



Published in final edited form as:

*Lifetime Data Anal.* 2019 January ; 25(1): 26–51. doi:10.1007/s10985-018-9420-0.

## A Class of Semiparametric Cure Models with Current Status

### Data:

#### Cure Models with Current Status Data

**Guoqing Diao** and

Department of Statistics, George Mason University

**Ao Yuan**

Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University

### Abstract

Current status data occur in many biomedical studies where we only know whether the event of interest occurs before or after a particular time point. In practice, some subjects may never experience the event of interest, i.e., a certain fraction of the population is cured or is not susceptible to the event of interest. We consider a class of semiparametric transformation cure models for current status data with a survival fraction. This class includes both the proportional hazards and the proportional odds cure models as two special cases. We develop efficient likelihood-based estimation and inference procedures. We show that the maximum likelihood estimators for the regression coefficients are consistent, asymptotically normal, and asymptotically efficient. Simulation studies demonstrate that the proposed methods perform well in finite samples. For illustration, we provide an application of the models to a study on the calcification of the hydrogel intraocular lenses.

### Keywords

Box-Cox transformation; Cure fraction; Empirical process; NPML; Proportional hazards cure model; Proportional odds cure model; Semiparametric efficiency

## 1 Introduction

Current status data occur in many biomedical and health studies, where the exact onset of the event of interest cannot be observed directly and we only know whether or not the event has occurred at the time the sample is collected. For example, in a tumorigenicity experiment that involves lung tumors in mice (Hoel and Walburg, 1972), only an indicator of lung tumor presence or absence at time of death was observed. A second example is the Calcification study of the hydrogel intraocular lenses (IOL), which is an infrequently reported complication of cataract treatment (Yu et al, 2001). The exact time to the occurrence of calcification is not observable. Instead, the calcification status was observed at the time of examination.

Current status data have been investigated extensively in literature. Groeneboom and Wellner (1992) derived the nonparametric maximum likelihood estimator (NPML) of the survival function of time to event. Methods for the regression analysis of current status data

include the Cox proportional hazards model (Huang, 1996), the proportional odds model (Rossini and Tsiatis, 1996; Huang and Rossini, 1997), and the additive hazard model (Lin et al, 1998). More recently, partly linear accelerated failure time (AFT) model and partly linear transformation models were proposed by Xue et al (2004) and Ma and Kosorok (2005), respectively. There is an inexplicit assumption in all the aforementioned methods that the event will eventually occur for all subjects given a sufficient long time of follow-up.

In many studies, it is observed that a proportion of subjects never experiences the event of interest, i.e., they are cured or not susceptible to the event of interest. The aforementioned standard survival analysis techniques may no longer be appropriate in such situations. Alternatively, survival models that allow for a cured subgroup are more desirable. Such models generally are referred to as “cure models”. Under the cure model, the survival function, denoted by  $S$ , is allowed to be improper, i.e.,  $S(\infty) > 0$ . There is an abundant literature on cure rate models incorporating a cure fraction for right censored data. Perhaps the most commonly used cure rate model is the mixture model first introduced by Berkson and Gage (1952). This model assumes that the underlying population consists of a cured sub-population and a sub-population not being cured. This mixture cure model and its extensions have been studied by many investigators, including Farewell (1982, 1986), Gray and Tsiatis (1989), Kuk and Chen (1992), Taylor (1995), Sy and Taylor (2000), Peng and Dear (2000), Betensky and Schoenfeld (2001), and Lu and Ying (2004), among others. A comprehensive discussion of the mixture cure model is given in Maller and Zhou (1996).

The mixture cure model, however, does have several drawbacks both from a frequentist and Bayesian perspective as noted by Chen et al (1999) and Ibrahim et al (2001). One particular drawback is that the mixture cure model does not appear to describe the underlying biological process generating the failure time in the context of cancer relapse. The test of the effects of one covariate (on the cure fraction and the hazard rate in the uncured sub-population) has a degree of freedom of two and consequently can potentially compromise the power of the test. Furthermore, the EM-algorithm is often used to estimate the unknown parameters in the mixture cure models, which may increase the computation burden especially when there are no closed form solutions in the M-step. Additionally, in the Bayesian inference, the mixture cure model can yield improper posterior distributions for many non-informative improper priors, including the uniform prior for regression coefficients.

Alternatively, Yakovlev and Tsodikov (1996) proposed the promotion time cure model, which was further studied by Tsodikov (1998), Chen et al (1999), and Tsodikov et al (2003), among others. This model has a proportional hazards structure, a desirable feature of survival models, yields proper posterior distributions under a wide class of non-informative improper priors for the regression coefficients, and also has biological motivations (Chen et al, 1999). In this model, the population survival function for a subject with covariates  $\mathbf{Z}$  is given by

$$S_{\text{pop}}(t | \mathbf{Z}) = \exp \{ - e^{\beta^T \mathbf{Z}} F(t) \}, \quad (1)$$

where  $F(t)$  is a distribution function. This model integrates the survival function for either cured subjects or non-cured subjects into one single formulation and assumes that the cured subjects have survival time equal to infinity. The hazard function corresponding to (1) is  $\lambda_{\text{pop}}(t | \mathbf{Z}) = e^{\beta^T \mathbf{Z}} f(t)$ , where  $f(t) = dF(t)/dt$ . The cure rate under model (1) is  $\lim_{t \rightarrow \infty} S_{\text{pop}}(t | \mathbf{Z}) = \exp(-e^{\beta^T \mathbf{Z}})$ . The promotional time cure model is also referred to as the proportional hazards cure model. As a matter of fact, the above promotional time cure model can be derived from a biological process generating the relapse time in cancer patients. For a cancer patient, let  $N$  denote the number of carcinogenic cells, which is assumed to follow a Poisson distribution with mean  $\theta$ . Here  $\theta$  may depend on some covariates, e.g.,  $\theta = e^{\beta^T \mathbf{Z}}$ . We denote by  $Y_j$  the random time for the  $j$ th carcinogenic cell to produce a detectable cancer mass. We assume that  $Y_j$ 's are i.i.d. with a distribution function  $F(\cdot)$  and are independent of  $N$ . It then can be shown that the probability of cancer free by time  $t$  for the patient is  $\exp\{-\theta F(t)\}$ , which takes the same form as the survival function in (1).

To further increase the flexibility of the promotional time cure model, Zeng et al (2006b) proposed the following class of cure rate models by using Box-Cox type of transformations

$$S_{\text{pop}}(t | \mathbf{Z}) = G_{\gamma} \{ e^{\beta^T \mathbf{Z}} F(t) \}, \quad (2)$$

where

$$G_{\gamma}(x) = \begin{cases} (1 + \gamma x)^{-1/\gamma}, & \gamma > 0 \\ e^{-x}, & \gamma = 0. \end{cases}$$

This class of models include the proportional hazards cure model and the proportional odds cure model as special cases. When  $\gamma = 0$ , model (2) reduces to the proportional hazards cure model; when  $\gamma = 1$ , model (2) has the proportional odds structure. This class of cure models have sound biological basis and provide flexible alternatives to the mixture cure models.

The studies of cure rate models for current status data are relatively limited. Lam and Xue (2005) considered a partly linear AFT model for susceptible subjects and used a sieve maximum likelihood approach. Ma (2008) and Ma (2009) used additive risk model and linear or partly linear Cox models (Cox, 1972), respectively, to model the event time in the susceptible population. In all three papers, the authors considered mixture cure rate models. Cook et al (2008) also considered parametric and semiparametric mixture cure models but focused on the estimation of covariates effects on the cure rate only. Liu and Shen (2009) considered the proportional hazards cure model for interval-censored data and proved the strong consistency of the maximum likelihood estimators under the Hellinger distance.

In this article, we investigate current status data with a cure fraction. Particularly, we consider the class of transformation models (2). Although this class of models have been well studied for right-censored data, such models have not been used to analyze current status data and the extension is far from trivial. Our research aims to fill in this gap both theoretically and practically. In the next section, we derive the likelihood function and propose using maximum likelihood approach for estimation. In section 3 we establish the asymptotic properties of the proposed maximum likelihood estimators (MLEs). In Section 4 we conduct simulation studies to evaluate the finite-sample properties of the estimators and also illustrate the proposed model with the Calcification data. We conclude with a brief discussion in Section 5 and provide technical details for the proofs in the Appendix.

## 2 Maximum Likelihood Estimation

Suppose that there are  $n$  i.i.d. observations. For  $i = 1, \dots, n$ , we observe  $\{Y_i, \delta_i = I(T_i \leq Y_i), \mathbf{Z}_i\}$ , where  $T_i$  is the failure time for member  $i$ ,  $Y_i$  is the random censoring or inspection time,  $\mathbf{Z}_i$  is a vector of bounded covariates, and  $I(\cdot)$  is the indicator function. The first component of  $\mathbf{Z}_i$  is 1. We assume that the inspection time  $Y_i$  is conditionally independent of  $T_i$  given  $\mathbf{Z}_i$ .

We consider model (2). The observed-data likelihood function for the parameters  $(\boldsymbol{\beta}, F)$  can be expressed, for fixed  $\gamma$ , as

$$L_n(\boldsymbol{\beta}, F) = \prod_{i=1}^n \left[ 1 - G_\gamma \left\{ e^{\boldsymbol{\beta}^T \mathbf{Z}_i F(Y_i)} \right\} \right]^{\Delta_i} \left[ G_\gamma \left\{ e^{\boldsymbol{\beta}^T \mathbf{Z}_i F(Y_i)} \right\} \right]^{1 - \Delta_i}.$$

We wish to obtain the maximum likelihood estimators of  $(\boldsymbol{\beta}, F)$  by maximize the above likelihood function over the parameter space

$$\Theta = \{(\boldsymbol{\beta}, F): \boldsymbol{\beta} \text{ is in a compact set, } \| \boldsymbol{\beta} \| \leq B, F(\cdot) \text{ is a cumulative distribution function on } [0, \infty)\}.$$

It is obvious that the MLEs exist since the likelihood function is bounded from above by 1 and  $\Theta$  is weakly compact. The MLEs, however, are not unique since the likelihood function depends on  $F$  only through the its values at the observed times  $Y_i, i = 1, \dots, n$ . Thus we focus on the maximization of  $L_n(\boldsymbol{\beta}, F)$  over all nondecreasing step functions with jumps at the  $Y_i$ 's for  $F(t)$ . We denote the MLEs of  $\boldsymbol{\beta}$  and  $F$  by  $\hat{\boldsymbol{\beta}}_n$  and  $\hat{F}_n$ . Furthermore, we claim that  $\hat{F}_n$  can only have jumps at those  $Y_i$ 's with  $\delta_i = 1$ . Let  $s_1 < \dots < s_m$  denote the ordered distinct time points of  $Y_i, i = 1, \dots, n$  with associated  $\delta_i = 1$ . For a subject with  $\delta_i = 0$  and  $s_k < Y_i < s_{k+1}$ , consider two distribution functions  $F_1$  and  $F_2$ , which have the same values at all time points except that  $F_1(Y_i) = F_1(s_k)$  and  $F_2(Y_i) > F_1(s_k)$ . Since  $G_\gamma(\cdot)$  is a decreasing function, it can be easily shown that  $L_n(\boldsymbol{\beta}, F_1) > L_n(\boldsymbol{\beta}, F_2)$ . Therefore,  $\hat{F}_n$  can have positive jumps only at  $s_k, k = 1, \dots, m$ . Similarly, we can show that  $\hat{F}_n(s_m) = 1$  since  $F$  is a proper distribution function. Therefore we maximize the likelihood function subject to the constraint that  $0 = F(s_1) < \dots < F(s_m) = 1$ .

