



Published in final edited form as:

*J Am Stat Assoc.* 2009 December 1; 104(487): 1168–1178.

## A Semiparametric Regression Cure Model for Interval-Censored Data

**Hao Liu [Assistant Professor]** and

Division of Biostatistics, Dan L. Duncan Cancer Center, BCM 305, Baylor College of Medicine, Houston, TX 77030, U.S.A. (E-mail: haol@bcm.edu)

**Yu Shen [Professor]**

Department of Biostatistics, M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Box 447, Houston, TX 77030, U.S.A. (Email: yshen@mdanderson.org)

### Abstract

Motivated by medical studies in which patients could be cured of disease but the disease event time may be subject to interval censoring, we presents a semiparametric non-mixture cure model for the regression analysis of interval-censored time-to-event data. We develop semiparametric maximum likelihood estimation for the model using the expectation-maximization method for interval-censored data. The maximization step for the baseline function is nonparametric and numerically challenging. We develop an efficient and numerically stable algorithm via modern convex optimization techniques, yielding a self-consistency algorithm for the maximization step. We prove the strong consistency of the maximum likelihood estimators under the Hellinger distance, which is an appropriate metric for the asymptotic property of the estimators for interval-censored data. We assess the performance of the estimators in a simulation study with small to moderate sample sizes. To illustrate the method, we also analyze a real data set from a medical study for the biochemical recurrence of prostate cancer among patients who have undergone radical prostatectomy. Supplemental materials for the computational algorithm are available online.

### Keywords

Convex optimization; Hellinger consistency; maximum likelihood estimation; primal-dual interior-point method; prostate cancer

## 1. INTRODUCTION

Medical advancements have made it possible for patients to be cured of certain types of diseases. Monitoring for a non-fatal disease event, such as cancer biochemical recurrence or AIDS drug resistance, is usually carried out through periodic clinical visits of patients (Lindsey and Ryan, 1998). The exact time of the disease event is only known to have occurred during two consecutive clinical visits if it did happen; however, a complication is that the disease event may never happen because the patient can be cured of the disease. This type of interval-censored event-time data with possibility of cure is commonly encountered in medical studies, especially in follow-up studies when the cure of non-fatal event is of clinical interest.

For example, it is of routine practice to monitor serum prostate-specific antigen (PSA) for cancer biochemical recurrence among patients who have undergone radical prostatectomy for

their clinically localized prostate cancer. Radical prostatectomy removes the entire prostate gland and presumably all tumor cells. Patients who are free of PSA for two to three years after surgery are generally considered of being cured of biochemical recurrence because the operation has successfully removed all PSA-generating tumor cells (Dilliogluligil et al., 1997; Pound et al., 1999). However, PSA may reappear in serum after surgery, which is called biochemical recurrence, indicating the operation is not success and the cancer has come back. Investigators would like to know why some patients are *cured* of the biochemical recurrence but not the others, and how these relate to patient's baseline characteristics such as cancer clinical stage and tumor histology. During the follow-up, biochemical recurrence is found out by detectable PSA level in serum as ascertained at periodic clinic visits for about every three months after surgery. The actual time of biochemical recurrence is only known to have occurred between two consecutive visits, or may not even happen for some patients who are cured of the disease. It is therefore of great interest to develop new method for analyzing the probability of cure for the interval-censored data in the regression-analysis framework for including the patient baseline characteristics.

Extensive statistical literature has been devoted to interval-censored data, but without incorporating the possibility of cure. The nonparametric maximum likelihood estimator (NPMLE) for the distribution function for interval-censored data was first studied by Peto (1973) and then by Turnbull (1976) using a self-consistency algorithm. Groeneboom and Wellner (1992) proposed an iterative convex minorant algorithm for current status data and case-2 interval-censored data. Wellner and Zhan (1997) discussed some efficient algorithms for obtaining the NPMLE for these cases. Strong consistency of the NPMLE was proved for interval-censored data under a  $L_1$  topology (Schick and Yu, 2000; Yu et al., 2001) and under the Hellinger metric (van der Vaart and Wellner, 2000). Semiparametric regression methods for interval-censored data have been investigated for the proportional hazards model (Finkelstein, 1986), the proportional odds model (Rabinowitz et al., 1995), and the accelerated failure time model (Betensky et al., 2001). Semiparametric models have also been proposed for regression analysis of current status data (Huang, 1996; Rossini and Tsiatis, 1996; Lin et al., 1998; Tian and Cai, 2006). In addition, Lam and Xue (2005) considered a mixture cure model for current status data using the estimating equation approach.

Two major approaches to model the cure rate are the mixture cure model and the non-mixture cure model (Chen et al., 1999). The mixture cure model is a mixture of two separate regression models for the cure rate of the cured population and the survival function for the non-cured population, which has been investigated extensively in the literature (Berkson and Gage, 1952; Yamaguchi, 1992; Taylor, 1995; Maller and Zhou, 1996; Sy and Taylor, 2000; Li et al., 2007, among others). Chen, Ibrahim and Sinha (1999) pointed out that the mixture cure model did not maintain the assumed survival model structure, e.g. proportional hazards, for the whole population. As an alternative, the non-mixture cure model was proposed to keep the proportional hazards structure for the whole population, while allowing for a straight-forward interpretation of the covariate effects on the probability of cure (Tsodikov, 1998; Chen et al., 1999; Tsodikov et al., 2003). As described by Chen et al. (1999) and Zeng et al. (2006), the non-mixture cure model has a biological interpretation of tumor cell growth, for which the model is also called the *promotion time cure model*. The works for both the mixture and the non-mixture cure models are largely based on right-censored data, not on interval-censored data.

In this article, we consider the regression analysis of interval-censored data for a semiparametric non-mixture cure model. Let  $T$  be a nonnegative random variable representing the event time and let  $\mathbf{Z}$  be the vector of covariates. The survival function of  $T$  conditional on the covariate  $\mathbf{Z}$  is assumed to satisfy the following cure model:

$$S(t|Z) = P(T \geq t|Z) = \exp\left(-e^{\alpha + \beta'Z} F(t)\right), \quad (1)$$

where  $(\alpha, \beta)$  are the regression coefficients, and  $F$  is a completely unspecified cumulative distribution function. For right-censored data, Tsodikov (1998) considered a cure model with a bounded function  $F$ , where the intercept is not allowed. The regression model (1) is a natural structure to keep the proportional hazards assumption while allowing a nonzero probability of cure as  $t \rightarrow \infty$ . When  $F$  is an unbounded cumulative hazard function, the model becomes the Cox proportional hazards model by setting  $\alpha = 0$ . Letting  $F$  be bounded by 1, we will show that the model (1) is identifiable and leads to an improper survival function.

The paper is organized as follows. In Section 2, we describe the model, the identifiability issue and the likelihood function. In Section 3, we develop the maximum likelihood estimation (MLE) based on an expectation-maximization (EM) approach for interval-censored data. The computation utilizes the expectation conditional-maximization (ECM) algorithm of Meng and Rubin (1993), which alternates the M-step for  $(\alpha, \beta)$  and the M-step for  $F$  in addition to the E-step. The M-step for  $F$  is the nonparametric MLE, which imposes challenge for numerical calculation. We apply a convex optimization method to develop a self-consistency-type algorithm in the spirit of Turnbull (1976). In Section 4, we prove strong consistency of the MLE estimators under the Hellinger distance using the modern empirical processes of van der Vaart and Wellner (1996). In Section 5, we describe the simulation studies to evaluate the performance of the proposed method, and illustrate the method through analysis of data for serum PSA recurrence among prostate cancer patients. We provide concluding remarks in Section 6 and the technical proofs in Supplemental Materials.

## 2. MODEL AND LIKELIHOOD

### 2.1. The Model

Under the non-mixture cure model (1), the overall survival distribution falls in the framework of a proportional hazards model with the cumulative hazard function  $e^{\alpha}F(\cdot)$ . The intercept  $\alpha$  in the regression part of the cure model allows for the easy interpretation of covariate effects while keeping a flexible baseline function. When  $t \rightarrow \infty$ , the probability of cure for a given  $Z$  is  $\exp(-e^{\alpha + \beta'Z})$ , the complimentary log-log regression. The proposed method can be easily extended to a general regression model by replacing  $\exp(\cdot)$  with a known positive convex function. When  $Z = \mathbf{0}$ , the baseline cure rate is simply  $\exp(-e^{\alpha})$ , which depends on the intercept  $\alpha$ .

Let  $E_0$  denote the expectation under the true values  $\alpha_0, \beta_0$  and  $F_0$  that satisfy model (1). We assume that  $E_0[\exp(\beta_0'Z)] < \infty$  and  $|\alpha_0| < \infty$ . It is clear that the baseline survival function  $S_0(t) = \exp(-F_0(t))$  is not a proper survival function because  $\lim_{t \rightarrow \infty} S_0(t) = 1/e$ . The marginal survival function  $P_0(T \geq t) = E_0[P_0(T \geq t|Z)]$  is also not a proper survival function by Jensen's inequality  $\lim_{t \rightarrow \infty} P_0(T \geq t) = E_0[\exp(-\exp(\alpha_0 + \beta_0'Z))] \geq \exp\{-E[\exp(\alpha_0 + \beta_0'Z)]\}$ , always greater than 0.

In order to ensure the identifiability of model (1), we require  $F$  to be a regular distribution function with  $\lim_{t \rightarrow \infty} F(t) = 1$ . To show the identifiability, let  $(\alpha_*, \beta_*, F_*)$  be a set of parameters

that satisfies (1), i.e.,  $S_*(t|Z) = \exp[-e^{\alpha_* + \beta_*'Z} F_*(t)]$ . Fixing  $\omega$  in a set of probability 1 in the underlying probability space, then  $S_*(t|Z(\omega)) = S(t|Z(\omega))$  implies that for each  $t \in (0, \infty)$ ,

$e^{\alpha_* + \beta_*'Z(\omega)} F_*(t) = e^{\alpha + \beta'Z(\omega)} F(t)$ . Letting  $t \rightarrow \infty$  yields  $e^{\alpha_* + \beta_*'Z(\omega)} = e^{\alpha + \beta'Z(\omega)}$  because  $\lim_{t \rightarrow \infty}$

$F_*(t) = \lim_{t \rightarrow \infty} F(t) = 1$ . This implies that  $\alpha_* + \beta_*' \mathbf{Z} = \alpha + \beta' \mathbf{Z}$  with probability 1. Suppose that if  $P(\mathbf{b}' \mathbf{Z} = a) = 1$  for a real number  $a$  and any real vector  $\mathbf{b}$ , then  $a = 0$  and  $\mathbf{b} = \mathbf{0}$ . This implies that  $\alpha_* = \alpha$  and  $\beta_*$  must be equal to  $\beta$ . Hence,  $F_*(t) = F(t)$  for each  $t \in (0, \infty)$ .

