.
- Bang H, Tsiatis AA. Estimating Medical Costs with Censored Data. *Biometrika* 2000;87:329–343.
- Bang H, Tsiatis AA. Median Regression with Censored Cost Data. *Biometrics* 2002;58:643–649. [PubMed: 12229999]
- Bassan B, Kochar S, Spizzichino F. Some Bivariate Notions of IFR and DMRL and Related Properties. *Journal of Applied Probability* 2002;39:533–544.
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. *Efficient and Adaptive Inference in Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press; 1993.
- Chen YQ. Additive Regression of Expectancy. *Journal of the American Statistical Association* 2007;102:153–166.
- Chen YQ, Cheng SC. Semiparametric Regression Analysis of Mean Residual Life with Censored Survival Data. *Biometrika* 2005;92:19–29.
- Chen YQ, Cheng SC. Linear Life Expectancy Regression with Censored Data. *Biometrika* 2006;93:303–313.
- Chen YQ, Jewell NP, Lei X, Cheng SC. Semiparametric Estimation of Proportional Mean Residual Life Model in Presence of Censoring. *Biometrics* 2005;61:170–178. [PubMed: 15737090]
- Cox, DR. *Renewal Theory*. London: Methuen; 1961.
- Dauxois J-Y. Estimating the Deviation from Exponentiality under Random Censorship. *Communications in Statistics Theory and Methods* 2003;32:2117–2137.
- El-Bassiouny AH, Alwasel IA. A Goodness of Fit Approach to Testing Mean Residual Times. *Applied Mathematics and Computation* 2003;143:301–307.
- Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. New York: John Wiley; 1991.
- Grambsch P, Therneau T. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* 1994;81:515–526.
- Hall, WJ.; Wellner, JA. Mean Residual Life. In: Csörgő, M.; Dawson, DA.; Rao, JNK.; Saleh, AKME., editors. *Proceedings of the International Symposium on Statistics and Related Topics*. Amsterdam: North Holland; 1984. p. 169-184.
- Hollander M, Proschan F. Tests for the Mean Residual Life. *Biometrika* 1975;62:585–593.
- Kotz S, Shanbhag DN. Some New Approaches to Probability Distribution. *Advances in Applied Probability* 1980;12:903–921.

- Lad T, Rubinstein L, Sadeghi A. The Benefit of Adjuvant Treatment of Resected Locally Advance Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology* 1988;6:9–17. [PubMed: 2891798]
- Liang KY, Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* 1986;73:13–22.
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society, Ser B* 2000;62:711–730.
- Lin DY, Wei LJ, Ying Z. Semiparametric Transformation Models for Point Processes. *Journal of the American Statistical Association* 2001;96:620–628.
- Maguluri G, Zhang CH. Estimation in the Mean Residual Life Regression Model. *Journal of the Royal Statistical Society, Ser B* 1994;56:477–489.
- Newey WK. Semiparametric Efficiency Bounds. *Journal of Applied Econometrics* 1990;5:99–135.
- Oakes D, Dasu T. A Note on Residual Life. *Biometrika* 1990;77:409–410.
- Piantadosi, S. *Clinical Trials: A Methodologic Perspective*. 2. New York: Wiley; 2005.
- Robins JM, Ritov Y. Towards a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semiparametric Models. *Statistics in Medicine* 1997;16:285–319. [PubMed: 9004398]
- Robins, JM.; Rotnitzky, A. Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In: Jewell, N.; Dietz, K.; Farewell, V., editors. *AIDS Epidemiology: Methodologic Issues*. Boston: Birkhauser; 1992. p. 297-331.
- Robins JM, Rotnitzky A. Comment on the Bickel and Kwon Article, ‘Inference for Semiparametric Models: Some Questions and an Answer’. *Statistica Sinica* 2001;11:920–936.
- Robins JM, Rotnitzky A, Zhao LP. Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association* 1994;89:846–866.
- Rojo J, Ghebremichael M. Estimation of Two Ordered Bivariate Mean Residual Life Functions. *Journal of Multivariate Analysis* 2006;97:431–454.
- Rotnitzky, A.; Robins, JM. Inverse Probability Weighting in Survival Analysis. In: Armitage, P.; Colton, T., editors. *Encyclopedia of Biostatistics*. 2. Chichester, UK: Wiley; 2005. p. 2619-2625.
- Rubin, D.; van der Laan, MJ. Doubly Robust Censoring Unbiased Transformations. 2006. available at http://works.bepress.com/mark_van_der_laen/57
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models,” (with discussions and rejoinder). *Journal of the American Statistical Association* 1999;94:1096–1146.
- Shaked, M.; Shantikumar, J. *Stochastic Orders and Their Applications*. Boston: Academic Press; 1993.
- Swartz GB. The Mean Residual Life Function. *IEEE Transactions on Reliability* 1973;R-22:108–109.
- Tsiatis, AA. *Semiparametric Theory and Missing Data*. New York: Springer-Verlag; 2006.
- van der Laan, MJ.; Robins, JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer Verlag; 2003.
- Zelefsky MJ, Yamada Y, Fuks Z, Zhang Z, Hunt M, Cahlon O, Park J, Shippy A. Long-Term Results of Conformal Radiotherapy for Prostate Cancer: Impact of Dose Escalation on Biochemical Tumor Control and Distant Metastases-Free Survival Outcomes. *International Journal of Radiation Oncology Biology Physics* 2008;71:1028–1033.
- Zhao H, Tsiatis AA. Efficient Estimation of the Distribution of Quality-Adjusted Survival Time. *Biometrics* 1999;55:1101–1107. [PubMed: 11315054]