The pool-adjacent-violator (PAV) approach (Barlow et al, 1972; Robertson et al, 1988) is often used in the optimization problems with monotone constraints, e.g., see Groeneboom and Wellner (1992), van der Laan and Jewell (2001), and Ma (2008, 2009). Alternatively, in this paper we use the following transformation

$$F(s_k) = 1 - e^{-\sum_{j=1}^k \alpha_j}, \quad k = 1, \dots, m-1$$

and the constrained optimization problem reduces to the maximization of the likelihood function over  $(\boldsymbol{\beta}, \alpha_1, \dots, \alpha_{m-1})$  without constraints. For small  $m$ , Newton-Raphson algorithm can be used to solve the score equations for  $\boldsymbol{\beta}$  and  $\{\alpha_k, k = 1, \dots, m-1\}$ . When  $m$  is large, we use the quasi-Newton method of Broyden, Fletcher, Goldfarb and Shanno (or the so-called BFGS algorithm) as described in Press et al (1992) to search for the optimum. The BFGS algorithm refines an approximation of the inverse Hessian matrix at each iteration via a rank-one update. Therefore, although the number of parameters under the proposed model increases with sample size, the computation burden for obtaining the NPMLEs increases linearly with sample size. In our experience, the BFGS algorithm performs well even when there are thousands of parameters. This algorithm has been successfully applied in optimization problems with a large number of parameters, e.g., see Diao et al (2013) and Yuan and Diao (2014). It is well known that for interval-censored data, the MLE of the distribution function may have zero mass at some inspection times even with  $n \rightarrow \infty$ . Numerically, the estimates of the jumps of  $F$  are negligible ( $< 10^{-6}$ ) at those time points. The same transformation technique was also used by Zeng et al (2006a) to estimate the cumulative hazard function under the semiparametric additive risks model for interval-censored data.

### 3 Asymptotic Properties

We first impose the following assumptions:

- (C1) Covariates  $\mathbf{Z}_i$  are bounded with probability one. Furthermore, if there exists a constant vector  $\boldsymbol{\beta}$  such that  $\boldsymbol{\beta}^T \mathbf{Z}_i = 0$  almost surely, then  $\boldsymbol{\beta} = \mathbf{0}$ .
- (C2) Conditional on  $\mathbf{Z}_i$ , the inspection time  $Y_i$  is independent of  $T_i$ .
- (C3) There exists a constant  $t_0 < \infty$  such that  $P(T_i > t_0 | \mathbf{Z}_i) = P(T_i = \infty | \mathbf{Z}_i) > 0$  almost surely. Furthermore, the inspection time  $Y_i$  is a random variable from a distribution with support  $[\tau_l, \tau_u]$  with  $0 < \tau_l < t_0 < \tau_u < \infty$  and  $P(Y_i > t_0 | \mathbf{Z}_i) > 0$ .
- (C4) The true values of  $\boldsymbol{\beta}$ , denoted by  $\boldsymbol{\beta}_0$ , belong to the interior of a known compact set,

$$\mathcal{B} = \{\boldsymbol{\beta}: \|\boldsymbol{\beta}\| \leq B \text{ for some constant } B\},$$

where  $\|\cdot\|$  is the Euclidean norm.

(C5) The true promotion time cumulative distribution function  $F_0$  belongs to the space

$$\mathcal{F} = \{F: F \text{ is a cumulative distribution function on } [0, \infty) \text{ and has strictly positive first order derivatives on } [\tau_l, t_0]\}.$$

**Remark 1**—Condition (C1) is the usual linear independence condition of  $\mathbf{Z}_i$  in regression settings. Conditions (C2), (C4), and (C5) are standard assumptions in semiparametric inferences with failure time data. Condition (C3) assumes that no new cases occur after  $t_0$  and that  $t_0$  is in the support of inspection time. Under this assumption, all censored observations beyond  $t_0$  are treated as “ $T_i = \infty$ ” (i.e., observed to be cured) and we can discriminate between cured and un-cured observations. The constant  $t_0$  serves as the “cure threshold” defined in Zeng et al (2006b). This condition is needed to ensure the identifiability of the unknown parameters  $(\beta, F)$ . Note that for a parametric cure rate model, the condition that some subjects are observed to be cured is not needed. More details on the identifiability of cure models are provided in Li et al (2001). For practical data analysis, we can set  $t_0 = s_m$  or choose a threshold value suggested by clinicians. In the mixture cure rate model setting, Ma (2008, 2009) set  $\hat{\Lambda}(t) = \infty$  for  $t > \max\{Y_j: j = 1, \dots, n\}$ , where  $\Lambda$  is the baseline cumulative hazard function for the un-cured sub-population. This technique is essentially the same as setting  $t_0 = s_m$ .

**Remark 2**—For right-censored failure time data, Zeng et al (2006b) imposed the assumption  $P(Y_i = \infty | \mathbf{Z}_i) > 0$ . However with current status data  $Y_i = \infty$  implies  $i = 1$  regardless of whether  $T_i = \infty$  or not therefore such observations are not informative.

Under conditions (C1) – (C3), the parameters  $\beta$  and  $F$  are identifiable. Suppose that two sets of parameters,  $(\beta, F)$  and  $(\tilde{\beta}, \tilde{F})$ , give the same likelihood function for the observed data, i.e., if for almost all  $(\mathbf{Z}, Y)$ ,

$$[1 - G_\gamma\{e^{\beta^T \mathbf{Z}} F(Y)\}]^\Delta [G_\gamma\{e^{\beta^T \mathbf{Z}} F(Y)\}]^{1-\Delta} = [1 - G_\gamma\{e^{\tilde{\beta}^T \mathbf{Z}} \tilde{F}(Y)\}]^\Delta [G_\gamma\{e^{\tilde{\beta}^T \mathbf{Z}} \tilde{F}(Y)\}]^{1-\Delta},$$

we claim that  $\beta = \tilde{\beta}$  and  $F(t) = \tilde{F}(t)$  for every  $t \in [\tau_l, t_0]$ . In fact, by condition (C3),  $F(t) = \tilde{F}(t) = 1$  for any  $t > t_0$ . Let  $Y = t_0$ , we have  $G_\gamma\{e^{\beta^T \mathbf{Z}}\} = G_\gamma\{e^{\tilde{\beta}^T \mathbf{Z}}\}$ . Then, from the monotonicity of  $G_\gamma$ , it follows that  $\beta^T \mathbf{Z} = \tilde{\beta}^T \mathbf{Z}$  for almost all  $\mathbf{Z}$ . Thus condition (C1) gives  $\beta = \tilde{\beta}$ . Furthermore, by letting  $\gamma = 1$  and  $Y = t, \tau_l < t < t_0$ , we have  $G_\gamma\{e^{\beta^T \mathbf{Z}} F(t)\} = G_\gamma\{e^{\tilde{\beta}^T \mathbf{Z}} \tilde{F}(t)\}$ . It then follows that  $F(t) = \tilde{F}(t)$  for any  $t \in [\tau_l, t_0]$ .

We establish the consistency of the MLEs in the following theorem.

**Theorem 1. (Consistency)**—Under conditions (C1), (C3) and (C4),  $\|\hat{\beta}_n - \beta_0\| \rightarrow 0$  and  $\sup_{t \in [\tau_l, t_0]} |\hat{F}_n(t) - F_0(t)| \rightarrow 0$  almost surely.

We outline the proof here and defer the details to Appendix A.1. Let

$$p(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) = \left[ 1 - G_\gamma \left( e^{\boldsymbol{\beta}^T \mathbf{Z}_{F(Y)}} \right) \right]^\Delta \left[ G_\gamma \left( e^{\boldsymbol{\beta}^T \mathbf{Z}_{F(Y)}} \right) \right]^{1-\Delta},$$

and

$$m(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) = \log \frac{p(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F)}{p(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_0, F_0)}$$

be the log-likelihood ratio. Let  $d((\boldsymbol{\beta}, F), (\boldsymbol{\beta}_0, F_0)) = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| + \sup_{t \in [\tau_l, b_0]} |F(t) - F_0(t)|$ , and  $\mathcal{M} = \{m(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) : (\boldsymbol{\beta}, F) \in \mathcal{B} \times \mathcal{F}\}$ . We show that  $\mathcal{M}$  is a  $P$ -Glivenko-Cantelli class, and  $m(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F)$  satisfies some other conditions. It follows from Theorem 5.8 in van der Vaart (2002) that

$$d((\hat{\boldsymbol{\beta}}_n, \hat{F}_n), (\boldsymbol{\beta}_0, F_0)) \xrightarrow{a.s.} 0.$$

This gives the desired result.

With the consistency result, we can establish the following result on convergence rate of the MLEs.

**Lemma 1. (Convergence rate)**—Under conditions (C1), (C3) and (C4),

$$d((\hat{\boldsymbol{\beta}}_n, \hat{F}_n), (\boldsymbol{\beta}_0, F_0)) = O_p(n^{-1/3})$$

Groeneboom and Wellner (1992) established that the best convergence rate for estimates of the nonparametric distribution function with one-sample current status data is  $n^{1/3}$ . Huang (1996) and Ma (2008, 2009) obtained similar results for the estimates of the baseline cumulative hazard function in the Cox model (Cox, 1972) and the mixture cure rate model, respectively. Lemma 1 indicates that the optimal convergence rate can be achieved under the proposed cure rate model. The detailed proof of Lemma 1 is given in Appendix A.2.