## 2.2. Data and Likelihood

We consider a setting in which the event time  $T$  may not be observed exactly, but instead is known to have occurred in an interval  $[L, R]$ . Here,  $L$  is the latest examination time before the event, and  $R$  is the earliest examination time after the event, where  $R = \infty$  if the event has not happened before the last follow-up. The data are  $n$  i.i.d copies of  $(L, R, \mathbf{Z})$ , denoted by  $(L_i, R_i, \mathbf{Z}_i)$  for patient  $i$ , where  $L_i < R_i$  for  $i = 1, \dots, n$ . Denote  $\boldsymbol{\theta}' = (\alpha, \beta')$  and  $\tilde{\mathbf{Z}}_i' = (1, \mathbf{Z}_i')$ . Then model (1) can be re-expressed as  $S_{\mathbf{Z}_i}(t) = S_{\boldsymbol{\theta}, F}(t|\mathbf{Z}_i) = \exp(-e^{\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i} F(t))$ , where the function  $S_{\mathbf{Z}_i}(t)$  is left continuous, with right limits. If  $R_i < \infty$ ,

$$P_{\boldsymbol{\theta}, F}(L_i \leq T_i \leq R_i | \mathbf{Z}_i) = \exp\left(-e^{\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i} F(L_i-)\right) - \exp\left(-e^{\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i} F(R_i+)\right),$$

where  $P_{\boldsymbol{\theta}, F}$  is the probability measure under the parameters  $\boldsymbol{\theta}$  and  $F$ . If  $R_i = \infty$ , then there are two possibilities for the  $i$ th patient: the patient is cured, or the event of interest for that patient occurs after the last examination time, which can be calculated by

$$P_{\boldsymbol{\theta}, F}(L_i \leq T_i \leq \infty | \mathbf{Z}_i) = e^{-\exp(\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i) F(L_i-)}.$$

The likelihood function for the  $n$  observed interval-censored data is then

$$L_n(\boldsymbol{\theta}, F) = \prod_{i=1}^n \left[ e^{-\exp(\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i) F(L_i-)} - e^{-\exp(\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i) F(R_i+)} \right]^{1(R_i < \infty)} \left[ e^{-\exp(\boldsymbol{\theta}' \tilde{\mathbf{Z}}_i) F(L_i-)} \right]^{1(R_i = \infty)}. \quad (2)$$

Note that unlike the standard Cox model, the *full* likelihood function derived under the cure model (1) needs to account for the probability of cure.

To obtain the maximum likelihood estimators (MLE) based on (2) for  $(\boldsymbol{\theta}, F)$ , we now show that the nonparametric MLE for  $F$  is only unique up to an equivalent class. This derivation is an extension of the method of Peto (1973) and Turnbull (1976) for the NPMLE of interval-censored data without covariates. Define a finite number of disjoint intervals  $\{[s_j, r_j]\}_{j=1}^{m+1}$  constructed as follows:  $s_j \in \{L_i : i = 1, \dots, n\}$  and  $r_j \in \{R_i : i = 1, \dots, n\}$ , the interval  $(s_j, r_j)$  does not contain any members of  $\{L_i, R_i, i = 1, \dots, n\}$ , and  $s_1 \leq r_1 < s_2 \leq r_2 < \dots < s_m \leq r_m < s_{m+1} < r_{m+1} = \infty$ . It is possible that  $s_j = r_j, j = 1, \dots, m$ . Here  $s_{m+1}$  is the observed longest follow-up time. Let  $\mathcal{C} = \bigcup_{j=1}^m [s_j, r_j]$ . Note that function  $x \mapsto e^{-\exp(\boldsymbol{\theta}' \mathbf{Z})x}$  is nonincreasing in  $x$ .

We first examine the maximization of the likelihood function with respect to  $F$  for an arbitrary but fixed  $\boldsymbol{\theta}$ . For given data and fixed  $\boldsymbol{\theta}$ , the maximization of the likelihood function (2) depends on  $F$  only through its values on  $\mathcal{C}$  but not on the outside of  $\mathcal{C}$ . For fixed values of

$\{F(s_j-), F(r_j+)\}_{j=1}^m$ , the likelihood function does not change with the value of  $F$  in  $(s_j, r_j)$ . Let  $p_j = F(r_j+) - F(s_j-)$  for  $j = 1, \dots, m$ . Then the vector  $\mathbf{p} \equiv (p_1, \dots, p_m)$  where

$\sum_{j=1}^m p_j = 1$  and  $p_j \geq 0$  defines an equivalence class of cumulative distribution functions that are constant outside  $\mathcal{C}$ . Thus, we can limit our search for the MLE of  $F$  in the equivalence class of the following step functions that have right-hand limits and are continuous from the left:

$$t \mapsto \sum_{j=1}^m p_j 1(t \geq r_j),$$

, with the constraint  $\sum_{j=1}^m p_j = 1$ . Without loss of generality, we still denote the function by  $F$  in the equivalence class. The function is continuous from the right  $F(t) = F(t+)$  and takes a constant value on  $[r_j, s_{j+1}]$ , which is equal to  $F(r_j)$  for  $j = 1, \dots, m$ . Further,  $F(0) = 0$  and  $F(r_m) = F(s_{m+1}) = 1$ . The function  $F$  has jumps of size  $p_1, \dots, p_m$  right before  $r_1, \dots, r_m$ , respectively. Subsequently, the survival function of  $T$  given  $Z_i$  can be written as:

$$S_{Z_i}(t) = S_{\theta, F}(t|Z_i) = \exp\left(-e^{\theta' Z_i} F(t)\right) = \exp\left[-e^{\theta' Z_i} \sum_{j=1}^m p_j 1(t \geq r_j)\right]. \tag{3}$$

In particular, for  $j = 1, \dots, m$ ,  $S_{Z_i}(s_j^-) = \exp[-e^{\theta' Z_i}(p_0 + p_1 + \dots + p_{j-1})]$ , and

$$S_{Z_i}(r_j+) = \exp\left[-e^{\theta' Z_i}(p_0 + p_1 + \dots + p_j)\right] = S_{Z_i}(s_{j+1}),$$

where  $p_0 \equiv 0$  for the convenience of notation.

The likelihood function (2) of the observed data is now a function of the parameters  $\mathbf{p} = (p_1, \dots, p_m)$  and  $\theta$ . Let  $\delta_{ij}$  indicate whether  $[s_j, r_j]$  belongs to  $[L_i, R_i]$  for  $j = 1, \dots, m+1$ . In particular,  $\delta_{i, m+1} = 1$  indicates whether the  $i$ th person has completed the follow-up without the occurrence of the event (i.e.  $R_i = \infty$ ). Using the above notation, we write the log-likelihood function of (2) as

$$l_n(\theta, F) \equiv l_n(\theta, \mathbf{p}) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^m \delta_{ij} (S_{Z_i}(s_j^-) - S_{Z_i}(r_j+)) + \delta_{i, m+1} S_{Z_i}(s_{m+1}) \right\}. \tag{4}$$

Since we assume a cure in our application, the data have at least one event-free subject ( $R_i = \infty$ ) whose follow-up time is longer than  $r_m$ . It follows that the term  $\delta_{i, m+1} S_{Z_i}(s_{m+1})$  does not vanish from (4). The likelihood contribution of event-free subjects is determined by comparing their follow-up times with  $r_m$ . Only the subjects who are event-free and have  $L_i > r_m$  contribute to the likelihood as the cured subjects. The random times  $\{r_m, s_{m+1}\}$  essentially serve the same role as the deterministic ‘‘cure threshold’’ proposed by Zeng, Yin, and Ibrahim (2006) for right-censored data. As with all cure modeling, the viability of this model requires the follow-up of a study to be sufficiently long so that  $s_{m+1}$  is large enough for some subjects to be considered cured in the given scientific context.

Due to the complexity of data notation, we use a hypothetical data example with 4 subjects to illustrate the derivation of the equivalent class for the likelihood function as depicted in Figure 1. The cases for patients  $i = 1, 2, 3$  are straightforward. We only describe the case for patient  $i = 4$ , where  $\delta_{4,1} = \delta_{4,2} = \delta_{4,3} = 1$ . The contribution of patient  $i = 4$  to the log-likelihood function (4) is  $\sum_{j=1}^2 \delta_{4,j} (S_{Z_4}(s_j^-) - S_{Z_4}(r_j+)) + \delta_{4,3} S_{Z_4}(s_3)$  which is equal to  $S_{Z_4}(s_1^-)$  in the equivalent class for the likelihood function (2).

In summary, we have derived the nonparametric MLE  $\hat{F}_n(\boldsymbol{\theta}) = \operatorname{argmax}_F l_n(\boldsymbol{\theta}, F)$  by an equivalence class argument for a fixed but arbitrary  $\boldsymbol{\theta}$ . The MLE for  $\boldsymbol{\theta}$  can then be obtained by maximizing the profile log-likelihood  $\boldsymbol{\theta} \mapsto l_n(\boldsymbol{\theta}, \hat{F}_n(\boldsymbol{\theta}))$ , which is equivalent to maximizing the function  $\boldsymbol{\theta} \mapsto \sup_F L_n(\boldsymbol{\theta}, F)$ . Maximizing the profile log-likelihood yields the semiparametric maximum likelihood estimator  $(\hat{\boldsymbol{\theta}}_n, \hat{F}_n)$ .