APPENDIX: REGULARITY CONDITIONS

The proofs of asymptotic properties and double robustness are provided in the supplemental technical report. Here, we only outline the regularity conditions.

(C1) $H(t)$ converges almost surely to a nonrandom and bounded function $\tilde{H}(t)$ uniformly in $t \in [0, \tau]$.

(C2) \mathbf{Z} is bounded and G is continuous.

(C3) $m_0(t)$ is continuously differentiable on $[0, \tau]$.

(C4) $\mathbf{A} := E \left[\int_0^\tau I(T_i > t) \{\mathbf{Z}_i - \bar{\mathbf{z}}_a(t)\}^{\otimes 2} \dot{g} \{m_0(t) + \beta'_0 \mathbf{Z}_i\} d \tilde{H}(t) \right]$ is nonsingular where $\bar{\mathbf{z}}_a(t)$ is the limit of $\mathbf{Z}_a(t; \beta_0)$.

(C5) $\mathbf{B} := E \left[\int_0^\tau I(T_i > t) \{\mathbf{Z}_i - \bar{\mathbf{z}}_b(t)\}^{\otimes 2} \dot{g} \{m_0(t) + \beta'_0 \mathbf{Z}_i\} d \tilde{H}(t) \right]$ is nonsingular where $\bar{\mathbf{z}}_b(t)$ is the limit of $\mathbf{Z}_b(t; \beta_0)$.

Table 1

Simulation results: IPCW estimator, independent censoring.

		$g(t) = \exp(t)$												
		$H_1(t)$				$H_2(t)$				$H_3(t)$				
β_0	n	C%	$\hat{\beta}_a$	SE	SEE	CP	$\hat{\beta}_a$	SE	SEE	CP	$\hat{\beta}_a$	SE	SEE	CP
0	100	10	-0.006	0.1230	0.1200	0.944	-0.001	0.1162	0.1162	0.953	-0.001	0.1212	0.1226	0.953
0	100	30	-0.007	0.1611	0.1531	0.934	-0.004	0.1498	0.1480	0.947	-0.001	0.1557	0.1478	0.937
0	200	10	-0.005	0.0856	0.0843	0.954	0.001	0.0816	0.0820	0.950	0.005	0.0857	0.0867	0.956
0	200	30	-0.002	0.1115	0.1085	0.945	0.005	0.1082	0.1049	0.941	-0.000	0.1062	0.1040	0.947
0.5	100	10	0.499	0.0919	0.0917	0.951	0.507	0.0920	0.0897	0.943	0.502	0.0962	0.0962	0.954
0.5	100	30	0.502	0.1055	0.1060	0.949	0.507	0.1083	0.1043	0.944	0.505	0.1089	0.1081	0.944
0.5	200	10	0.502	0.0641	0.0648	0.951	0.503	0.0643	0.0633	0.951	0.505	0.0686	0.0684	0.953
0.5	200	30	0.501	0.0763	0.0755	0.950	0.503	0.0736	0.0736	0.951	0.503	0.0762	0.0760	0.951