Before stating the result for the asymptotic normality of  $\hat{\boldsymbol{\beta}}_n$ , we derive the efficient score function for  $\boldsymbol{\beta}$ . The log-likelihood for a single observation  $X \equiv (Y, \mathbf{Z})$ , denoted by  $l(\boldsymbol{\beta}, F)$ , takes the form

$$l(\boldsymbol{\beta}, F) = \Delta \log [1 - G_\gamma \{e^{\boldsymbol{\beta}^T \mathbf{Z}_{F(Y)}}\}] + (1 - \Delta) \log [G_\gamma \{e^{\boldsymbol{\beta}^T \mathbf{Z}_{F(Y)}}\}].$$

The score function for  $\boldsymbol{\beta}$  is the first derivative of  $l(\boldsymbol{\beta}, F)$  with respect to  $\boldsymbol{\beta}$  and takes the form

$$l_{\beta}(\beta, F) = \mathbf{Z}F(Y)Q(X; \beta, F),$$

where

$$Q(X; \beta, F) = e^{\beta^T \mathbf{Z}} G_{\gamma}' \{ e^{\beta^T \mathbf{Z}} \mathbf{Z} F(Y) \} \left[ -\frac{\Delta}{1 - G_{\gamma} \{ e^{\beta^T \mathbf{Z}} \mathbf{Z} F(Y) \}} + \frac{1 - \Delta}{G_{\gamma} \{ e^{\beta^T \mathbf{Z}} \mathbf{Z} F(Y) \}} \right]$$

and  $G_{\gamma}'$  is the first derivative of  $G_{\gamma}$ . For a function  $g$  in  $[\tau_b, \tau_u]$  with bounded total variation, the score function for  $F$  along the direction of  $g$  is given by

$$I_F(\beta, F)[g] = g(Y)Q(X; \beta, F).$$

Project the score function of  $\beta$  onto the space generated by  $I_F(\beta, F)[g]$ , and for  $\mathbf{g} = (g_1, \dots, g_d)$  with  $d = \dim(\beta)$ , denote  $I_F(\beta, F)[\mathbf{g}] = (I_F(\beta, F)[g_1], \dots, I_F(\beta, F)[g_d])^T$ . We obtain the efficient score function for  $\beta$  at the true parameter

$$I_{\beta}^*(\beta_0, F_0) = I_{\beta}(\beta_0, F_0) - I_F(\beta_0, F_0)[\mathbf{g}^*],$$

where

$$\mathbf{g}^*(Y) = F_0(Y) \frac{E(\mathbf{Z}Q^2(\mathbf{X}; \beta_0, F_0) | Y)}{E(Q^2(\mathbf{X}; \beta_0, F_0) | Y)}.$$

Write  $I_{\beta}(\beta_0, F_0) = E[l_{\beta}^*(\beta_0, F_0)l_{\beta}^{*T}(\beta_0, F_0)]$ . In the appendix, we will show that  $I_{\beta}(\beta_0, F_0)$  is positive definite and component-wise bounded.

With the derived convergence rate and the efficient score function for  $\beta$ , we can obtain the following asymptotic normality results for  $\hat{\beta}_n$ .

**Theorem 2. (Asymptotic normality)**—Under conditions (C1)–(C4),  $\sqrt{n}(\hat{\beta}_n - \beta_0)$  converges weakly to a zero-mean normal distribution with variance  $I_{\beta}^{-1}(\beta_0, F_0)$ .

Theorem 2 states that although the convergence rate of  $(\hat{\beta}_n, \hat{F}_n)$  is considerably slower than the  $n^{1/2}$  rate for right censored data, we can still obtain the  $\sqrt{n}$  consistency and asymptotic normality of  $\hat{\beta}_n$ . Furthermore, we conclude that  $\hat{\beta}_n$  is the most efficient estimator for  $\beta_0$  as



$I_{\beta}^{-1}(\beta_0, F_0)$  attains the semiparametric efficiency bound for  $\beta_0$ . The proof of Theorem 2 is provided in Appendix A.3.

Let  $\mathbb{B}(\cdot)$  be the two-sided Brownian motion originating from zero: a mean zero Gaussian process on  $R$  with  $\mathbb{B}(0) = 0$ , and  $E(\mathbb{B}(s) - \mathbb{B}(h))^2 = |s - h|$  for all  $s, h \in R$ . Denote  $\xrightarrow{D}$  for convergence in distribution. Let  $w_n$  be the random variable that given the  $\mathbf{Z}_i$ 's,  $w_n$  takes  $e^{-\hat{\beta}_n^T \mathbf{Z}_i} G_{\gamma}^{-1}(1 - i/n) - e^{-\hat{\beta}_n^T \mathbf{Z}_{i+1}} G_{\gamma}^{-1}(1 - (i + 1)/n)$  with probability  $1/n$ . Denote  $w = w(\mathbf{Z})$  for its asymptotic version conditioning on  $\mathbf{Z}$ . Let  $f_0(\cdot)$  be the derivative of  $F_0(\cdot)$ .

To derive the asymptotic distribution of  $\hat{F}_n$ , we make two additional assumptions.

(C6) Function  $G_{\gamma}$  satisfies  $0 < \lim_{x \rightarrow \infty} G_{\gamma}^{-1}(1/x)/x < \infty$ .

(C7) The density function of  $Y$ ,  $g(t) > 0$  with  $t \in [\tau_1, t_0]$ .

**Theorem 3**—Assume (C1)–(C7), then  $\forall t \in [\tau_1, t_0]$ ,

$$n^{1/3}(\hat{F}_n(t) - F_0(t)) \xrightarrow{D} \left( \frac{4\eta^2(t)f_0(t)}{g(t)} \right)^{1/3} \arg \min_{h \in R} \{ \mathbb{B}(h) + h^2 \},$$

where  $\eta_2(t) = E[(I(T = t)w(\mathbf{Z}) - F_0(t))^2]$ .

The distribution of  $\arg \min_{h \in R} \{ \mathbb{B}(h) + h^2 \}$  is called Chernoff distribution, the density function of which was derived in Groeneboom (1989). It has no closed form and is not easy to evaluate. Kosorok (2008) proposed a sampling method to evaluate this distribution. The proof of Theorem 3 is in Appendix A.4.

**Remark 3**—The inference above is based on the selected model with the transformation parameter  $\gamma$  fixed. To select the best model, we may choose the transformation which minimizes the Akaike information criterion (AIC), which is defined as twice the negative log-likelihood function plus twice the number of parameters. It is theoretically possible to accounting for the variation due to the model selection procedure. The computation, however, may be demanding or unstable. Furthermore, whether this kind of variation should be accounted for is up to debate (Box and Cox, 1982). In the subsequent simulation studies, we will fix the transformation whereas in the data application, we will use the AIC to select the best transformation.

## 4 Numerical Studies

### 4.1 Simulation

We conducted simulation studies to examine the finite sample performance of our proposed methodology. We generated data from model (2) with  $\mathbf{Z} = (1, Z_1, Z_2)^T$ , where  $Z_1$  is a uniform random variable in  $[0, 1]$ , and  $Z_2$  is a Bernoulli random variable with a success probability 0.5. The true parameter values of  $\beta = (\beta_0, \beta_1, \beta_2)$  are set to be  $(-0.5, 1, -0.5)$ ;

and  $F(t) = 1 - \{\exp(-t) - \exp(-t_0)\} I(t - t_0) / \{1 - \exp(-t_0)\}$ , where  $t_0 = 4$ . The inspection time was set to be the minimum of 4 and an exponential variable with mean 2. We considered five different models by varying the values of  $\gamma$  from 0 to 1. The average cure rate ranged from 0.45 to 0.56 as  $\alpha$  changed from 0 to 1, whereas the censoring proportion among those uncured subjects was about 0.26 for all five models. We considered sample sizes of 200 and 400 for each model. For each simulation set-up, we generated 1,000 data sets. The quasi-Newton algorithm implemented in C language was applied to maximize the log-likelihood function. The initial values of the regression parameters were set to be 0 and the initial jump sizes of  $F$  were set to be  $1/m$  at  $s_1, \dots, s_m$ . The convergence of the quasi-Newton algorithm was very fast and it took about 0.3 second to analyze one data set with 400 subjects on a Dell PowerEdge 2900 server.

Table 1 summarizes the results for each combination of  $\gamma$  and  $n$ . The column labeled as “Mean” denotes the sampling mean of the parameter estimator from the true parameter value; “SE” is the sampling standard error of the parameter estimator; “SEE” is the mean of the standard error estimator, which is calculated based on the inverse of the observed information; and “CP(%)” is the coverage probability of the 95% confidence interval based on the asymptotic normal approximation. The proposed estimators appear to be unbiased. The estimate of the standard error reflects accurately the true variation, and for moderate sample size and cluster size the confidence intervals have proper coverage probabilities. As the sample size increases from 200 to 400, the standard deviations of the estimates shrink at approximately the  $\sqrt{2}$  rate.

The next line of simulation studies concerned about the performance of the commonly used proportional hazards cure model and the proportional odds cure model when data were generated from a different model. Specifically, we used the same setting for generating data in the simulation studies described earlier. The model with  $\gamma = 1/2$  corresponds to a model between the proportional hazards cure model and the proportional odds cure models. For each simulation setting, we generate 1,000 replications with sample size of  $n = 400$ . The results are summarized in Table 2. We observe when the true model is the proportional hazards cure model, the regression parameters from the proportional odds cure model are overestimated, whereas the parameters from the proportional hazards cure model are underestimated when the proportional proportional odds cure model is true. In general, when the transformation parameter is between 0 and 1, the proportional hazards cure model tends to bias towards 0, whereas the opposite is observed for the proportional odds cure model. Note that both models estimate the direction of the coefficients correctly under misspecified transformation. We also observe that standard error estimates of the regression coefficients corresponding to the covariates appear to be correct although the parameter estimates are biased.

## 4.2 Application to the Calcification Study

As an illustration, we applied our proposed model to the Calcification study of Yu et al (2001). The objective of the study was to investigate the effects of certain clinical factors on the time to IOL calcification after implantation. The status of calcification for each of the 379 patients in the study was determined by an experienced ophthalmologist at a random

time ranging from 0 to 36 months after implantation. Covariates of interest include incision length, gender, and age at implantation. One individual with missing measurement on incision length was excluded from the analysis. Of the remaining 378 individuals, 48 experienced calcification before examination time. The nonparametric maximum likelihood estimates of the survival functions for males and females at the end of study were 0.844 and 0.5, respectively.

In our data analysis, we included incision length, gender, and age at implantation as covariates in the cure models. Let  $Z_1$  be the incision length,  $Z_2$  be gender that takes values 0 and 1 for female and male patients, respectively, and  $Z_3$  be the age at implantation divided by 10. We fit the proposed model (2) with  $\mathbf{Z} = (1, Z_1, Z_2, Z_3)^T$ . We compared the proportional hazards cure model and the proportional odds cure model, and those models in between. As suggested by Zeng et al (2006b), we selected the proportional odds cure model which maximizes the Akaike information criterion (AIC). The AIC is defined as the twice log-likelihood function minus twice the number of parameters. The results under the proportional odds cure model are presented in Table 3. We see that males and older people are more likely to be cured and increase in incision length leads to an decrease in cure rate, although their effects are not significant. Our findings agree well with those of Lam and Xue (2005). Furthermore, to give a graphical comparison, Figure 1 plots the separate nonparametric estimates of the survival curve and the model-fitted survival curve for each gender group. The model-fitted survival function is calculated as the empirical average of the predicted survival functions. The dotted and dot-dashed lines in Figure 1 present the predicted survival functions; these in general agree quite well with the nonparametric estimates of the survival curves. The big difference between the fitted survival function and the nonparametric estimate of the survival function for males at the tail is due to that the largest inspection time in male patients with the occurrence of calcification was 17.6 months. Therefore the nonparametric estimate of the survival function for males is flat whereas the MLE of  $F_0$  in the fitted model has jumps beyond that time point. Figure 1 indicates that the proportional odds cure model fits the data well.

## 5 Discussion

We have considered a class of semiparametric cure models by imposing the Box-Cox type transformation on the population survival function for the analysis of current status data with a cured proportion in the population. This class includes the well-known proportional hazards and proportional odds structures as two special cases and has great potential in clinical trials and for modeling survival data with a cure fraction. We have developed efficient likelihood-based estimation and inference procedures and shown that the MLEs of the regression parameters are consistent, asymptotically normal and asymptotically efficient. The inference in this paper is based on the selected model with the transformation parameter fixed. As suggested by Zeng et al (2006b) we use the AIC to select the best model. It would be of interest to estimate the variances of the MLEs of the regression parameters accounting for the variation due to the model selection procedure in semiparametric inference. Future research is warranted.