### 3. COMPUTATIONAL ALGORITHM AND PROPERTIES

We now develop the computational algorithm to obtain the MLE from (4). It is numerically difficult to maximize the log-likelihood function (4), while the dimension of  $\hat{F}_n$  can increase with the sample size  $n$ . To overcome this difficulty, we first consider an EM method for interval-censored data as an extension of the method for the nonparametric estimation when the covariates are absent (Kalbeisch and Prentice, 2002, page 80–81). The resulting expected log-likelihood function has the appealing property of being concave. We then replace the M-step by two conditional M-steps that alternate the M-step for  $\boldsymbol{\theta}$  and the M-step for  $\boldsymbol{p}$ , which allow us to utilize different optimization techniques for maximizing the expected log-likelihood function for  $\boldsymbol{\theta}$  and  $\boldsymbol{p}$ , separately. This is essentially the ECM algorithm of Meng and Rubin (1993), but we further exploit convex optimization techniques for a numerically stable and efficient algorithm for the M-step. Our algorithm forces the search of  $\boldsymbol{p}$  in the interior of a constrained convex parameter space for obtaining MLE.

#### 3.1. EM Method: Expectation-step

To simplify the notation in (4), for  $i = 1, \dots, n, j = 1, \dots, m$ , we define

$$\begin{aligned} \gamma_{i,j}(\boldsymbol{\theta}, \boldsymbol{p}) &\equiv S_{Z_i}(s_{j-}) - S_{Z_i}(r_{j+}) = S_{Z_i}(s_j) \left[ 1 - \exp\left(-p_j e^{\boldsymbol{\theta}' \boldsymbol{Z}_i}\right) \right] > 0, \\ \gamma_{i,m+1}(\boldsymbol{\theta}, \boldsymbol{p}) &\equiv \gamma_{i,m+1}(\boldsymbol{\theta}) = S_{Z_i}(s_{m+1}) = \exp\left(-e^{\boldsymbol{\theta}' \boldsymbol{Z}_i}\right). \end{aligned}$$

The log likelihood function (4) is then

$$l_n(\boldsymbol{\theta}, \boldsymbol{p}) = \sum_{i=1}^n \log \left( \sum_{j=1}^{m+1} \delta_{ij} \gamma_{ij}(\boldsymbol{\theta}, \boldsymbol{p}) \right), \tag{5}$$

where  $\sum_{j=1}^{m+1} \gamma_{i,j}(\boldsymbol{\theta}, \boldsymbol{p}) = 1$  for each  $i = 1, \dots, n$ .

By the equivalence of (2) and (4), the interval-censored data are now represented by  $\boldsymbol{\Delta}_i = (\delta_{i,1}, \dots, \delta_{i,m}, \delta_{i,m+1})$  for  $i = 1, \dots, n$ . We may view  $\boldsymbol{\Delta}_i$  as the incomplete data arising from an  $(m+1)$ -class multinomial distribution with index 1. Specifically, we consider the complete data as denoted by  $X = \{\boldsymbol{X}_i \equiv (X_{i,1}, \dots, X_{i,m}, X_{i,m+1}), \boldsymbol{Z}_i : i = 1, \dots, n\}$ . Conditional on the covariate  $\boldsymbol{Z}_i$ ,  $\boldsymbol{X}_i$  is assumed to have an  $m+1$  class multinomial distribution with index 1, and for  $j = 1, \dots, m+1$ ,

$$\Pr(X_{ij}=1|\boldsymbol{Z}_i) = \gamma_{ij}(\boldsymbol{\theta}, \boldsymbol{p}) > 0.$$

The  $\boldsymbol{X}_i$  for  $i = 1, \dots, n$  are assumed to be independent of each other so the complete data log-likelihood is



$$l_x(\theta, \mathbf{p}) = \sum_{i=1}^n \left\{ \sum_{j=1}^m X_{ij} \left[ \log S_{z_i}(s_j) + \log \left( 1 - \exp(-p_j e^{\theta' \tilde{z}_i}) \right) \right] - X_{i,m+1} e^{\theta' \tilde{z}_i} \right\}. \tag{6}$$

From the incomplete data  $\{\Delta_i, i = 1, \dots, n\}$ , we know that the  $i$ th outcome falls into one of the classes  $j$  as indexed by  $\delta_{ij}$  but we do not know the specific class into which it actually falls, except for the trivial cases. For example, in Figure (1), the event for patient 4 is known to have occurred in either  $(s_1, r_1)$ ,  $(s_2, r_2)$  or  $(s_3, r_3)$ , but in exactly which one is unknown. Under this missing mechanism, the log-likelihood of the incomplete data is identical to (4).

The E-step is to calculate the expected log-likelihood function of (6). Given data  $\{\Delta_i, \mathbf{Z}_i : i = 1, \dots, n\}$  and current parameter estimation  $(\theta^{(k)}, \mathbf{p}^{(k)})$  of the  $k$ -th step,

$$X_{ij}^{(k)} \equiv E[X_{ij} | \theta^{(k)}, \mathbf{p}^{(k)}, \Delta_i, \mathbf{Z}_i] = \frac{\delta_{ij} \gamma_{ij}(\theta^{(k)}, \mathbf{p}^{(k)})}{\sum_{j'=1}^{m+1} \delta_{ij'} \gamma_{ij'}(\theta^{(k)}, \mathbf{p}^{(k)})} \geq 0, \quad j=1, \dots, m+1. \tag{7}$$

Then the expected log-likelihood at the  $k$ th step is

$$\begin{aligned} l^{(k)}(\theta, \mathbf{p}) &\equiv E(l_x(\theta, \mathbf{p}) | \theta^{(k)}, \mathbf{p}^{(k)}, \Delta_i, \mathbf{Z}_i, i=1, \dots, n) \\ &= \sum_{i=1}^n \left\{ \sum_{j=1}^m X_{ij}^{(k)} \left[ -e^{\theta' \tilde{z}_i} \sum_{j'=1}^m p_{j'} 1(s_j \leq s_{j-1}) + \log \left( 1 - \exp(-p_j e^{\theta' \tilde{z}_i}) \right) \right] \right. \\ &\quad \left. - X_{i,m+1}^{(k)} e^{\theta' \tilde{z}_i} \right\}, \end{aligned}$$

where  $s_0$  is set to be  $-\infty$  for notational convenience. Denote

$c_{ij}^{(k)} = \sum_{j'=1}^m X_{ij'}^{(k)} 1(s_j \leq s_{j'-1}) \geq 0$ , which is independent of the parameters  $\theta$  and  $\mathbf{p}$ . The expected log-likelihood can be expressed as

$$l^{(k)}(\theta, \mathbf{p}) = \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[ -c_{ij}^{(k)} p_j e^{\theta' \tilde{z}_i} + X_{ij}^{(k)} \log \left( 1 - \exp(-p_j e^{\theta' \tilde{z}_i}) \right) \right] - X_{i,m+1}^{(k)} e^{\theta' \tilde{z}_i} \right\}. \tag{8}$$

The following proposition shows that the expected log-likelihood is a concave function with respect to both  $\theta$  and  $p_j > 0, j = 1, \dots, m$ , which implies that any local maximizer of the function is also a global maximizer on the set defined by  $\sum_{j=1}^m p_j = 1, p_j > 0, j = 1, \dots, m$  and  $\theta$  in an open convex set.

**Proposition 1**—The function  $l^{(k)}(\theta, \mathbf{p})$  is a concave function for each  $p_j > 0, j = 1, \dots, m$ , and for each  $\theta$  that belongs to an open compact set.

**Proof:** The function  $l^{(k)}(\theta, \mathbf{p})$  is a summation of functions of  $(\theta, \mathbf{p}) \mapsto -p_j e^{\theta' \tilde{z}_i}$  and  $(\theta, \mathbf{p}) \mapsto \log(1 - \exp(-p_j e^{\theta' \tilde{z}_i}))$ , which are concave by noting that  $(x, y) \mapsto -x \exp(y)$  is concave and  $(x, y) \mapsto \log(1 - \exp(-xe^y))$  is strictly concave for  $x > 0$  and  $-\infty < y < \infty$ , respectively. A summation of concave functions is still a concave function.

The M-step will be followed to maximize the expected log-likelihood function using the ECM algorithm: the maximization with respect to  $\theta$  for fixed  $p$ , and the maximization with respect to  $p$  for fixed  $\theta$ , iteratively.

### 3.2. Maximization Step for Covariate Coefficients $\theta$

The number of covariates is often much smaller than  $n$ , so it is straightforward to maximize  $l^{(k)}(\theta, p)$  directly with respect to  $\theta$  for a fixed  $p$ , for example, by a quasi-Newton method such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Boyd and Vandenberghe, 2004, Chapter 9). The first partial derivative of  $l^{(k)}(\theta, p)$  with respect to  $\theta$  is

$$\nabla_{\theta} l^{(k)}(\theta, p) = \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[ -c_{ij}^{(k)} + \frac{X_{ij}^{(k)}}{\exp(p_j e^{\theta' \tilde{Z}_i}) - 1} \right] p_j - X_{i,m+1}^{(k)} \right\} \tilde{Z}_i e^{\theta' \tilde{Z}_i}.$$

The second partial derivative of  $l^{(k)}(\theta, p)$  with respect to  $\theta$  is

$$\nabla_{\theta}^2 l^{(k)}(\theta, p) = - \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[ c_{ij}^{(k)} + \frac{X_{ij}^{(k)} \left[ 1 + \exp(p_j e^{\theta' \tilde{Z}_i}) (p_j e^{\theta' \tilde{Z}_i} - 1) \right]}{[\exp(p_j e^{\theta' \tilde{Z}_i}) - 1]^2} \right] p_j + X_{i,m+1}^{(k)} \right\} \tilde{Z}_i \tilde{Z}_i' e^{\theta' \tilde{Z}_i}.$$

We may apply the inequality  $1 - x < e^{-x}$  for  $x < 1$  to confirm that the gradient is positive definite whenever the covariate matrix  $Z_i Z_i'$  is positive definite. For every  $i, j$ , the second term in the above display is always nonnegative when  $p_j e^{\theta' \tilde{Z}_i} \geq 1$ . When  $p_j e^{\theta' \tilde{Z}_i} < 1$ , this term is still positive by noting that  $p_j e^{\theta' \tilde{Z}_i} - 1 > -\exp(-p_j e^{\theta' \tilde{Z}_i})$ . Proposition 1 also infers that for each fixed  $p$ ,  $l^{(k)}(\theta, p)$  is strictly concave when the matrix  $n^{-1} \sum_{i=1}^n Z_i Z_i'$  is positive definite and  $\theta$  is in a bounded convex set  $\mathcal{B}$  in  $\mathbb{R}^{d+1}$ . The concavity implies that a local maximizer is also the unique global maximizer.