		$g(t) = t$												
		$H_1(t)$				$H_2(t)$				$H_3(t)$				
β_0	n	C%	$\hat{\beta}_a$	SE	SEE	CP	$\hat{\beta}_a$	SE	SEE	CP	$\hat{\beta}_a$	SE	SEE	CP
0	100	10	0.002	0.0421	0.0407	0.941	0.002	0.0425	0.0418	0.948	-0.002	0.0605	0.0607	0.948
0	100	30	-0.002	0.0570	0.0538	0.940	-0.003	0.0552	0.0532	0.941	0.003	0.0736	0.0720	0.941
0	200	10	0.000	0.0291	0.0287	0.943	-0.001	0.0301	0.0284	0.949	-0.002	0.0425	0.0431	0.951
0	200	30	0.000	0.0398	0.0381	0.940	0.000	0.0388	0.0378	0.943	-0.001	0.0520	0.0514	0.941
0.5	100	10	0.502	0.0748	0.0749	0.950	0.498	0.0773	0.0769	0.941	0.502	0.0951	0.0946	0.939
0.5	100	30	0.490	0.0973	0.0908	0.924	0.489	0.0991	0.0924	0.916	0.486	0.1117	0.1081	0.940
0.5	200	10	0.503	0.0532	0.0528	0.949	0.499	0.0537	0.0543	0.943	0.499	0.0674	0.0671	0.951
0.5	200	30	0.498	0.0658	0.0646	0.931	0.495	0.0646	0.0657	0.948	0.499	0.0788	0.0770	0.941

NOTE: $\hat{\beta}_a$ represents the mean of the (1,000) point estimates of β_0 . SE represents the sample standard error of the estimates. SEE represents the mean of the standard error of $\hat{\beta}_a$. CP represents the empirical normal 95% coverage probability.

Table 2

Simulation results: IPCW estimator, dependent censoring.

		$g(t) = \exp(t)$												
		$H_1(t)$				$H_2(t)$				$H_3(t)$				
β_0	n	C%	$\hat{\beta}_b$	SE	SEE	CP	$\hat{\beta}_b$	SE	SEE	CP	$\hat{\beta}_b$	SE	SEE	CP
0	100	10	0.009	0.1240	0.1203	0.946	0.008	0.1191	0.1172	0.944	0.016	0.1151	0.1227	0.962
0	100	30	-0.001	0.1532	0.1535	0.949	0.003	0.1570	0.1500	0.932	0.001	0.1446	0.1489	0.953
0	200	10	0.007	0.0836	0.0850	0.950	0.009	0.0821	0.0827	0.947	0.007	0.0855	0.0863	0.953
0	200	30	0.000	0.1076	0.1093	0.947	0.003	0.1046	0.1072	0.943	0.008	0.0957	0.1059	0.969
0.5	100	10	0.505	0.0879	0.0890	0.957	0.505	0.0895	0.0865	0.942	0.502	0.0937	0.0938	0.950
0.5	100	30	0.504	0.0970	0.0992	0.955	0.505	0.1077	0.0964	0.941	0.502	0.0994	0.1010	0.953
0.5	200	10	0.504	0.0628	0.0632	0.952	0.507	0.0614	0.0617	0.953	0.505	0.0662	0.0669	0.952
0.5	200	30	0.504	0.0710	0.0700	0.948	0.502	0.0668	0.0688	0.960	0.507	0.0705	0.0718	0.957

		$g(t) = t$												
		$H_1(t)$				$H_2(t)$				$H_3(t)$				
β_0	n	C%	$\hat{\beta}_b$	SE	SEE	CP	$\hat{\beta}_b$	SE	SEE	CP	$\hat{\beta}_b$	SE	SEE	CP
0	100	10	0.003	0.0422	0.0412	0.946	0.003	0.0412	0.0408	0.947	0.006	0.0609	0.0615	0.943
0	100	30	0.002	0.0555	0.0551	0.947	0.000	0.0522	0.0529	0.952	-0.007	0.0737	0.0744	0.949
0	200	10	0.001	0.0283	0.0290	0.956	-0.002	0.0287	0.0288	0.948	0.004	0.0440	0.0435	0.948
0	200	30	-0.001	0.0380	0.0393	0.959	0.004	0.0375	0.0377	0.952	0.002	0.0523	0.0532	0.952
0.5	100	10	0.525	0.0909	0.0804	0.914	0.519	0.1015	0.0831	0.924	0.539	0.1187	0.1011	0.903
0.5	100	30	0.518	0.1193	0.1148	0.928	0.494	0.1399	0.1147	0.912	0.520	0.1456	0.1361	0.925
0.5	200	10	0.517	0.0593	0.0566	0.935	0.517	0.0618	0.0596	0.939	0.524	0.0756	0.0718	0.928
0.5	200	30	0.516	0.0793	0.0832	0.950	0.510	0.0906	0.0884	0.950	0.517	0.0952	0.1000	0.955

NOTE: $\hat{\beta}_b$ represents the mean of the (1,000) point estimates of β_0 . SE represents the sample standard error of the estimates. SEE represents the mean of the standard error of $\hat{\beta}_b$. CP represents the empirical normal 95% coverage probability.