We have developed an efficient algorithm implemented in C language for calculating the MLEs and standard error estimates of the MLEs under the proposed models. The computation is substantially faster than the PAV approach under the mixture cure model in Ma (2009). This significant improvement makes it possible to apply our methods to genetic studies in which linkage or association tests are performed at hundreds of thousands of genetic markers. Our computer program is available upon request.

As noted by Taylor (1995), Fang et al (2005), and Lam et al (2005), applications of cure models are restricted to problems in which there is strong scientific evidence for the existence of the cured population or a clear level plateau of the survival function is observed. Identifiability problem may arise when the cure probability is very close to one or zero. In our experience, the numerical computation of the MLEs may not be stable if the cure fraction is small.

## Acknowledgments

NIH grant 1R15CA150698-01

The authors wish to thank Dr. Donglin Zeng and the referees for their helpful comments and suggestions, which lead to a considerable improvement in the presentation of this manuscript. The authors also would like to thank Drs. A. F. K. Yu, K. F. Lam, and HongQi Xue for providing the Calcification data.

## References

- Barlow, RE, Bartholomew, DJ, Bremner, JM, Brunk, HD. *Statistical Inference Under Order Restrictions*. Wiley; New York: 1972.
- Berkson J, Gage RP. 1952; Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*. 47:501–515.
- Betensky RA, Schoenfeld DA. 2001; Nonparametric estimation in a cure model with random cure times. *Biometrics*. 57(1):282–286. [PubMed: 11252611]
- Box GEP, Cox DR. 1982; An analysis of transformation revisited, rebutted. *Journal of The American Statistical Association*. 77:209–210.
- Chen MH, Ibrahim JG, Sinha D. 1999; A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*. 94:909–919.
- Cook RJ, White BJG, Grace YY, Lee KA, Warkentin TE. 2008; Analysis of a nonsusceptible fraction with current status data. *Statistics in Medicine*. 27:2715–2730. [PubMed: 17943708]
- Cox DR. 1972; Regression model and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B*. 34:187–220.
- Diao G, Zeng D, Yang S. 2013; Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics*. 69(4):840–849. [PubMed: 24328712]
- Fang H, Li G, Sun J. 2005; Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics*. 32(1):59–75.
- Farewell VT. 1982; The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 38:1041–1046. [PubMed: 7168793]
- Farewell VT. 1986; Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*. 14:257–262.
- Gray RJ, Tsiatis AA. 1989; A linear rank test for use when the main interest is in differences in cure rates. *Biometrics*. 45(3):899–904. [PubMed: 2790128]
- Groeneboom P. 1989; Brownian motion with a parabolic drift and airy functions. *Probability theory and related fields*. 81(1):79–109.
- Groeneboom, P, Wellner, JA. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser; Basel: 1992.

- Hoel DG, Walburg HE. 1972; Statistical analysis of survival experiments. *Journal of National Cancer Institute*. 49:361–372.
- Huang J. 1996; Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*. 24(2):540–568.
- Huang J, Rossini AJ. 1997; Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*. 92(439):960–967.
- Ibrahim, JG, Chen, MH, Sinha, D. *Bayesian Survival Analysis*. Springer; New York: 2001.
- Kosorok, MR. *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*. Institute of Mathematical Statistics; 2008. Bootstrapping the grenander estimator; 282–292.
- Kuk AYC, Chen CH. 1992; A mixture model combining logistic-regression with proportional hazards regression. *Biometrika*. 79(3):531–541.
- van der Laan MJ, Jewell NP. 2001; The NPMLE for doubly censored current status data. *Scandinavian Journal of Statistics*. 28(3):537–547.
- Lam K, Xue H. 2005; A semiparametric regression cure model with current status data. *Biometrika*. 92(3):573–586. DOI: 10.1093/biomet/92.3.573
- Lam KF, Fong DYT, Tang O. 2005; Estimating the proportion of cured patients in a censored sample. *Statistics in Medicine*. 24(12):1865–1879. DOI: 10.1002/sim.2137 [PubMed: 15900587]
- Li CS, Taylor JMG, Sy JP. 2001; Identifiability of cure models. *Statistics & Probability Letter*. 54:389–395.
- Lin DY, Oakes D, Ying Z. 1998; Additive hazards regression with current status data. *Biometrika*. 85(2):289–298.
- Liu H, Shen Y. 2009; A semiparametric regression cure model for interval-censored data. *Journal of The American Statistical Association*. 104:1168–1178. [PubMed: 20354594]
- Lu W, Ying Z. 2004; On semiparametric transformation cure models. *Biometrika*. 91(2):331–343.
- Ma S. 2008 Additive risk model for current status data with a cured subgroup. *Annals of the Institute of Statistical Mathematics*.
- Ma S. 2009; Cure model with current status data. *Statistica Sinica*. 19(1):233–249.
- Ma S, Kosorok MR. 2005; Penalized log-likelihood estimation for partly linear transformation models with current status data. *Annals of Statistics*. 33(5):2256–2290. DOI: 10.1214/009053605000000444
- Maller, R, Zhou, X. *Survival Analysis with Long-Term Survivors*. Wiley; New York: 1996.
- Peng Y, Dear KBG. 2000; A nonparametric mixture model for cure rate estimation. *Biometrics*. 56(1): 237–243. [PubMed: 10783801]
- Press, WH, Teukolsky, SA, Vetterling, WT, Flannery, BP. *Numerical Recipes in C: the art of scientific computing*. 2. Cambridge University Press; 1992.
- Robertson, T, Wright, F, Dykstra, R. *Order Restricted Statistical Inference*. Wiley; New York: 1988.
- Rossini AJ, Tsiatis AA. 1996; A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*. 91(434):713–721.
- Sy JP, Taylor JMG. 2000; Estimation in a cox proportional hazards cure model. *Biometrics*. 56(1):227–236. [PubMed: 10783800]
- Taylor J. 1995; Semiparametric estimation in failure time mixture-models. *Biometrics*. 51(3):899–907. [PubMed: 7548707]
- Tsodikov A. 1998; A proportional hazards model taking account of long-term survivors. *Biometrics*. 54:1508–1516. [PubMed: 9883549]
- Tsodikov AD, Ibrahim JG, Yakovlev AY. 2003; Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*. 98(464):1063–1078. DOI: 10.1198/01622145030000001007 [PubMed: 21151838]
- van der Vaart, A, Wellner, J. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag; New York: 1996.
- van der Vaart, AW. *Lectures on Probability Theory and Statistics*. Lecture Notes in Math. Vol. 1781. Springer; New York: 2002. Semiparametric statistics; 331–457.

- Xue H, Lam K, Li G. 2004; Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of the American Statistical Association*. 99(466):346–356. DOI: 10.1198/016214504000000313
- Yakovlev, AY, Tsodikov, AD. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific; New Jersey: 1996.
- Yu A, Kwan K, Chan D, Fong D. 2001; Clinical features of 46 eyes with calcified hydrogel intraocular lenses. *Journal of Cataract and Refractive Surgery*. 27:1596–1606. [PubMed: 11687358]
- Yuan M, Diao G. 2014; Semiparametric odds rate model for modeling short-term and long-term effects with application to a breast cancer genetic study. *The international journal of biostatistics*. 10(2): 231–249. [PubMed: 24815054]
- Zeng D, JC, YS. 2006a; Semiparametric additive risks model for interval-censored data. *Statistica Sinica*. 16:287–302.
- Zeng D, Yin G, Ibrahim J. 2006b; Semiparametric transformation models for survival data with a cure fraction. *Journal of American Statistical Association*. 101:670–684.

## Appendix

### A.1. Proof of Theorem 1

Let  $p(Y, \mathbf{Z}, \boldsymbol{\beta}, F) = [1 - G_\gamma(e^{\boldsymbol{\beta}^T \mathbf{Z}} F(Y))] [G_\gamma(e^{\boldsymbol{\beta}^T \mathbf{Z}} F(Y))]^{1-}$ , and

$$q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) = \log \frac{p(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F)}{p(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_0, F_0)}$$

be the log-likelihood ratio. Let  $P$  be the probability measure of  $p(Y, \mathbf{Z}, \boldsymbol{\beta}_0, F_0)$ ,  $Pq(Y, \mathbf{Z}, \boldsymbol{\beta}, F) = \int q(Y, \mathbf{Z}, \boldsymbol{\beta}, F) dP(Y, \mathbf{Z}, \Delta)$ , the true mean of  $q$ ; and let

$P_n q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) = (1/n) \sum_{i=1}^n q(Y_i, \mathbf{Z}_i, \Delta_i | \boldsymbol{\beta}, F)$ , the empirical mean of  $q$  based on the data

$\{(Y_i, \mathbf{Z}_i, \Delta_i) : i = 1, \dots, n\}$ . Note that  $Pq(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$  is the negative Kullback-Leibler divergence of  $p(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$  from  $p(Y, \mathbf{Z}, \boldsymbol{\beta}_0, F_0)$ , and as a function of  $(\boldsymbol{\beta}, F)$ , it is always non-positive, attaining its maximum value of 0 at  $(\boldsymbol{\beta}, F) = (\boldsymbol{\beta}_0, F_0)$ . Let  $\Theta$  be the range of  $\boldsymbol{\beta}$  and  $\mathcal{F}$  be the collection of all distribution functions on  $R^+$ . Recall that  $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)$  is the MLE of  $(\boldsymbol{\beta}_0, F_0)$ , i.e.,

$$(\boldsymbol{\beta}_0, F_0) \in \arg \max_{(\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F})} P q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F), \quad \text{and} \quad (\hat{\boldsymbol{\beta}}_n, \hat{F}_n) \in \arg \max_{(\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F})} P_n q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F).$$

Let  $d((\boldsymbol{\beta}, F), (\boldsymbol{\beta}_0, F_0))$  as given before the Lemma (recall definitions of  $\tau_I$  and  $t_0$  given in condition (C3)). By our model specification,  $Pq(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$  is continuous with respect to  $(\boldsymbol{\beta}, F)$ ; also, with (C3), the model is identifiable, and so  $(\boldsymbol{\beta}_0, F_0)$  is the unique maximizer of  $Pq(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$ . Thus for all  $\eta > 0$ ,

$$\sup_{(\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F}) : d((\boldsymbol{\beta}, F), (\boldsymbol{\beta}_0, F_0)) > \eta} P q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}, F) < P q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_0, F_0)$$

and by definition of MLE,  $P_n q(Y, \mathbf{Z}, \Delta | \hat{\boldsymbol{\beta}}_n, \hat{F}_n) \geq P_n q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_0, F_0)$ ,

so if we show that  $\mathcal{Q} = \{q(Y, \mathbf{Z}, \boldsymbol{\beta}, F) : (\boldsymbol{\beta}, F) \in (\boldsymbol{\Theta}, \mathcal{F})\}$  is a  $P$ -Glivenko-Cantelli class, then by Theorem 5.8 in van der Vaart (2002, p.386),

$$d((\hat{\boldsymbol{\beta}}_n, \hat{F}_n), (\boldsymbol{\beta}_0, F_0)) \xrightarrow{a.s.} 0.$$

This gives the desired result. Note that  $\hat{F}_n$  is estimated by data points  $s_1, \dots, s_m$ .