### 3.3. Maximization Step for $F$ : Optimality Conditions

The next step in the ECM algorithm is to maximize the expected log-likelihood function with respect to  $p$  conditional on  $\theta$ . Although the maximization in  $p$  is numerically challenging, we note that the function  $p \mapsto l^{(k)}(\theta, p)$  is strictly concave by Proposition 1 when not all  $p_j \in [0, 1]$  are zero for  $j = 1, \dots, m$ , and not all  $X_{ij}^{(m)}$  are zero for  $j = 1, \dots, m$ . Thus, if we can find a local maximizer for  $p$ , it is also the unique global maximizer for  $p$  in the conditional maximization step. We first explore the optimality conditions for the maximization step for  $p$  by convex optimization techniques (Boyd and Vandenberghe, 2004, Section 5.5). To develop a numerically efficient and stable algorithm, we apply the primal-dual interior-point method for constrained optimization (Boyd and Vandenberghe, 2004, page 609–613), which yields a self-consistency algorithm for  $p$  in the sense of Turnbull (1976) and keeps the point mass  $p_j$  positive.

For notational convenience, denote for  $j = 1, \dots, m$ ,

$$a_j = \sum_{i=1}^n \left[ e^{\theta' \tilde{Z}_i} \sum_{j'=1}^m X_{ij'}^{(k)} 1(s_j \leq s_{j'-1}) \right],$$



where  $X_{ij}^{(k)}$  is defined in (7) and  $k$  is the index for the current EM iteration. Here both  $\theta$  and  $k$  are fixed, so  $a_j, e^{\theta' \mathbf{Z}_i}$  and  $X_{ij}^{(k)}, i=1, \dots, n, j=1, \dots, m$ , are positive constants in the maximization step for  $F$ . Denote  $\mathbf{a} = (a_1, \dots, a_m)'$ . Note that the last term in (8) is not a function of  $\mathbf{p}$ , so the part in the expected log-likelihood that involves  $\mathbf{p}$  is

$$f(\mathbf{p}) \equiv -\mathbf{a}^T \mathbf{p} + \sum_{i=1}^n \sum_{j=1}^m X_{ij}^{(k)} \log \left[ 1 - \exp(-e^{\theta' \mathbf{Z}_i} p_j) \right],$$

with the constraint that  $\sum_{j=1}^m p_j = 1$  and  $0 \leq p_j \leq 1, j = 1, \dots, m$ . Because  $f(\mathbf{p})$  is strictly concave in  $\mathbf{p}$ , the maximization of  $f$  becomes a concave optimization problem with constraints:

$$\max f(\mathbf{p}) \quad \text{subject to} \quad \sum_{j=1}^m p_j = 1, \text{ and } p_j \geq 0, \quad \text{for } j=1, \dots, m. \tag{9}$$

The first-order partial derivative of  $f$  with respect to  $p_j$  is given by

$$\dot{f}_j(\mathbf{p}) = \frac{\partial}{\partial p_j} f(\mathbf{p}) = -a_j + \sum_{i=1}^n \frac{X_{ij}^{(k)} e^{\theta' \mathbf{Z}_i}}{\exp(e^{\theta' \mathbf{Z}_i} p_j) - 1} \quad \text{for } j=1, \dots, m.$$

The second-order partial derivative of  $f$  with respect to  $p_j$  is given by

$$\ddot{f}_j(\mathbf{p}) = \frac{\partial^2}{\partial p_j^2} f(\mathbf{p}) = - \sum_{i=1}^n \frac{X_{ij}^{(k)} e^{2\theta' \mathbf{Z}_i} \exp(e^{\theta' \mathbf{Z}_i} p_j)}{(\exp(e^{\theta' \mathbf{Z}_i} p_j) - 1)^2} \quad \text{for } j=1, \dots, m.$$

The second-order partial derivative of  $f$  with respect to  $p_j$  and  $p_{j'}$  is zero for  $j \neq j', j, j' = 1, \dots, m$ , i.e.,  $\partial^2 / \partial p_j \partial p_{j'} f(\mathbf{p}) = 0$ . The Hessian matrix  $\nabla^2 f$  is thus a diagonal matrix, which we will exploit to derive a simple and efficient algorithm.

First we characterize the optimal solution of the concave function with the Lagrangian function. Let  $v$  and  $\tau_j, j = 1, \dots, m$  be the Lagrange multipliers for the equality constraint and the inequality constraint, respectively. Let  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$  and  $\mathbf{1} = (1, \dots, 1)^T$ . The Lagrange function with parameters  $\boldsymbol{\tau}$  and  $v$  for the constrained maximization problem (9) is  $\Gamma(\mathbf{p}, \boldsymbol{\tau}, v) = f(\mathbf{p}) + \boldsymbol{\tau}^T \mathbf{p} + v(\mathbf{1} - \mathbf{1}^T \mathbf{p})$ . For a concave optimization, the necessary and sufficient conditions for a point  $\mathbf{p}^*$  to be the optimal solution are the Karush-Kuhn-Tucker (KKT) conditions (Boyd and Vandenberghe, 2004, page 243–244). Specifically, a point  $\mathbf{p}^*$  in  $[0, 1]^m, \boldsymbol{\tau}^* = (\tau_1^*, \dots, \tau_m^*)$  in  $\mathbb{R}^m$  and  $v^*$  in  $\mathbb{R}$  is an optimal solution if and only if

$$\begin{aligned} \sum_{j=1}^m p_j^* &= 1, \quad p_j^* \geq 0, \quad \tau_j^* \geq 0, \quad \tau_j^* p_j^* = 0, \\ \dot{f}_j(\mathbf{p}^*) + \tau_j^* - v^* &= 0, \quad j=1, \dots, m. \end{aligned} \tag{10}$$

The maximization problem now is equivalent to solving the KKT conditions to find the solution  $(\mathbf{p}^*, \boldsymbol{\tau}^*, v^*)$  from (10). The nonlinear equations in (10) have a simple form because of the diagonal Hessian matrix. Eliminating  $\tau_j^*$  from these equations, the KKT conditions become

$$\sum_{j=1}^m p_j^* = 1, \quad p_j^* \geq 0, \quad v^* \geq \dot{f}_j(\mathbf{p}^*), \quad [v^* - \dot{f}_j(\mathbf{p}^*)] p_j^* = 0, \quad j=1, \dots, m.$$

If  $p_j^* = 0$ , then  $\dot{f}_j(\mathbf{p}^*) = \infty$ , which is impossible by the second inequality in the above display. Thus,  $p_j^* > 0$ , and the maximizers  $\mathbf{p}^*$  and  $v^*$  must satisfy:

$$\dot{f}_j(\mathbf{p}^*) - v^* = 0, \quad \sum_{j=1}^m p_j^* = 1, \quad \text{and } p_j^* > 0, \quad j=1, \dots, m.$$

This implies that  $p_j^*$  must be less than 1 because of the constraint  $\sum_{j=1}^m p_j^* = 1$ . The KKT conditions are the necessary and sufficient conditions for the maximizers to satisfy our constrained convex maximization. The sufficient condition suggests that the initial values start from the ‘‘interior’’, i.e.,  $0 < p_j < 1$  for each  $j$ . The solutions  $\mathbf{p}^*$  and  $v^*$  are found directly from the equations for the equality constraint, while the interior-point method forces the inequality constraints  $0 < p_j^* < 1, j=1, \dots, m$ . The necessary KKT conditions infer that the interior point constraints continue to hold strictly for each ECM iteration.

### 3.4. Maximization Step for $F$ : Self-Consistent Algorithm

We now describe the details of the algorithm for the maximization step of (9) using the primal-dual interior-point method, which avoids using a generic computer maximizer. The key step is to apply Newton’s method to solve the nonlinear equations (10) for the modified KKT conditions with a scalar parameter  $\eta > 0$  that controls the inequality constraints. Specifically, Newton’s method approximates the nonlinear equations by the Taylor’s expansion and solves the following equations for the Newton steps  $(\Delta \mathbf{p}, \Delta \boldsymbol{\tau}, \Delta v)$ :

$$\begin{bmatrix} \nabla^2 f(\mathbf{p}) & \mathbf{I} & -\mathbf{1} \\ \text{diag}(\boldsymbol{\tau}) & \text{diag}(\mathbf{p}) & \mathbf{0} \\ -\mathbf{1}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{p} \\ \Delta \boldsymbol{\tau} \\ \Delta v \end{bmatrix} = - \begin{bmatrix} \nabla f(\mathbf{p}) + \text{diag}(\boldsymbol{\tau}) - v \mathbf{1} \\ \text{diag}(\boldsymbol{\tau}) \mathbf{p} - (\mathbf{1}/\eta) \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \quad (11)$$

where  $\nabla^2 f(\mathbf{p})$  is the Hessian Matrix,  $\text{diag}(\mathbf{p})$  is the diagonal matrix with diagonal entries  $p_1, \dots, p_m$ ,  $\mathbf{I}$  is the identity matrix, and  $\mathbf{1}$  is the vector with all components one (Boyd and Vandenberghe, 2004, page 610).

We can solve (11) explicitly for  $(\Delta \mathbf{p}, \Delta \boldsymbol{\tau}, \Delta v)$  because the Hessian matrix  $\nabla^2 f$  is diagonal

$$\dot{f}_j(\mathbf{p}) \Delta p_j + \Delta \tau_j - \Delta v = -(\dot{f}_j(\mathbf{p}) + \tau_j - v), \quad j=1, \dots, m \quad (12)$$

$$\tau_j \Delta p_j + \Delta \tau_j p_j = -(\tau_j p_j - 1/\eta), \quad j=1, \dots, m \quad (13)$$

$$\sum_{j=1}^m \Delta p_j = \sum_{j=1}^m p_j - 1 = 0. \quad (14)$$

Denote  $\tau_j^+ = \tau_j + \Delta\tau_j$  for  $j=1, \dots, m$ , and  $v^+ = v + \Delta v$ . Equation (12) gives

$$\tau_j^+ = v^+ - \ddot{f}_j(\mathbf{p})\Delta p_j - \dot{f}_j(\mathbf{p}). \quad (15)$$

Equation (13) gives  $\tau_j\Delta p_j = -\tau_j^+ p_j + 1/\eta$ . Plugging (15) into the equation, we get

$$[\tau_j - p_j \ddot{f}_j(\mathbf{p})] \Delta p_j = -p_j v^+ + p_j \dot{f}_j(\mathbf{p}) + 1/\eta.$$

When  $\tau_j - p_j \ddot{f}_j(\mathbf{p}) \neq 0$ , we have

$$\Delta p_j = \frac{-p_j v^+ + p_j \dot{f}_j(\mathbf{p}) + 1/\eta}{\tau_j - p_j \ddot{f}_j(\mathbf{p})}. \quad (16)$$

Using equation (14),

$$\sum_{j=1}^m \frac{p_j [\dot{f}_j(\mathbf{p}) - v^+] + 1/\eta}{\tau_j - p_j \ddot{f}_j(\mathbf{p})} = 0.$$

This gives

$$v^+ = \frac{\sum_{j=1}^m [(1/\eta + p_j \dot{f}_j(\mathbf{p})) / (\tau_j - p_j \ddot{f}_j(\mathbf{p}))]}{\sum_{j=1}^m [p_j / (\tau_j - p_j \ddot{f}_j(\mathbf{p}))]},$$

provided that the denominator is not zero. The Newton step  $\Delta p_j$  is then calculated by (16).