Table 3

Simulation results: comparison of IPCW, AIPCW, and QPS estimators.

		$g(t) = \exp(t)$								
		Baseline HW ($D_1 = -0.5, D_2 = 0.5$)			Baseline HW ($D_1 = -1/3, D_2 = 1$)					
β_0	n	C%	AIPCW1	AIPCW2	IPCW	QPS1	AIPCW1	AIPCW2	IPCW	QPS1
			$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE	$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE
0	100	10	-0.004	-0.007	1.01	0.95	-0.013	-0.018	0.99	0.92
0	100	30	-0.013	-0.030	0.76	0.76	-0.030	-0.049	0.88	0.67
0	200	10	-0.004	-0.006	1.00	0.94	-0.008	-0.011	1.00	0.90
0	200	30	-0.009	-0.017	0.92	0.76	-0.018	-0.031	0.84	0.67
0.5	100	10	0.501	0.500	0.99	0.98	0.499	0.496	0.97	0.95
0.5	100	30	0.508	0.506	0.97	0.93	0.493	0.484	0.95	0.81
0.5	200	10	0.500	0.500	1.00	0.98	0.501	0.500	0.99	0.96
0.5	200	30	0.504	0.503	1.00	0.92	0.500	0.496	0.97	0.80

		$g(t) = t$								
		Baseline HW ($D_1 = -0.5, D_2 = 0.5$)			Baseline HW ($D_1 = -1/3, D_2 = 1$)					
β_0	n	C%	AIPCW1	AIPCW2	IPCW	QPS2	AIPCW1	AIPCW2	IPCW	QPS2
			$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE	$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE
0	100	10	-0.003	-0.004	0.99	0.96	-0.013	-0.018	0.98	0.91
0	100	30	-0.004	-0.010	0.91	0.74	-0.015	-0.033	0.84	0.73
0	200	10	-0.002	-0.003	1.00	0.96	-0.011	-0.014	0.99	0.90
0	200	30	-0.004	-0.007	0.96	0.82	-0.012	-0.025	0.61	0.68
0.5	100	10	0.502	0.501	1.00	1.00	0.511	0.508	1.00	0.95
0.5	100	30	0.491	0.491	0.99	0.94	0.522	0.511	0.91	0.86
0.5	200	10	0.498	0.498	1.00	0.99	0.518	0.517	1.00	0.95

$g(t) = \exp(t)$												
Baseline HW ($D_1 = -0.5, D_2 = 0.5$)						Baseline HW ($D_1 = -1/3, D_2 = 1$)						
	AIPCW1	AIPCW2	IPCW	QPS1	AIPCW1	AIPCW2	IPCW	QPS1	AIPCW1	AIPCW2	IPCW	QPS1
β_0	$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE	$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE	$\hat{\beta}_1$	$\hat{\beta}_2$	RE	RE
0.5	200	30	0.495	0.495	0.99	0.96	0.54	0.526	0.522	0.97	0.82	0.63

NOTE: AIPCW1 and AIPCW2 are both based on (22), but with $\hat{F}(t|Z)$ estimated by Kaplan-Meier and by assuming model (5), respectively. IPCW is equivalent to \hat{f}_a (i.e., the solution of (7)). QPS1 and QPS2 are the quasi-partial likelihood estimators in Chen and Cheng (2005) and Chen and Cheng (2006), respectively. RE, relative efficiency of the estimator with respect to AIPCW1.

Table 4

Regression parameter estimates with standard errors for the lung cancer clinical trial data.

Model	Regression parameter estimates			
	Treatment (Z_1)		Cell type (Z_2)	
	$\hat{\beta}_1$	SE	$\hat{\beta}_2$	SE
Transformed mean residual life model: $g(t) = \exp(t)$	0.726	0.278	-0.524	0.257
Transformed mean residual life model: $g(t) = t$	383.220	180.169	-275.804	160.738
Transformed mean residual life model: $g(t) = ((t + 1)^2 - 1)/2$	11.768	4.892	-8.602	4.592

NOTE: $g(t) = \exp(t)$ corresponds to the proportional mean residual life model, $g(t) = t$ corresponds to the additive mean residual life model, and $g(t) = ((t + 1)^2 - 1)/2$ corresponds to the Box-Cox transformed mean residual life model. SE, standard error estimate.