Now we show that  $\mathcal{Q}$  is  $P$ -Glivenko-Cantelli. For any function  $g$ , let  $\|g\|_{L_1(P)} = \int |g(y, \mathbf{z}, \delta)| dP(y, \mathbf{z}, \delta)$ , and  $N_{[]}(\varepsilon, \mathcal{Q}, L_1(P))$  be the minimum number of  $\varepsilon$ -brackets needed to cover  $\mathcal{Q}$  under norm  $\|\cdot\|_{L_1(P)}$ . We first show that  $N_{[]}(\varepsilon, \mathcal{Q}, L_1(P))$  is finite  $\forall \varepsilon > 0$ .

By our specification of  $p(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$  and with (C1), it can be checked that  $q(Y, \mathbf{Z}, \boldsymbol{\beta}, F)$  is boundedly differentiable with respect to  $\boldsymbol{\beta}$ , and boundedly Gâteaux differentiable with respect to  $F$ . Consequently, by Taylor expansion, there are constants  $0 < C_j < \infty$  ( $j = 1, 2$ ), such that

$$\|q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_1, F_1) - q(Y, \mathbf{Z}, \Delta | \boldsymbol{\beta}_2, F_2)\|_{L_1(P)} \leq C_1 \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| + C_2 \|F_1 - F_2\|_{L_1(P)},$$

$$\forall (\boldsymbol{\beta}_1, F_1), (\boldsymbol{\beta}_2, F_2) \in (\boldsymbol{\Theta}, \mathcal{F}). \text{ So,}$$

$$N_{[]}(\varepsilon, \mathcal{Q}, \|\cdot\|_{L_1(P)}) \leq N_{[]}(\frac{\varepsilon}{2C_1}, \boldsymbol{\Theta}, \|\cdot\|) \times N_{[]}(\frac{\varepsilon}{2C_2}, \mathcal{F}, \|\cdot\|_{L_1(P)}), \quad \forall \varepsilon < 0.$$

By (C4),  $N_{[]}(\varepsilon(2C_1), \boldsymbol{\Theta}, \|\cdot\|) = O(1/\varepsilon^d)$ , with  $d = \dim(\boldsymbol{\Theta})$ . Since  $\mathcal{F}$  is a collection of bounded monotone functions, by Theorem 2.7.5 in van der Vaart and Wellner (1996, p.159), for some constant  $0 < C < \infty$ ,

$$N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{L_1(P)}) \leq \exp\left\{\frac{C}{\varepsilon}\right\}, \quad \forall \varepsilon > 0.$$

Thus, for some generic constant  $0 < C < \infty$ ,

$$\begin{aligned} N_{[]}(\varepsilon, \mathcal{Q}, \|\cdot\|_{L_1(P)}) &\leq N_{[]}(\frac{\varepsilon}{2C_1}, \boldsymbol{\Theta}, \|\cdot\|) \times N_{[]}(\frac{\varepsilon}{2C_2}, \mathcal{F}, L_1(P)) \\ &\leq \frac{C}{\varepsilon^d} \exp\left\{\frac{C}{\varepsilon}\right\} < \infty, \quad \forall \varepsilon > 0, \end{aligned}$$

and so by Theorem 2.4.1 in van der Vaart and Wellner (1996, p.122),  $\mathcal{Q}$  is a Glivenko-Cantelli class with respect to  $P$ .

## A.2 Proof of Lemma 1

Denote  $S = (Y, \mathbf{Z}, \delta)$ , let  $p(S|\boldsymbol{\beta}, F)$ ,  $P$  and  $P_n$  as given in the proof of Theorem 1,  $l(\boldsymbol{\beta}, F|S) = \log p(S|\boldsymbol{\beta}, F)$  be the log-likelihood, and  $\mathbf{D}_n$  be the set of all the observed data. Define

$$\mathbb{M}_n(\boldsymbol{\beta}, F) = P_n \ell(\boldsymbol{\beta}, F | S) \quad \text{and} \quad \mathbb{M}(\boldsymbol{\beta}, F) = P \ell(\boldsymbol{\beta}, F | S).$$

Since  $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n) \in (\Theta, \mathcal{F})$  is the MLE of  $(\boldsymbol{\beta}_0, F_0)$ , it is the M-estimator based on the log-likelihood  $\ell(\boldsymbol{\beta}, F | \mathbf{D}_n)$  on the parameter space  $(\Theta, \mathcal{F})$ , and so

$$\mathbb{M}_n(\hat{\boldsymbol{\beta}}_n, \hat{F}_n) \geq \mathbb{M}_n(\boldsymbol{\beta}, F) \geq \mathbb{M}_n(\boldsymbol{\beta}, F) - O_p(r_n^{-2}), \text{ for any } (\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F}) \text{ and any positive sequence } r_n \rightarrow \infty.$$

Let  $d(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{F}_n - F_0) = \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \|\hat{F}_n - F_0\|_{L_2}$ . Since conditions of Theorem 1 are satisfied, in its proof change  $\sup_{t \in [\tau, t_0]} |\hat{F}_n(t) - F_0(t)|$  to  $\|\hat{F}_n - F_0\|_{L_2}$ , we get

$$d(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{F}_n - F_0) \xrightarrow{P} 0.$$

Note that  $(\boldsymbol{\beta}_0, F_0) = \arg \sup_{(\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F})} \mathbb{M}(\boldsymbol{\beta}, F)$ . Denote by

$$\ell_{\boldsymbol{\beta}}^{(1)}(\boldsymbol{\beta}_0, F_0 | S), \ell_F^{(1)}(\boldsymbol{\beta}_0, F_0 | S)[F - F_0], \ell_{\boldsymbol{\beta}\boldsymbol{\beta}}^2(\boldsymbol{\beta}_0, F_0 | S), \ell_{\boldsymbol{\beta}, F}^2(\boldsymbol{\beta}_0, F_0 | S)[F - F_0], \ell_{F, \boldsymbol{\beta}}^2(\boldsymbol{\beta}_0, F_0 | S)[F - F_0], \ell_{F, F}^{(2)}(\boldsymbol{\beta}_0, F_0 | S)[F - F_0, F - F_0]$$

the partial derivatives of  $\ell(\boldsymbol{\beta}, F | S)$  with respect to  $(\boldsymbol{\beta}, F)$ . The derivative with respect to  $F$  is in the Gâteaux sense. By our specification of model, these quantities exist. Note that

$$E_{(\boldsymbol{\beta}_0, F_0)} \ell_{\boldsymbol{\beta}}^{(1)}(\boldsymbol{\beta}_0, F_0 | S) = \mathbf{0}, \text{ and } E_{(\boldsymbol{\beta}_0, F_0)} \ell_F^{(1)}(\boldsymbol{\beta}_0, F_0 | S)[F - F_0] = 0. \text{ Denote by } \mathcal{Q}^{(2)}(\boldsymbol{\beta}, F | S)[F - F_0, F - F_0] \text{ the } (d + 1) \times (d + 1) \text{ matrix of all the second order partial derivatives.}$$

Therefore,

$$\mathbb{M}(\boldsymbol{\beta}, F) - \mathbb{M}(\boldsymbol{\beta}_0, F_0) = \frac{1}{2} E_{(\boldsymbol{\beta}_0, F_0)} \left( (\boldsymbol{\beta} - \boldsymbol{\beta}_0, 1)^T \mathcal{Q}^{(2)}(\bar{\boldsymbol{\beta}}, \bar{F} | S)[F - F_0, F - F_0] (\boldsymbol{\beta} - \boldsymbol{\beta}_0, 1) \right) < 0$$

and is of order  $O(d^2(\boldsymbol{\beta} - \boldsymbol{\beta}_0, F - F_0))$ , where  $(\bar{\boldsymbol{\beta}}, \bar{F})$  is an intermediate value between  $(\boldsymbol{\beta}, F)$  and  $(\boldsymbol{\beta}_0, F_0)$ . Under some general conditions, the above can be upper bounded by  $-Cd^2(\boldsymbol{\beta} - \boldsymbol{\beta}_0, F - F_0)$  for some  $0 < C < \infty$ , in small neighborhood of  $(\boldsymbol{\beta}_0, F_0)$ . So for any  $0 < \eta_n \rightarrow 0$  and any  $\tau$  with  $\eta_n < \tau < \infty$ , for some  $0 < C < \infty$ ,

$$\tau/2 < d(\boldsymbol{\beta} - \boldsymbol{\beta}_0, F - F_0) \leq \tau, (\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F}) \implies \mathbb{M}(\boldsymbol{\beta}, F) - \mathbb{M}(\boldsymbol{\beta}_0, F_0) \leq -C\tau^2.$$

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

$$E^* \sup_{\tau/2 < d(\boldsymbol{\beta} - \boldsymbol{\beta}_0, F - F_0) \leq \tau, (\boldsymbol{\beta}, F) \in (\Theta, \mathcal{F})} \sqrt{n} | (\mathbb{M}_n - \mathbb{M})(\boldsymbol{\beta}, F) - (\mathbb{M}_n - \mathbb{M})(\boldsymbol{\beta}_0, F_0) | \leq \phi_n(\tau),$$

(A.0)



for some decreasing function  $\phi_n(\cdot)$  to be given.

For this, let  $N_{[\cdot]}(\varepsilon, (\Theta, \mathcal{F}), L_2(P))$  be defined in the proof of Theorem 1, we showed  $N_{[\cdot]}(\varepsilon, (\Theta, \mathcal{F}), L_2(P)) = O(1/\varepsilon^d \exp\{C/\varepsilon\}) = O(\exp\{C/\varepsilon\})$ . Let  $C$  be some generic finite positive constant, and

$$\begin{aligned} J_{[\cdot]}(\tau, (\Theta, \mathcal{F}), L_2(P)) &= \int_0^\tau \sqrt{1 + \log N_{[\cdot]}(\varepsilon, (\Theta, \mathcal{F}), L_2(P))} d\varepsilon \\ &= \int_{-\infty}^{\log \tau} e^t \sqrt{1 + Ce^{-t}} dt \leq C \int_{-\infty}^{\log \tau} e^t e^{-t/2} dt = C\tau^{1/2}. \end{aligned}$$

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

Let  $\mathcal{Q}_1 = \{q(\cdot|\beta, F): (\beta, F) \in (\Theta, \mathcal{F}), \tau/2 < d(\beta - \beta_0, F - F_0) < \tau\}$ ,  $\mathbb{G}_n q = \sqrt{n}(P_n - P)q$ ,  $\|\mathbb{G}_n q\|_{\mathcal{Q}_1} = \sup_{q \in \mathcal{Q}_1} \|\mathbb{G}_n q\|$ . Then the left hand side of (A.0) is upper bounded by  $2\|\mathbb{G}_n\|_{\mathcal{Q}_1}$ .