After calculating  $\tau_j^+$  by (15), we obtain the Newton steps  $\Delta v = v^+ - v$  and  $\Delta\tau_j = \tau_j^+ - \tau_j$ . Newton's method is an iteration procedure. Given the current iteration  $(\mathbf{p}, \boldsymbol{\tau}, v)$ , the next iteration is determined by

$$p_j \leftarrow p_j + \psi \Delta p_j, \quad \tau_j \leftarrow \tau_j + \psi \Delta\tau_j, \quad v \leftarrow v + \psi \Delta v, \quad (17)$$

where  $\psi$  is calculated by a standard backtracking line search, which further controls the inequality constraints (supplemental materials available online).

Equations (17), similar in the spirit to the self-consistent algorithm of Turnbull (1976), are explicit and simple self-updating equations. Equation (15) and Equation (16) provide accurate calculations for the Newton step  $(\Delta\mathbf{p}, \Delta\boldsymbol{\tau}, \Delta v)$ . Another computational advantage of the algorithm is that it provides a numerically stable solution to the maximization problem, even if the dimension of  $p$  is large.

### 3.5. Summary of the Iterative Algorithm for MLE

We now summarize the algorithm that combines the ECM algorithm with the primaldual interior-point method for obtaining the MLE for the cure model with interval-censored data:

1. Take a random initial value  $\theta^{(0)}$  and a random initial value  $\mathbf{p}^{(0)}$  that satisfies

$$\sum_{j=1}^m p_j^{(0)} = 1 \text{ and } p_j^{(0)} > 0$$

2. CM1: Obtain  $\theta^{(k+1)}$  by maximizing  $l^{(k)}(\theta, \mathbf{p}^{(k)})$  with respect to  $\theta$ ,

$$l^{(k)}(\theta, \mathbf{p}) = \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[ -c_{ij}^{(k)} p_j e^{\theta' \tilde{Z}_i} + X_{ij}^{(k)} \log \left( 1 - \exp(-p_j e^{\theta' \tilde{Z}_i}) \right) \right] - X_{i,m+1}^{(k)} e^{\theta' \tilde{Z}_i} \right\},$$

where for  $i = 1, \dots, n$ ,  $c_{ij}^{(k)} = \sum_{j'=1}^m X_{ij'}^{(k)} 1(s_j \leq s_{j'-1})$ ,

$$\gamma_{ij}(\theta, \mathbf{p}) = S_{z_i}(s_j) \left[ 1 - \exp(-p_j e^{\theta' \tilde{Z}_i}) \right], \quad j=1, \dots, m,$$

$$\gamma_{i,m+1}(\theta, \mathbf{p}) = \exp(-e^{\theta' \tilde{Z}_i}),$$

$$S_{z_i}(s_j) = \exp \left( -e^{\theta' \tilde{Z}_i} \sum_{j'=0}^{j-1} p_{j'} \right), \quad j=1, \dots, m,$$

$$X_{ij}^{(k)} = \frac{\delta_{ij} \gamma_{ij}(\theta^{(k)}, \mathbf{p}^{(k)})}{\sum_{j'=1}^m \delta_{ij'} \gamma_{ij'}(\theta^{(k)}, \mathbf{p}^{(k)})}, \quad j=1, \dots, m+1.$$

3. CM2 (obtain  $\mathbf{p}^{(k+1)}$  by the primal-dual interior method): given  $\theta^{(k+1)}$ , define

$$b_i^{(k+1)} = \exp(\tilde{Z}_i' \theta^{(k+1)}) \text{ and update for } j = 1, \dots, m,$$

$$X_{ij}^{(k+1)} = \frac{\delta_{ij} \gamma_{ij}(\theta^{(k+1)}, \mathbf{p}^{(k)})}{\sum_{j'=1}^m \delta_{ij'} \gamma_{ij'}(\theta^{(k+1)}, \mathbf{p}^{(k)})}.$$

Start with the initial value  $\tilde{\mathbf{p}} = \mathbf{p}^{(k)}$ ; choose  $\sigma > 1$  and  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$  where  $\tau_j > 0$ . Repeat the following to update  $\tilde{\mathbf{p}}$  until convergence to get  $\mathbf{p}^{(k+1)}$ :

- a. Determine  $\eta = \sigma/\hat{\xi}$ , where  $\hat{\xi} = \sum_{j=1}^m \tilde{p}_j \tau_j / m$ .

- b. Evaluate

$$f_j(\tilde{\mathbf{p}}) = \sum_{i=1}^n b_i^{(k+1)} \left[ \frac{X_{ij}^{(k)}}{\exp(b_i^{(k+1)} \tilde{p}_j) - 1} - \sum_{j'=1}^m X_{ij'}^{(k+1)} 1(s_j \leq s_{j'-1}) \right],$$

$$\begin{aligned} \ddot{f}_j(\tilde{\boldsymbol{p}}) &= - \sum_{i=1}^n \frac{X_{ij}^{(k+1)} \exp(2\tilde{Z}_i' \theta^{(k+1)} \exp(b_i^{(k+1)} \tilde{p}_j))}{(\exp(b_i^{(k+1)} \tilde{p}_j) - 1)^2}, \\ v^+ &= \frac{\sum_{j=1}^m [(1/\eta + \tilde{p}_j \dot{f}_j(\tilde{\boldsymbol{p}})) / (\tau_j - \tilde{p}_j \dot{f}_j(\tilde{\boldsymbol{p}}))]}{\sum_{j=1}^m [\tilde{p}_j / (\tau_j^{(k)} - \tilde{p}_j \dot{f}_j(\tilde{\boldsymbol{p}}))]}, \\ \Delta p_j &= \frac{\tilde{p}_j (f_j(\tilde{\boldsymbol{p}}) - v^+)^{1/\eta}}{\tau_j - \tilde{p}_j \dot{f}_j(\tilde{\boldsymbol{p}})}, \\ \tau_j^+ &= v^+ - \dot{f}_j(\tilde{\boldsymbol{p}}) \Delta p_j - \ddot{f}_j(\tilde{\boldsymbol{p}}). \end{aligned}$$

- c. Update  $\tilde{\boldsymbol{p}}$ ,  $\boldsymbol{\tau}$  and  $v$  by (17) with the standard backtracking line search as described in supplemental materials available online.
- 4. Repeat steps 2–3 until  $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\| + \|\boldsymbol{p}^{(k+1)} - \boldsymbol{p}^{(k)}\| < \epsilon_1$  and the expected log-likelihood converges for a pre-specified small  $\epsilon_1 > 0$ .

Note that the convergence of the algorithm can be established by the general results for the constrained ECM algorithms that have been shown in Meng and Rubin (1993) and Nettleton (1999); see also Little and Rubin (2002, Theorem 8.1, page 173).

#### 4. HELLINGER CONSISTENCY OF MLE

The technical challenge for studying the strong consistency of the MLE  $(\hat{\boldsymbol{\theta}}_n, \hat{F}_n)$  for the proposed cure model comes from interval-censored data, for which a classical approach via the Kullback-Leibler information cannot be easily applied. Instead, we prove the strong consistency of  $(\hat{\boldsymbol{\theta}}_n, \hat{F}_n)$  to the true values  $\boldsymbol{\theta}_0$  and  $F_0$  under the Hellinger distance, which provides a global consistency. The Hellinger distance is an  $L_2$ -distance between the square roots of two probability densities  $q_1$  and  $q_2$ ,

$$h(q_1, q_2) = \left( \frac{1}{2} \int (\sqrt{q_1} - \sqrt{q_2})^2 dv \right)^{1/2} = \left( 1 - \int \sqrt{q_1 q_2} dv \right)^{1/2},$$

which does not depend on the dominating measure  $v$ . This is a true distance that satisfies  $h(q_1, q_2) \leq 1$  and  $h(q_1, q_2) = 0$  if and only if  $q_1 = q_2$  a.e.  $v$ . For the nonparametric estimation for interval-censored data, van der Vaart and Wellner (2000) proved the strong consistency of MLE under the Hellinger distance. We extend their method to the semiparametric MLE under the cure model for a general case of interval-censored data, for which the proof is largely based on the theory of modern empirical processes (van der Vaart and Wellner, 1996, Chapter 2.4).

We first describe the setup of the general case of interval-censored data, which is similar to these described by Schick and Yu (2000) and van der Vaart and Wellner (2000), but with a modification to allow for the possibility of cure. This setup is necessary in order to take into account of the randomness of the event inspection times, but it leads to the same likelihood function (2), for which we have developed a computationally efficient algorithm for the MLE  $(\hat{\boldsymbol{\theta}}_n, \hat{F}_n)$ .

Let  $K$  be a positive random integer that denotes the number of inspection times for one person, and  $\boldsymbol{Y}_K = \{Y_{K,1}, \dots, Y_{K,K}\}$  denote the inspection times for the event of interest, where  $Y_{K,1} < \dots < Y_{K,K}$ . In practice, it is natural to assume  $E_0(K) < \infty$ . The random observation times for one person is a triangular array  $\boldsymbol{Y} = \{Y_{k,j} : j = 1, \dots, k, k = 1, 2, \dots\}$ , where  $k$  is the realization of  $K$ . Let  $\boldsymbol{Y}_k = (Y_{k,1}, \dots, Y_{k,k})$  be the  $k$ -th row of the array  $\boldsymbol{Y}$  with the realization denoted by  $y_k = (y_{k,1}, \dots, y_{k,k})$ .