Note that since  $\mathcal{Q}_1$  is a subset of the  $\mathcal{Q}$  defined in the proof of Theorem 1, so  $N_{[\cdot]}(\varepsilon, \mathcal{Q}_1, L_2(P)) \leq N_{[\cdot]}(\varepsilon, \mathcal{Q}, L_2(P)) \leq N_{[\cdot]}(\varepsilon, (\Theta, \mathcal{F}), L_2(P))$  and consequently,  $J_{[\cdot]}(\tau, \mathcal{Q}_1, L_2(P)) \leq J_{[\cdot]}(\tau, (\Theta, \mathcal{F}), L_2(P)) \leq C\tau^{1/2}$ .

Also, it is easy to see that, with our specification of the likelihood,  $Pq^2 < C\tau^2$  and  $\|q\|_\infty < C$  for all  $q \in \mathcal{Q}_1$ , for some  $0 < C < \infty$ . Thus, by Lemma 3.4.2 in van der Vaart and Wellner (1996, p.324),

$$E^* \|\mathbb{G}_n\|_{\mathcal{Q}_1} \leq J_{[\cdot]}(\tau, \mathcal{Q}, L_2(P)) \left( 1 + \frac{J_{[\cdot]}(\tau, \mathcal{M}, L_2(P))}{\tau^2 \sqrt{n}} C \right) \leq C\tau^{1/2} (1 + \tau^{-3/2} n^{-1/2}),$$

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

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

Now, all conditions of Theorem 3.4.1 in van der Vaart and Wellner (1996, p.322) are satisfied, so by this Theorem,  $r_n d(\hat{\beta}_n - \beta_0, \hat{F}_n - F_0) = n^{1/3} d(\hat{\beta}_n - \beta_0, \hat{F}_n - F_0) = O_p(1)$ .

### A.3 Proof of Theorem 2

Recall that  $I_\beta(\beta, F)$  is the score for  $\beta$ ,  $I_F(\beta, F)[g] = I(\beta, F + \lambda g) / \lambda|_{\lambda=0}$  is the score operator for  $F$  at direction  $g$ , with  $g(y) = \int_0^y h(x) dF(x)$  and  $h \in L_2^0(F)$ , where

$L_2^0(F) = \{h: \int h dF = 0, \int h^2 dF < \infty\}$ . Denote  $\mathcal{G} = \{g: g(\cdot) = \int h dF, h \in L_2^0(F)\}$ . Note that

$l_{\beta}^*(\beta_0, F_0) = l_{\beta}(\beta_0, F_0) - l_F(\beta_0, F_0)[\mathbf{g}^*]$  be the efficient score, where  
 $\mathbf{g}^* = (g_1^*, \dots, g_d^*)^T$  ( $d = \dim(\beta)$ ),

$$l_F(\beta_0, F_0)[\mathbf{g}^*] = (l_F(\beta_0, F_0)[g_1], \dots, l_F(\beta_0, F_0)[g_d])^T,$$

and  $\mathbf{g}^*$  is determined by

$$\begin{aligned} \mathbf{0} &= E\left((l_{\beta}(\beta_0, F_0) - l_F(\beta_0, F_0)[\mathbf{g}^*])l_F(\beta_0, F_0)[g]\right) \quad (\text{A.1}) \\ &= E\left(E\left[(\mathbf{Z}F_0(Y) - \mathbf{g}^*(Y))Q^2(\mathbf{X}; \beta_0, F_0) \mid Y\right]\right), \quad \forall g \in \mathcal{G}. \end{aligned}$$

This gives  $\mathbf{g}^*(Y) = F_0(Y)E[\mathbf{Z}Q^2(\mathbf{X}; \beta_0, F_0) \mid Y] / E[Q^2(\mathbf{X}; \beta_0, F_0) \mid Y]$ .

Let  $l_n(\beta, F) = \log L_n(\beta, F)$  be the log-likelihood,  $l_{\beta, n}(\beta, F) = l_n(\beta, F) / \beta$  and  $l_{F, n}(\beta, F)[g]$  be the sample score operator for  $F$  at direction  $g$ , and define  $l_{F, n}(\beta, F)[\mathbf{g}^*]$  similarly. Since  $(\hat{\beta}_n, \hat{F}_n)$  is the MLE, we have  $l_{\beta, n}(\hat{\beta}_n, \hat{F}_n) = \mathbf{0}$  and  $l_{F, n}(\hat{\beta}_n, \hat{F}_n)[\mathbf{g}^*] = \mathbf{0}$ . Additionally, it is obvious that  $Pl_{\beta}(\beta_0, F_0) = \mathbf{0}$  and  $Pl_F(\beta_0, F_0)[\mathbf{g}^*] = \mathbf{0}$ . Let  $\mathcal{M}_1 = \{l_{\beta}(\beta, F): \beta \in \mathcal{B}, F \in \mathcal{F}\}$  and  $\mathcal{M}_2 = \{l_F(\beta, F)[\mathbf{g}^*]: \beta \in \mathcal{B}, F \in \mathcal{F}\}$ . Since  $\mathcal{B}$  is bounded and  $\mathcal{F}$  is collection of bounded monotone functions, using the entropy computations in the proof of Theorem 1, it is easy to show that they are Donsker classes. It can be shown that for fixed  $X$ ,  $l_{\beta}(\beta, F)$  and  $l_F(\beta, F)$  are Lipschitz in  $(\beta, F)$ , so  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are also Donsker classes. Since by Theorem 1,  $\|\hat{\beta}_n - \beta_0\| + \sup_{t \in [\tau_l, t_0]} |\hat{F}_n - F_0| \rightarrow 0$  almost surely, condition (C1) and the given model imply that  $l_{\beta}(\beta, F)$  and  $l_F(\beta_0, F_0)[\mathbf{g}^*]$  are continuous in  $(\beta, F)$  and with finite second moments, thus by Corollary 2.3.12 in van der Vaart and Wellner (1996), we have

$$\left| \sqrt{n}(P_n - P)l_{\beta}(\hat{\beta}_n, \hat{F}_n) - \sqrt{n}(P_n - P)l_{\beta}(\beta_0, F_0) \right| = o_p(1)$$

where the  $o_p(1)$  is in the vector sense, and

$$\left| \sqrt{n}(P_n - P)l_F(\hat{\beta}_n, \hat{F}_n)[\mathbf{g}^*] - \sqrt{n}(P_n - P)l_F(\beta_0, F_0)[\mathbf{g}^*] \right| = o_p(1).$$

From these facts we get

$$-\sqrt{n}Pl_{\beta}(\hat{\beta}_n, \hat{F}_n) - \sqrt{n}P_n l_{\beta}(\beta_0, F_0) = o_p(1) \quad (\text{A.2})$$

and

$$-\sqrt{n}Pl_F(\hat{\beta}_n, \hat{F}_n)[\mathbf{g}^*] - \sqrt{n}Pl_F(\beta_0, F_0)[\mathbf{g}^*] = o_p(1). \quad (\text{A.3})$$

Denote  $l_{\beta, \beta}(\beta, F) = \kappa(\beta, F) \wedge (\beta \beta^T)$ ,

$$l_{\beta, F}(\beta, F)[g] = \partial l_{\beta}(\beta, \int (1 + \lambda g)dF) / \partial \lambda |_{\lambda=0}$$

and define  $l_{F, \beta}(\beta, F)[g]$  and  $l_{F, F}(\beta, F)[g_1, g_2]$  similarly. Using Taylor expansion, and note that  $\|\hat{\beta}_n - \beta_0\|^2 + \|\hat{F}_n - F_0\|^2$  is bounded, so by Lemma 1 and dominated convergence  $E\|\hat{\beta}_n - \beta_0\|^2 + \|\hat{F}_n - F_0\|^2 = O(n^{-2/3})$ , and so

$$\begin{aligned} & Pl_{\beta}(\hat{\beta}_n, \hat{F}_n) - Pl_{\beta}(\beta_0, F_0) - Pl_{\beta, \beta}(\beta_0, F_0)(\hat{\beta}_n - \beta_0) - Pl_{\beta, F}(\beta_0, F_0)(\hat{F}_n - F_0) \\ &= OE(\|\hat{\beta}_n - \beta_0\|^2 + \|\hat{F}_n - F_0\|^2) = O(n^{-2/3}). \end{aligned}$$

The above and (A.2) give

$$-P_n l_{\beta}(\beta_0, F_0) - Pl_{\beta, \beta}(\beta_0, F_0)(\hat{\beta}_n - \beta_0) - Pl_{\beta, F}(\beta_0, F_0)(\hat{F}_n - F_0) = o_p(n^{-1/2}). \quad (\text{A.4})$$

Similarly,

$$\begin{aligned} & Pl_F(\hat{\beta}_n, \hat{F}_n)[\mathbf{g}^*] - Pl_F(\beta_0, F_0)[\mathbf{g}^*] - Pl_{F, \beta}(\beta_0, F_0)[\mathbf{g}^*](\hat{\beta}_n - \beta_0) - Pl_{F, F}(\beta_0, F_0)[\mathbf{g}^*, \hat{F}_n - F_0] \\ & (\|\hat{\beta}_n - \beta_0\|^2 + \|\hat{F}_n - F_0\|^2) = O_p(n^{-2/3}). \end{aligned}$$

The above and (A.3) give

$$-P_n l_F(\beta_0, F_0)[\mathbf{g}^*] - Pl_{F, \beta}(\beta_0, F_0)[\mathbf{g}^*](\hat{\beta}_n - \beta_0) - Pl_{F, F}(\beta_0, F_0)[\mathbf{g}^*, \hat{F}_n - F_0] = o_p(n^{-1/2}).$$

(A.5)

It is known that  $P(l_{\beta}(\beta_0, F_0)l_{\beta}^T(\beta_0, F_0)) = -Pl_{\beta, \beta}(\beta_0, F_0)$ . By the same way,  $-Pl_{\beta, F}(\beta_0, F_0)[g] = Pl_{\beta}(\beta_0, F_0)l_F(\beta_0, F_0)[g]$  for all  $g$ , and  $-Pl_{F, F}(\beta_0, F_0)[g_1, g_2] = Pl_F(\beta_0, F_0)[g_1]l_F(\beta_0, F_0)[g_2]$  for all  $g_1$  and  $g_2$ . Note that (A.1) holds for any  $g$ . Thus,

$$\begin{aligned}
& P l_{\beta, F}(\boldsymbol{\beta}_0, F_0)[\hat{F}_n - F_0] - P l_{F, F}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*, \hat{F}_n - F_0] \\
&= -P(l_{\beta, F}(\boldsymbol{\beta}_0, F_0) l_{F, F}(\boldsymbol{\beta}_0, F_0)[\hat{F}_n - F_0]) + P(l_{F, F}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*] l_{F, F}(\boldsymbol{\beta}_0, F_0)[\hat{F}_n - F_0]) \\
&= -P\left((l_{\beta, F}(\boldsymbol{\beta}_0, F_0) - l_{F, F}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*]) l_{F, F}(\boldsymbol{\beta}_0, F_0)[\hat{F}_n - F_0]\right) = \mathbf{0}.
\end{aligned}$$

Now, subtracting (A.4) from (A.5) we get

$$\sqrt{n}P(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*])(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \sqrt{n}P_n(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*]) + o_p(1).$$

(A.6)

Note that, since  $P(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*]) l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*] = \mathbf{0}$ , so

$$\begin{aligned}
& P(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*]) = -P(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0)^T - l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*]^T) \\
&= -P\left(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0)(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*])^T\right) \\
&= -P\left((l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*])(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0) - l_{F, \beta}(\boldsymbol{\beta}_0, F_0)[\mathbf{g}^*])^T\right) \\
&= -P\left(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0)(l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0))^T\right) = -l_{\beta, \beta}(\boldsymbol{\beta}_0, F_0).
\end{aligned}$$

This and (A.6) give

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = I_{\beta, \beta}^{-1}(\boldsymbol{\beta}_0, F_0) \sqrt{n}P_n l_{\beta, \beta}^*(\boldsymbol{\beta}_0, F_0) + o_p(1),$$

which gives the desired result.

#### A.4 Proof of Theorem 3

Since  $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)$  is the MLE, we have

$$\hat{F}_n = \arg \max_{F \in \mathcal{F}} \sum_{i=1}^n \left( \Delta_i \log \left[ 1 - G_{\gamma} \{ e^{\hat{\boldsymbol{\beta}}_n^T \mathbf{Z}_i F(Y_i)} \} \right] + (1 - \Delta_i) \log \left[ G_{\gamma} \{ e^{\hat{\boldsymbol{\beta}}_n^T \mathbf{Z}_i F(Y_i)} \} \right] \right).$$

Let  $R(y|\mathbf{z}, F) = 1 - G_{\gamma} \{ \hat{\boldsymbol{\beta}}^T \mathbf{z} F(y) \}$ , then  $R(\cdot|\mathbf{z}, F)$  is a distribution function. Let  $\mathcal{R} = \{R(\cdot|\mathbf{z}, F) : F \in \mathcal{F}\}$ , and  $\hat{R}_n(\cdot|\mathbf{z})$  be the MLE of  $R(\cdot|\mathbf{z}, F)$ , i.e.,

$$\hat{R}_n = \arg \max_{R \in \mathcal{R}} \sum_{i=1}^n (\Delta_i \log R(Y_i | \mathbf{Z}_i, F) + (1 - \Delta_i) \log [1 - R(Y_i | \mathbf{Z}_i, F)])$$

and it is straightforward that  $\hat{R}_n(\cdot | \mathbf{z}) = R(y | \mathbf{z}, \hat{F}_n)$ .

Recall that the common current status model with log-likelihood

$$\sum_{i=1}^n (\Delta_i \log F(Y_i) + (1 - \Delta_i) \log [1 - F(Y_i)])$$

and it is known (e.g., see Example 3.2.15 in van der Vaart and Wellner (1996)) that the NPMLE  $\check{F}_n(\cdot)$  of  $F_0(\cdot)$  is the slope function of the greatest convex minorant of  $F_n(\cdot)$ , which is the cumulative-sum diagram on  $[0, 1]$  with jump  $\Delta_i/n$  at  $Y_i$  i.e.

$$F_n(t) = \frac{1}{n} \sum_{j \leq i} \Delta_j, \quad t \in n^{-1}(i-1, i], \quad (i = 1, \dots, n).$$

and  $\check{F}_n(\cdot)$  is the slope of greatest convex minorant of  $F_n(\cdot)$ , and for  $a \in R$ ,

$$\check{F}_n(Y_i) \leq a \text{ if and only if } \arg \min_s \{ \check{V}_n(s) - a \check{G}_n(s) \} \geq Y_i,$$

where  $\check{V}_n(s) = n^{-1} \sum_{i=1}^n \Delta_i I(Y_i \leq s)$ , and  $\check{G}_n(s) = n^{-1} \sum_{i=1}^n I(Y_i \leq s)$ .

In our case, similarly,  $\hat{R}_n(\cdot | \mathbf{Z})$  is the greatest convex minorant of  $R_n(\cdot | \mathbf{Z})$ , cumulative-sum diagram on  $[0, 1]$ , where the latter has constant jump of  $\Delta_i/n$ . Thus the corresponding  $F_n(\cdot)$  in this case must be a weighted cumulative-sum diagram with weight  $w_i$  at the  $Y_i$ 's, where the  $w_i$ 's are determined by  $w_i \geq 0, \sum_{i=1}^n w_i = n$ . Then  $\hat{F}_n(Y_i) = \sum_{j \leq i} w_j/n = W_i/n$ . Assume the  $e^{\hat{\beta}_n^T \mathbf{Z}_i} w_i$ 's are arranged in an increasing order. With (C6),  $W_n = e^{-\hat{\beta}_n^T \mathbf{Z}_n} G_\gamma^{-1}(1/n)$  with  $0 < \lim_n W_n/n < \infty$  in probability. Since  $W_i = 0$  for  $i = 0$ , we can arrange the  $W_i$ 's such that  $\{w_i/n = (W_{i+1} - W_i)/n : i = 1, \dots, n\}$  are probability weights. Then as for the common current status model, the log-likelihood in our case will be maximized with the  $W_i$ 's satisfying

$$W_1 = 0, \quad 1 - G_\gamma \left( e^{\hat{\beta}_n^T \mathbf{Z}_i + 1} W_{i+1} \right) = 1 - G_\gamma \left( e^{\hat{\beta}_n^T \mathbf{Z}_i} W_i \right) + \frac{\Delta_i}{n}, \quad (i = 1, \dots, n-1).$$

Thus,  $e^{\hat{\beta}_n^T \mathbf{Z}_i} W_i = G_\gamma^{-1}(1 - i/n)$ , and

$$w_i = G_\gamma^{-1} \left(1 - \frac{i+1}{n}\right) e^{\hat{\beta}_n^T \mathbf{Z}_{i+1}} - G_\gamma^{-1} \left(1 - \frac{i}{n}\right) e^{\hat{\beta}_n^T \mathbf{Z}_i}.$$

In our case, the above equality may not be achieved for some of the  $i$ 's.

Let

$$F_n(y) = \frac{1}{n} \sum_{j \leq i} \Delta_i w_i, \quad y \in n^{-1}(i-1, i], \quad (i = 1, \dots, n).$$

then  $\hat{F}_n(\cdot)$  is the slope of greatest convex minorant of  $F_n(\cdot)$ , and for  $a \in R$ ,

$$\hat{F}_n(Y_i) \leq a \text{ if and only if } \arg \min_s \{V_n(s) - aG_n(s)\} \geq Y_i,$$

where  $V_n(s) = n^{-1} \sum_{i=1}^n \Delta_i w_i I(Y_i \leq s)$  and  $G_n(s) = n^{-1} \sum_{i=1}^n I(Y_i \leq s)$ .

Then as in Example 3.2.15 in van der Vaart and Wellner (1996), at every fixed point  $t$ ,

$$\hat{F}_n(t) \leq a \text{ if and only if } \arg \min_s \{V_n(s) - aG_n(s)\} - t \geq 0.$$

To evaluate the distribution of  $n^{1/3}(\hat{F}_n(t) - F_0(t))$ , we need to compute the probability of the event  $\{n^{1/3}(\hat{F}_n(t) - F_0(t)) \leq r\}$  for each  $r \in R$ . Take  $a = F_0(t) + nr^{-1/3}$ , then  $\{n^{1/3}(\hat{F}_n(t) - F_0(t)) \leq r\} = \{\hat{F}_n(t) \leq a\} = \{n^{1/3}(\arg \min_s \{V_n(s) - aG_n(s)\} - t) \leq 0\}$ , and by the change of variable  $s \mapsto t + n^{-1/3}h$ ,

$$\begin{aligned}
 & n^{1/3} \left( \arg \min_s \{V_n(s) - aG_n(s)\} - t \right) \\
 &= \arg \min_h \{V_n(t + n^{-1/3}h) - (F_0(t) + rn^{-1/3})G_n(t + n^{-1/3}h)\} \\
 &= \arg \min_h \{P_n(\Delta w I(Y \leq t + n^{-1/3}h)) - (F_0(t) + rn^{-1/3})P_n I(Y \leq t + n^{-1/3}h)\} \\
 &= \arg \min_h \{P_n([\Delta w - F_0(t)]I[Y \leq t + n^{-1/3}h]) - rn^{-1/3}P_n[I(Y \leq t + n^{-1/3}h)]\} \\
 &= \arg \min_h \{P_n([\Delta w - F_0(t)]I[t \leq Y \leq t + n^{-1/3}h]) + P_n([\Delta w - F_0(t)]I(Y \leq t]) - rn^{-1/3}P_n \\
 & \quad (I(t \leq Y \leq t + n^{-1/3}h)) - rn^{-1/3}P_n I(Y \leq t)\} \\
 &= \arg \min_h \{P_n([\Delta w - F_0(t)]I[t \leq Y \leq t + n^{-1/3}h]) - rn^{-1/3}P_n I[t \leq Y \leq t + n^{-1/3}h]\} \\
 &= \arg \min_h \{n^{2/3}P_n([\Delta w - F_0(t)]I[t \leq Y \leq t + n^{-1/3}h]) - rn^{1/3}P_n I[t \leq Y \leq t + n^{-1/3}h]\} \\
 &= \arg \min_h \{n^{2/3}(P_n - P)([\Delta w - F_0(t)]I[t \leq Y \leq t + n^{-1/3}h]) + n^{2/3}P([\Delta w - F_0(t)]I[t \leq Y \\
 & \quad \leq t + n^{-1/3}h]) - rn^{1/3}P_n I[t \leq Y \leq t + n^{-1/3}h]\} \\
 &:= \arg \min_h \{B_{1,n}(h) + B_{2,n}(h) + B_{3,n}(h)\}.
 \end{aligned}$$

(A.6)

Let  $g(\cdot)$  be the density function of  $Y$ ,

$$B_{3,n}(h) \sim -rn^{1/3}P[I(t \leq Y \leq t + n^{-1/3}h)] \sim -rn^{1/3}g(t)n^{-1/3}h = -rg(t)h.$$

As a random variable, given the  $\mathbf{Z}_i$ 's,  $w$  takes

$$e^{-\hat{\beta}_n^T \mathbf{Z}_i} G_\gamma^{-1}(1 - i/n) - e^{-\hat{\beta}_n^T \mathbf{Z}_{i+1}} G_\gamma^{-1}(1 - (i+1)/n) \text{ with probability } 1/n. \text{ Without confusion,}$$

we also denote  $w = w(\mathbf{Z})$  for its asymptotic version conditioning on  $\mathbf{Z}$ . Note that  $Y$  and  $(T, \mathbf{Z})$  are independent, and  $\|\hat{\beta}_n - \beta_0\| = O_p(n^{-1/2})$ , and with condition (C6), we show that

$$P(w) = E[e^{-\beta_0^T \mathbf{Z}}] \sum_{i=1}^n \frac{1}{n} \left( G_\gamma^{-1}\left(1 - \frac{i+1}{n}\right) - G_\gamma^{-1}\left(1 - \frac{i}{n}\right) \right) = E[e^{-\beta_0^T \mathbf{Z}}] \frac{G_\gamma^{-1}\left(\frac{1}{n}\right)}{n} - 1.$$

and

$$\begin{aligned}
 & B_{2,n}(h) \sim n^{2/3}P([\Delta - F_0(t)]I[t \leq Y \leq t + n^{-1/3}h]) \\
 &= n^{2/3} \int_t^{t+n^{-1/3}h} (F_0(y) - F_0(t))g(y)dy - n^{2/3}f_0(t)g(t) \int_t^{t+n^{-1/3}h} (y-t)dy \\
 &= n^{2/3}f_0(t)g(t)n^{-2/3} \frac{1}{2}h^2 = \frac{1}{2}f_0(t)g(t)h^2.
 \end{aligned}$$