The event time  $T$  is only known to have occurred in one of the intervals  $[Y_{K,j-1}, Y_{K,j}]$ ,  $j = 1, \dots, K$ , or  $[Y_{K,K}, Y_{K,K+1}]$ , where  $Y_{K,0} \equiv 0$  and  $Y_{K,K+1} \equiv \infty$ . We denote  $\Lambda_k = (\Lambda_{k,1}, \dots, \Lambda_{k,k+1})$ , where  $\Lambda_{k,j} = 1_{[Y_{k,j-1}, Y_{k,j}]}(T)$  for  $j = 1, \dots, k$ , and  $\Lambda_{k,k+1} = 1_{[Y_{k,k}, Y_{k,k+1}]}(T)$ , with the realization denoted by  $\lambda_k = (\lambda_{k,1}, \dots, \lambda_{k,k+1})$ .

We now formulate the likelihood under this setup. Assume that conditional on  $\mathbf{Z}$ ,  $(K, \mathbf{Y})$  is independent of the event time  $T$ . We also make the noninformative inspection assumption that the distribution of  $(K, \mathbf{Y}_K, \mathbf{Z})$  does not depend on the event time  $T$  and the parameters of interest  $\theta$  and  $F$ . Given  $\mathbf{Z}$ , the conditional survival function of  $T$  follows cure model (1). Conditional on  $(K, \mathbf{Y}_K)$  and  $\mathbf{Z}$ , the random vector  $\Lambda_k$  has a multinomial distribution, i.e.,  $(\Lambda_k | K, \mathbf{Y}_K, \mathbf{Z}) \sim \text{Multinomial}_{K+1}(1, \Delta S_K(\mathbf{Z}))$ , where the vector

$$\Delta S_K(\mathbf{Z}) \equiv (1 - S(Y_{K,1} | \mathbf{Z}), \dots, S(Y_{K,K-1} | \mathbf{Z}) - S(Y_{K,K} | \mathbf{Z}), S(Y_{K,K} | \mathbf{Z})).$$

Let  $\mathbf{V} = (K, \Lambda_k, \mathbf{Y}_K, \mathbf{Z})$  with the realization denoted by  $\mathbf{v} = (k, \lambda_k, \mathbf{y}_k, \mathbf{z})$ . Denote  $\tilde{\mathbf{z}}^T = (1, \mathbf{z}^T)$ . The distribution of  $\mathbf{V}$  has a version of density  $dP_V / d\mu$  with respect to a dominating measure  $\mu$  determined by the joint distribution of  $(K, \mathbf{Y}_K, \mathbf{Z})$ , which does not depend on the parameters of interest  $\theta$  and  $F$  by the non-informative inspection assumption. A version of density of the distribution of  $\mathbf{V}$ , which is the likelihood function for one observation  $\mathbf{v}$  that we will use, is thus given by

$$p_{\theta, F}(\mathbf{v}) = \sum_{j=1}^k \lambda_{k,j} \left[ \exp\left(-e^{\theta'} \tilde{\mathbf{z}} F(y_{k,j-1})\right) - \exp\left(-e^{\theta'} \tilde{\mathbf{z}} F(y_{k,j})\right) \right] + \lambda_{k,k+1} \left[ \exp\left(-e^{\theta'} \tilde{\mathbf{z}} F(y_{k,k})\right) \right]. \tag{18}$$

The data are  $n$  i.i.d.  $\mathbf{V}_1, \dots, \mathbf{V}_n$  copies of  $\mathbf{V}$ , where  $\mathbf{V}_i = (K_i, \Lambda_{K_i}^{(i)}, \mathbf{Y}_{K_i}^{(i)}, \mathbf{Z}_i)$ ,  $i = 1, \dots, n$ . Define by  $\mathbb{P}_n g = n^{-1} \sum_{i=1}^n g(\mathbf{V}_i)$  the empirical distribution for a measurable real function  $g$  on the sample space  $\mathcal{V}$ . The log-likelihood function for  $(\theta, F)$  of  $\tilde{\mathbf{V}} = (\mathbf{V}_1, \dots, \mathbf{V}_n)$  is

$$\frac{1}{n} l_n(\theta, F | \tilde{\mathbf{V}}) \equiv \mathbb{P}_n(l_{\theta, F}(\mathbf{V})) = \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{j=1}^{K_i} \Lambda_{K_i, j}^{(i)} \log \left[ \exp\left(-e^{\theta'} \tilde{\mathbf{z}}_i F(Y_{K_i, j-1}^{(i)})\right) - \exp\left(-e^{\theta'} \tilde{\mathbf{z}}_i F(Y_{K_i, j}^{(i)})\right) \right] - \Lambda_{K_i, K_i}^{(i)} e^{\theta'} \tilde{\mathbf{z}}_i F(Y_{K_i, K_i}^{(i)}) \right\}, \tag{19}$$

where

$$l_{\theta, F}(\mathbf{V}) = \sum_{j=1}^K \Lambda_{K, j} \log \left[ \exp\left(-e^{\theta'} \tilde{\mathbf{z}} F(Y_{K, j-1})\right) - \exp\left(-e^{\theta'} \tilde{\mathbf{z}} F(Y_{K, j})\right) \right] - \Lambda_{K, K} e^{\theta'} \tilde{\mathbf{z}} F(Y_{K, K}).$$

The log-likelihood function (19) is same as the logarithm of the original likelihood function (2) by noting that  $\Lambda_{K_i, j}^{(i)} = 1_{\{Y_{K_i, j-1} \leq T_i < Y_{K_i, j}\}}$ ,  $j = 1, \dots, K_i$ , and  $\Lambda_{K_i, K_i}^{(i)} = 1_{\{Y_{K_i, K_i} \leq T_i \leq \infty\}}$ . Therefore,  $L_i = Y_{K_i, j-1}$  and  $R_i = Y_{K_i, j}$  if  $Y_{K_i, j-1} \leq T_i < Y_{K_i, j}$ , and if  $T_i \geq Y_{K_i, K_i}$ ,  $L_i = Y_{K_i, K_i}$  and  $R_i = \infty$ . This setup is flexible enough to allow a “data-generating” mechanism for the interval-censoring data  $[L_i, R_i]$ , where the observation stops after a random number  $K_i$  of inspections.



There is no need to make specific distribution assumptions about the inspection process. This setup is also convenient for the technical proof of the Hellinger consistency.

We will work with the log-likelihood function (19) to prove the Hellinger consistency, and need the following regularity conditions:

C1. Parameter  $\theta$  is restricted in a compact set  $\Theta$  in  $\mathbb{R}^{d+1}$ , and the baseline distribution function  $F$  is in the set  $\mathcal{B}$  of all sub-distribution functions on  $(0, \infty)$ .

C2.  $E_0(\|\mathbf{Z}\|_2) < \infty$  for the Euclidean norm  $\|\cdot\|_2$  in  $\mathbb{R}^p$ . For the true parameter  $\theta_0$ ,  $E(\exp(e^{\theta_0' \mathbf{Z}})) < \infty$ .

C3. The true parameter function  $F_0$  is strictly increasing, and satisfies

$$E_0 \left\{ \frac{1}{\min_{j=1, \dots, K} (F_0(Y_{K,j}) - F_0(Y_{K,j-1}))^2} | K \right\} < \infty.$$

Condition C3 is a technical condition for preventing  $E_0(1/p_{\theta_0, F_0})$  from escaping to infinity. It regulates the variability of the inspection process  $\{Y_{K,1}, \dots, Y_{K,K}\}$ . The following theorem summarizes the strong consistency property of  $(\hat{\theta}_n, \hat{F}_n)$ . The detail of the technical proofs is provided in Supplemental Materials.

### Theorem 1

Under the regularity conditions C1–C3, the maximum likelihood estimate  $\hat{\theta}_n$  and  $\hat{F}_n$  satisfies

$$h(p_{\hat{\theta}_n, \hat{F}_n}, p_{\theta_0, F_0}) \rightarrow_{a.s.} 0.$$

## 5. NUMERICAL STUDIES

### 5.1. Simulation Study

To examine the empirical properties of the proposed method, we perform a simulation study. The event time  $T$  is generated under the cure model (1) for given  $\mathbf{Z}$ ,

$$S(t|\mathbf{Z}) = \exp\left(-e^{\alpha_0 + \beta_0' \mathbf{Z}} F_0(t)\right).$$

There is a connection between the cure model (1) and the mixture cure model as pointed out by Chen et al. (1999). The overall survival function can be written in the form of the mixture cure model:

$$S(t|\mathbf{Z}) = \pi(\mathbf{Z}) + (1 - \pi(\mathbf{Z}))S^*(t|\mathbf{Z}),$$

where  $\pi(\mathbf{Z}) = \exp[-\exp(\alpha_0 + \beta_0' \mathbf{Z})]$  is the covariate-specific probability of cure, and

$$S^*(t|\mathbf{Z}) = \frac{\exp(-e^{\alpha_0 + \beta_0' \mathbf{Z}} F_0(t)) - \exp(-e^{\alpha_0 + \beta_0' \mathbf{Z}})}{1 - \exp(-e^{\alpha_0 + \beta_0' \mathbf{Z}})}.$$

The conditional survival function  $S^*$  is a proper survival function.

To simulate  $T$  under (1), we consider a setting with one covariate  $Z$ , which has a uniform distribution on  $(-1:0, 1:0)$ . The first step is to generate a random variate distributed as Bernoulli with  $\pi(Z)$  for a given single covariate  $Z$ . If it is 1, then  $T = \infty$ . If it is 0, then generate a random variate  $T^*$  distributed as the proper survival function  $S^*(t/Z)$ . To generate  $T^* \sim S^*(t/Z)$ , we apply the accept-reject method for simulating the random variate (Devroye, 1986, page 40–65). For simplicity, we choose the baseline distribution function  $F_0(t) = 1 - \exp(-t)$ .

The sample size is chosen to be 50, 100 or 200. The true value for  $\beta_0$  is chosen to be 0.0, 0.5 or 1.0. When  $\beta_0 = 0.0$ , the probability of cure is  $\exp(-e^{\alpha_0})$ . The value of  $\alpha_0$  is chosen to be either 0.0 or 1.0, which corresponds to a baseline cure rate of either 0.37 or 0.07, respectively. For small sample sizes, care has been taken to ensure that there is at least one cured patient in the simulated dataset. The examination times are generated independent of  $Z$  and  $T$ , from a homogeneous Poisson process. The inter-examination times are independent and identically distributed as exponential with mean  $\zeta$  of either 1/5 year or 1/2 year, corresponding to short or long inter-examinations, respectively. The length of study is either 3 years for short inter-examination or 5 years for long inter-examination. The simulation replicates are 1,000 for each scenario.

The results of the Monte Carlo simulations are summarized in Table 1, including the bias, the empirical estimate of the mean squared error (MSE) and its average standard error. The biases of estimates for  $\alpha$  and  $\beta$  are very small even for a small sample size, indicating that the algorithm performs surprisingly well. The empirical estimates for the standard error, as well as the mean squared error, decrease with the sample sizes. The MSE of the estimator under the long inter-examination time is slightly higher than those under the short inter-examination time.

## 5.2. Data Analysis

The proposed method was motivated by the medical study for patients with clinically localized prostate cancer, who were followed and monitored for biochemical recurrence of cancer after undergoing radical prostatectomy. As a widely used surgical procedure, radical prostatectomy takes away the entire prostate gland and surrounding tissue to remove all tumor cells, leading to a very high cure rate for early prostate cancer (Kufe et al., 2003, Chapter 111). Although no cells should remain to produce prostate-specific antigen (PSA), clinical studies have shown that about 20% to 30% of these patients could still experience biochemical recurrence of cancer, defined as a detectable serum PSA level ( $\geq 0.1$  ng/mL) after surgery. Clinical studies have also shown that the biochemical recurrence of cancer often occurs within 3 years after surgery, which is a much earlier event than the clinical recurrence of cancer (Dilliogluligil et al., 1997; Pound et al., 1999). A patient who has remained free of serum PSA three years after his surgery is generally considered to be cured of biochemical recurrence of prostate cancer. It is therefore important to know what preoperative clinical characteristics might be prognostic for the rate of cure for biochemical recurrence and how long it is likely to occur after surgery.

For this purpose, we obtained the clinical data of 260 men who underwent radical prostatectomy for clinically localized prostate cancer performed by a single surgeon at an academic hospital in Houston, Texas. The average age for the men at the operation was 59.5 (median age 59). After the operation, these men were followed up periodically for serum PSA level examination for about every 3 months. All patients were alive at their last follow-up. The longest follow-up duration  $s_{m+1}$  was 4.1 years, and all biochemical recurrences were observed within  $r_m = 3.3$  years after surgery. This indicated that the patients' follow-up times were adequate for the analysis of the cure rate for biochemical recurrence. The exact time of biochemical recurrence was only known to have occurred between the previous examination and the latest examination for a patient. The possible prognostic factors for biochemical recurrence included the

following: pathologic stage at surgery (T2, clinically organ-confined; T3, locally advanced), preoperative PSA level, Gleason score, and positive surgical margins (PSM). The preoperative PSA level was logarithmically transformed. The Gleason score graded histologically the prostate tumor cells, ranging from 2 to 10 (the higher the worse). We classified the Gleason score into 2 categories: less or equal to 7, and above 7 (the baseline). Positive surgical margins (PSM), defined as the presence of cancer cells at the inked margin of a resected tumor specimen, might indicate incomplete excision of tumor, but it was not always clear in the medical literature whether PSM was associated with biochemical recurrence. Of note, being cured of biochemical recurrence is different from being cured of prostate cancer, which may be confounded with over-diagnosis of prostate cancer.

The estimated covariate coefficients were shown in Table 2. The estimated baseline function  $\hat{F}$  was plotted in Figure 2. For computational convenience, we used the nonparametric bootstrap method to estimate the standard errors, where the replicates of the bootstrap samples were 1,000. The intercept, stage T2, and Gleason score  $\leq 7$  in the regression model were not significantly different from zero, but their magnitude were negative, indicating an increased cure rate ( $\approx 66.3\%$ ) for patients with the following clinical characteristics: tumor of stage T2, preoperative PSA level 1.0, Gleason score  $\leq 7$ , and a negative surgical margin. Only the preoperative PSA level was significantly associated with the cure of biochemical recurrence at the 0.05 level. The positive estimated coefficient indicated a lower preoperative PSA level would lead to a higher probability of cure.

## 6. DISCUSSION

We have studied the semiparametric MLE for a non-mixture cure model for interval-censored data. Besides its attractive properties and biological interpretation as explained by Tsodikov (1998) and Chen, Ibrahim and Sinha (1999), one advantage of the non-mixture cure model for interval-censored data is the parsimonious model of the covariate effects on both cure rate and time to event. This is especially appealing when the primary goal of a study is to evaluate risk factors for the cure rate. In addition, the proposed semiparametric non-mixture cure model  $S(t|\mathbf{Z}) = \exp(-\phi(\boldsymbol{\theta}'\mathbf{Z})F(t))$  is flexible enough to allow a general function  $\phi(\cdot)$  to replace  $\exp(\cdot)$  in the regression part. Our algorithm can be easily modified for the general function  $\phi$  as long as it is nonnegative and convex. It is important to assess the goodness-of-fit for the posited model (1), perhaps based on the likelihood function developed herein. Another obvious extension within the regression framework is to allow for time-dependent covariates, which may be useful for checking the proportionality assumption in model (1). Although it will not be conceptually difficult to develop the estimators, studying their properties requires substantial effort.

Interval-censored data provides unique opportunity and challenge for analyzing the cure rate in the medical studies. We developed the likelihood function through the classic equivalence class approach of Turnbull (1976). To estimate cure rate, the likelihood function requires that at least one event-free subject be followed longer than  $r_m$ . In contrast to the mixture cure model, the definition of cure fraction for the non-mixture cure model is unchanged even in the case of insufficient follow-up. However, for sound data analysis, one should always consult with clinical investigators to assess whether the value of  $r_m$  in a given data is appropriate for the scientific context.

Another major contribution in this work for interval-censored data is the development of a computational algorithm that incorporated some modern optimization techniques with the EM method for solving MLE. The maximization step is a constrained optimization that requires the sophisticated interior-point method to keep searching in the restricted parameter subspace. Combined with the theoretically sound primal-dual method, our algorithm provides a computationally robust solution for the nonparametric estimation of the function  $F$ . We further

worked out an efficient and numerically stable algorithm for the M-step for  $F$ , which is self-consistency in the sense of Turnbull (1976). Our experience is that the algorithm converges very quickly in each M-step even when the dimension of  $\boldsymbol{p}$  is large, although the convergence for the entire ECM iterations still needs extra computation time.

The assumption of an independent and non-informative inspection process is for the consistency proof. That assumption is, however, likely satisfied in our data example. This is because biochemical recurrence after radical prostatectomy has been shown to occur much earlier than symptomatic clinical recurrence (Pound et al., 1999). Therefore, it is unlikely that the inspection process would be altered among the asymptomatic patients. The asymptotic normality and efficiency of  $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$  may be obtained by deriving the efficient score function of  $\boldsymbol{\theta}$  while treating  $F$  as the nuisance function, where the detail will be reported elsewhere. We conjecture that the rate of convergence of  $\hat{\boldsymbol{\theta}}_n$  is  $n^{-1/2}$  and that of  $\hat{F}_n$  is  $n^{-1/3}$ . It is difficult to obtain the asymptotic distribution for  $n^{1/3}(\hat{F}_n - F_0)$ , while the asymptotic normality of  $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$  is relatively easy to prove by modifying the empirical process techniques for the case-2 interval-censored data as in the work of van de Geer (2000, page 228–230). The nonparametric bootstrap is a valid approach to calculating the standard error for  $\hat{\boldsymbol{\theta}}_n$ . Alternatively, the asymptotic variance-covariance matrix of estimator  $\hat{\boldsymbol{\theta}}_n$  and  $\hat{F}_n$  may be estimated by the supplemented ECM (S-ECM) algorithm of van Dyk et al. (1995), which is a computational byproduct of the ECM algorithm.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The Authors thank Dr. J. M. Levitt and Dr. K. M. Slawin for providing the data, and the associated editor and three referees for their helpful comments. The research of Hao Liu was supported in part by NIH grant P30-CA125123. The research of Yu Shen was supported in part by NIH grant R01-CA079466.

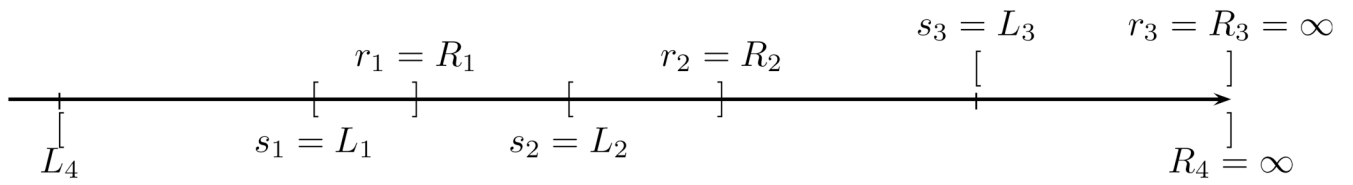
## REFERENCES

- Berkson J, Gage RP. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* 1952;47:501–515.
- Betensky RA, Rabinowitz D, Tsiatis AA. Computationally simple accelerated failure time regression for interval censored data. *Biometrika* 2001;88(3):703–711.
- Boyd, SP.; Vandenberghe, L. *Convex Optimization*. Cambridge, UK: Cambridge University Press; 2004.
- Chen M-H, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* 1999;94(447):909–919.
- Devroye, L. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag; 1986.
- Dillioglulugil O, Leibman BD, Kattan MW, Seale-Hawkins C, Wheeler TM, Scardino PT. Hazard rates for progression after radical prostatectomy for clinically localized prostate cancer. *Urology* 1997;50(1):93–99. [PubMed: 9218025]
- Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics* 1986;42(4):845–854. [PubMed: 3814726]
- Groeneboom, P.; Wellner, JA. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel: Birkhauser Verlag; 1992.
- Huang J. Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics* 1996;24(2):540–568.
- Kalbeisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. second edition. Hoboken, NJ: Wiley-Interscience; 2002.
- Kufe, DW.; Pollock, RE.; Weichselbaum, RR.; Bast, RC.; Gansler, TS.; Holland, JF.; Frei, E., editors. *Holland-Frei Cancer Medicine* 6. 6th edition. BC Decker Inc; 2003.

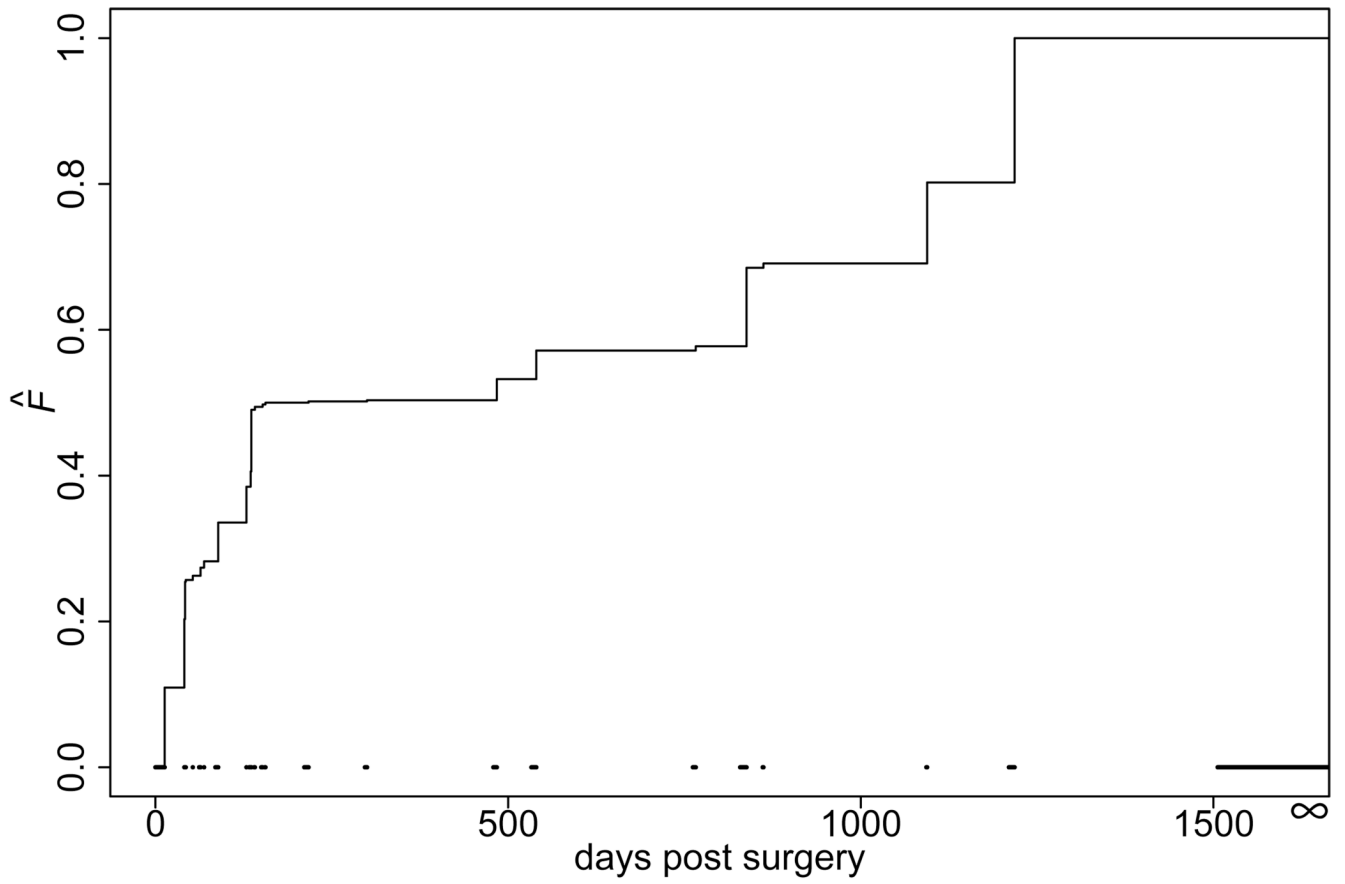
- Lam KF, Xue H. A semiparametric regression cure model with current status data. *Biometrika* 2005;92(3):573–586.
- Li Y, Tiwari RC, Guha S. Mixture cure survival models with dependent censoring. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 2007;69(3):285–306.
- Lin DY, Oakes D, Ying Z. Additive hazards regression with current status data. *Biometrika* 1998;85(2):289–298.
- Lindsey JC, Ryan LM. Tutorial in biostatistics: methods for interval-censored data. *Statistics in Medicine* 1998;17:219–238. [PubMed: 9483730]
- Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2nd ed edition. Hoboken, N.J: Wiley; 2002.
- Maller, R.; Zhou, X. *Survival Analysis with Long-term Survivors*. New York: John-Wiley; 1996.
- Meng X-L, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 1993;80(2):267–278.
- Nettleton D. Convergence properties of the em algorithm in constrained parameter spaces. *Canadian Journal of Statistics* 1999;27(3):639–648.
- Peto R. Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society. Series C. Applied Statistics* 1973;22:86–91.
- Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. *Journal of the American Medical Association* 1999;281(17):1591–1597. [PubMed: 10235151]
- Rabinowitz D, Tsiatis A, Aragon J. Regression with interval censored data. *Biometrika* 1995;82:501–513.
- Rossini AJ, Tsiatis AA. A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association* 1996;91(434):713–721.
- Schick A, Yu Q. Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics* 2000;27(1):45–55.
- Sy JP, Taylor JMG. Estimation in a cox proportional hazards cure model. *Biometrics* 2000;56(1):227–236. [PubMed: 10783800]
- Taylor JMG. Semi-parametric estimation in failure time mixture models. *Biometrics* 1995;51(3):899–907. [PubMed: 7548707]
- Tian L, Cai T. On the accelerated failure time model for current status and interval censored data. *Biometrika* 2006;93(2):329–342.
- Tsodikov A. A proportional hazards model taking account of long-term survivors. *Biometrics* 1998;25(54):1508–1516. [PubMed: 9883549]
- Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association* 2003;98(464):1063–1078.
- Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B. Methodological* 1976;38(3):290–295.
- van de Geer, SA. *Applications of Empirical Process Theory*. Cambridge, UK: Cambridge University Press; 2000.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. New York: Springer; 1996.
- van der Vaart, AW.; Wellner, JA. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In: Gine, E.; M, D.; Wellner, J., editors. *High Dimensional Probability II*. Boston: Birkhäuser; 2000. p. 113-132.
- van Dyk DA, Meng X-L, Rubin DB. Maximum likelihood estimation via the ECM algorithm: computing the asymptotic variance. *Statistica Sinica* 1995;5(1):55–75.
- Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* 1997;92(439):945–959.

- Yamaguchi K. Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in japan. *Journal of the American Statistical Association* 1992;87(418):284–292.
- Yu QQ, Wong GYC, Li LX. Asymptotic properties of self-consistent estimators with mixed interval-censored data. *Annals of the Institute of Statistical Mathematics* 2001;53(3):469–486.
- Zeng D, Yin G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association* 2006;101(474):670–684.





**Figure 1.**  
A hypothetical example with  $n = 4$ . Each pair of brackets [ ] corresponds to a patient data.



**Figure 2.**  
NPMLE of the function  $\hat{F}$ . The dashed lines at the bottom are the non-overlapping intervals  $(s_j, r_j]$  for  $j = 1, \dots, m + 1$ .

Summary of the simulation results including bias and mean squared error (MSE) for the MLE estimator  $\hat{\alpha}$  and  $\hat{\beta}$ , where  $\zeta$  is the mean inter-examination time,  $n$  is the sample size, and the true values are  $\alpha_0$  and  $\beta_0$ .

**Table 1**

$n$	$\alpha_0$	$\beta_0$	$\zeta = 0.2$						$\zeta = 0.5$											
			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\alpha}$			$\hat{\beta}$								
			Bias	MSE	MSE	Bias	MSE	MSE	Bias	MSE	MSE	Bias	MSE	MSE						
50	0.0	0.0	-0.007	0.043	0.010	0.113	0.023	0.042	-0.012	0.122	0.5	0.002	0.040	0.032	0.120	0.036	0.043	0.044	0.133	
	1.0	0.0	-0.003	0.043	0.042	0.136	0.052	0.055	0.133	0.174	1.0	0.0	0.031	0.040	0.004	0.073	0.087	0.046	-0.002	0.125
	0.5	0.5	0.042	0.042	0.021	0.092	0.110	0.055	0.113	0.152	1.0	0.030	0.048	0.037	0.093	0.116	0.069	0.165	0.151	
100	0.0	0.0	-0.009	0.019	0.011	0.053	0.023	0.019	-0.014	0.058	0.5	0.005	0.020	0.011	0.057	0.027	0.020	0.038	0.063	
	1.0	0.0	0.004	0.019	0.021	0.056	0.041	0.026	0.103	0.082	1.0	0.0	0.018	0.022	-0.007	0.034	0.067	0.023	-0.012	0.058
	0.5	0.5	0.021	0.019	0.012	0.038	0.078	0.026	0.089	0.086	1.0	0.012	0.022	0.020	0.043	0.070	0.024	0.098	0.056	
200	0.0	0.0	-0.006	0.009	0.004	0.025	0.024	0.010	-0.003	0.028	0.5	0.002	0.009	0.007	0.027	0.026	0.010	0.034	0.032	
	1.0	0.0	0.001	0.011	0.011	0.028	0.042	0.013	0.085	0.045	1.0	0.0	0.003	0.010	0.002	0.017	0.054	0.012	-0.028	0.026
	0.5	0.5	0.008	0.009	0.006	0.018	0.066	0.014	0.074	0.057	1.0	0.005	0.010	0.009	0.020	0.053	0.011	0.061	0.023	

NOTE: MSE was the empirical mean-squared error. The average standard errors of MSE for  $(\hat{\alpha}, \hat{\beta})$  were (0.036, 0.087) when  $\zeta = 0.2$ , and were (0.048, 0.128) when  $\zeta = 0.5$ .

**Table 2**

Estimates and standard error (se) for analyzing the biochemical recurrence study.

	Intercept	Stage T2	log PSA level	Gleason $\leq 7$	PSM
estimates	-0.34	-0.24	0.62	-0.31	0.06
bootstrap se	0.41	0.74	0.29	0.33	0.37

PSM: Positive surgical margins.