Now we evaluate  $B_{1,n}(h)$ . For  $h < 0$ , the notation  $I(t - Y \leq t + hn^{-1/3})$  means  $I(t + hn^{-1/3} - Y \leq t)$ . Let

$$f_{n,h}(Y_i, \mathbf{Z}_i) = n^{1/6}(\Delta w(\mathbf{Z}_i) - F_0(t))I(t \leq Y_i \leq t + hn^{-1/3}).$$

Then

$$B_{1,n}(h) = n^{-1/2} \sum_{i=1}^n [f_{n,h}(Y_i, \mathbf{Z}_i) - Pf_{n,h}(Y_i, \mathbf{Z}_i)].$$

Let  $\mathcal{F}_n = \{f_{n,h}(\cdot, \cdot) : |h| \leq K\}$ , for some  $K > 0$ . By (C1) and (C4), the  $w(\mathbf{Z}_i)$ 's are uniformly bounded, thus  $\mathcal{F}_n$  has an envelope function  $F_n(y, x, z) = n^{1/6}CK|t - Kn^{-1/3} - Y \leq t + Kn^{-1/3}$ , with some  $0 < C < \infty$ . Obviously,

$$PF_n^2(Y, \mathbf{Z}) = O(1), \quad P[F_n^2 I(F_n > \eta\sqrt{n})] \rightarrow 0, \quad \forall \eta > 0,$$

and for some  $0 < C < \infty$ ,

$$\begin{aligned} P(f_{n,s} - f_{n,h})^2 &= n^{1/3} E[(\Delta_i w(\mathbf{Z}_i) - F_0(t))^2 I(t + (s \wedge h)n^{-1/3} \leq Y_i \leq t + (s \vee h)n^{-1/3})] \\ &\sim Cn^{1/3} \int_{t + (s \wedge h)n^{-1/3}}^{t + (s \vee h)n^{-1/3}} h g(y) dy \sim O(1) |s - h|. \end{aligned}$$

It follows that for any totally bounded semimetric  $\rho(\cdot, \cdot)$  on  $[-K, K]$ ,

$$\sup_{(s,h): \rho(s,h) \leq \delta_n} P(f_{n,s} - f_{n,h})^2 \sim O(1) |\delta_n| \rightarrow 0, \quad \forall \delta_n \rightarrow 0.$$

Thus the three conditions in Theorem 2.11.21 of van der Vaart and Wellner (1996) are satisfied.

Let  $\mathcal{F}_{n,\delta} = \{f_{n,s} - f_{n,h} : \rho(s, h) \leq \delta\}$ . Then  $\mathcal{F}_{n,\delta}$  and  $\mathcal{F}_{n,\delta}^2$  are  $P$ -measurable in the sense of Definition 2.33 in van der Vaart and Wellner (1996). Furthermore, for fixed  $h$ ,

$$\begin{aligned} Pf_{n,h} &= n^{1/6} E[(\Delta w(\mathbf{Z}) - F_0(t))I(t \leq Y \leq t + hn^{-1/3})] \\ &\sim n^{1/6} E[\{(\Delta - F_0(t)) + o(1)\}I(t \leq Y \leq t + hn^{-1/3})] \\ &= O(n^{1/6} n^{-2/3} h^2) + o(1) n^{1/6} g(t) n^{-1/3} h = O(n^{-1/2} h^2) + o(n^{-1/6} h) \rightarrow 0, \end{aligned}$$



and

$$\begin{aligned} Pf_{n,s}f_{n,h} &= n^{1/3}E\left[(\Delta w(\mathbf{Z}) - F_0(t))^2 I(t \leq Y \leq t + (s \wedge h)n^{-1/3})\right] \\ &= n^{1/3} \int_t^{t+(s \wedge h)n^{-1/3}} E\left[(I(T \leq Y)w(\mathbf{Z}) - F_0(t))^2 \mid Y = y\right]g(y)dy \\ &\sim n^{1/3}E\left[(I(T \leq t)w(\mathbf{Z}) - F_0(t))^2\right]g(t)(s \wedge h)n^{-1/3} = \eta^2(t)g(t)(s \wedge h). \end{aligned}$$

Therefore as  $n \rightarrow \infty$ ,  $(Pf_{n,s}f_{n,h} - Pf_{n,s}Pf_{n,h}) \rightarrow \eta^2(t)g(t)(s \wedge h)$ . Note that  $E[\mathbb{B}(s) - \mathbb{B}(h)]^2 = |s - h|$ , where  $\mathbb{B}(\cdot)$  is the two-sided Brownian motion process, originating from zero. It follows that  $Pf_{n,s}f_{n,h} - Pf_{n,s}Pf_{n,h}$  converges to the covariance function of the process  $g^{1/2}(t)\eta(t)\mathbb{B}(\cdot)$ .

For a probability measure  $Q$ , let  $N(\epsilon, \mathcal{F}_n, L_2(Q))$  be the number of  $\epsilon$ -balls needed to cover  $\mathcal{F}_n$  under the metric of  $L_2(Q)$ . It is easy to see that

$$\sup_Q N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q)) = O(1/\epsilon),$$

where the supreme is over all probability measures. Thus,  $\forall \eta_n \rightarrow 0$ ,

$$\begin{aligned} &\sup_Q \int_0^{\eta_n} \sqrt{N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q))} d\epsilon \\ &\sim \int_0^{\eta_n} \sqrt{-\log \epsilon} d\epsilon = 2 \int_{\sqrt{-\log \eta_n}}^{\infty} x e^{-x^2} dx \rightarrow 0. \end{aligned}$$

By Theorem 2.11.22 in van der Vaart and Wellner (1996), we obtain

$$B_{1,n}(\cdot) \xrightarrow{D} \sqrt{\eta^2(t)g(t)\mathbb{B}(\cdot)}, \quad \text{in } l^\infty([-K, K]),$$

where  $\mathcal{F}^\infty([-K, K])$  is the space of all bounded real functions on  $[-K, K]$  equipped with the supremum metric.

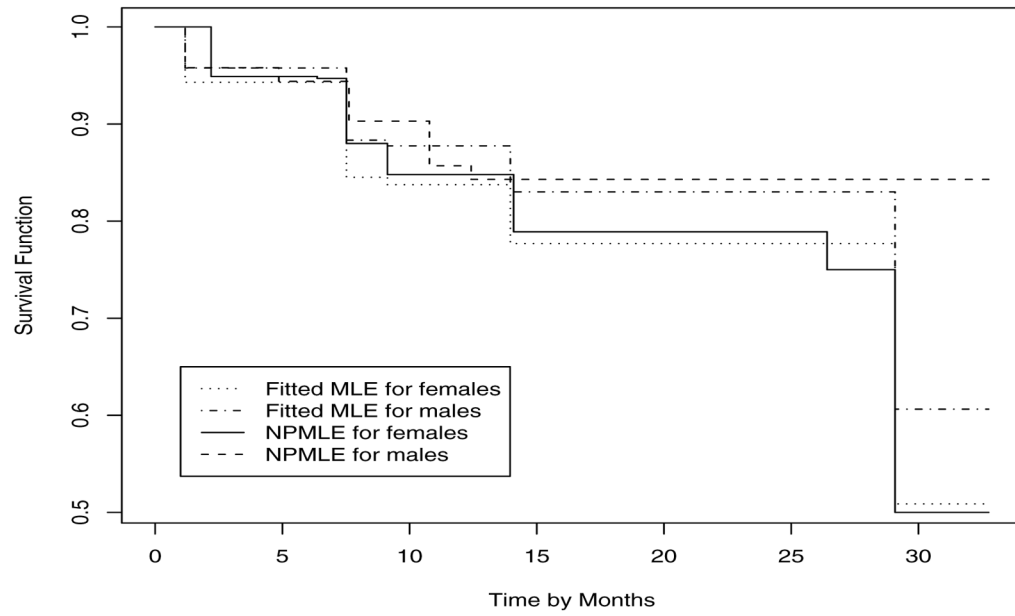
Now collecting results from (A.6), we have that the event  $\{n^{1/3}(\hat{F}_n(t) - F_0(t)) - t\}$  is asymptotically equivalent to

$$\begin{aligned} & \left\{ \arg \min_h \left\{ \sqrt{\eta^2(t)g(t)}\mathbb{B}(h) + \frac{1}{2}f_0(t)g(t)h^2 - rg(t)h \right\} \geq 0 \right\} \\ &= \left\{ \arg \max_h \left\{ -\sqrt{\frac{\eta^2(t)}{g(t)}}\mathbb{B}(h) - \frac{1}{2}f_0(t)h^2 + rh \right\} \geq 0 \right\} \\ &= \left\{ \arg \max_h \left\{ \sqrt{\frac{\eta^2(t)}{g(t)}}\mathbb{B}(h) - \frac{1}{2}f_0(t)h^2 + rh \right\} \geq 0 \right\}. \end{aligned}$$

In the above we used the fact that  $-\mathbb{B}(h) \stackrel{D}{=} \mathbb{B}(h)$ . Using problem 3.2.5 in van der Vaart and Wellner (1996), the above is re-written as

$$\begin{aligned} & \left\{ \left( \frac{4\eta^2(t)f_0(t)}{g(t)} \right)^{1/3} \arg \max_h \{ \mathbb{B}(h) - h^2 \} + r \geq 0 \right\} \\ &= \left\{ \left( \frac{4\eta^2(t)f_0(t)}{g(t)} \right)^{1/3} \arg \min_h \{ -\mathbb{B}(h) + h^2 \} \geq -r \right\} \\ &= \left\{ \left( \frac{4\eta^2(t)f_0(t)}{g(t)} \right)^{1/3} \arg \min_h \{ \mathbb{B}(h) + h^2 \} \geq -r \right\} \\ &= \left\{ \left( \frac{4\eta^2(t)f_0(t)}{g(t)} \right)^{1/3} \arg \min_h \{ \mathbb{B}(h) + h^2 \} \leq r \right\}, \quad \forall t, r. \end{aligned}$$

In the above we used the fact that  $-\mathbb{B}(\cdot)$  and  $\mathbb{B}(\cdot)$  have the same distribution, and that  $W = \arg \min_h \{ \mathbb{B}(h) + h^2 \}$  is symmetrically distributed about 0, thus  $P\{CW \leq -b\} = P\{CW \geq b\}$ . This gives the desired result for  $h \in [-K, K]$ . Below we prove that the result is actually true on  $\mathcal{R}$ . Let  $\hat{h}_n = \arg \min_h \{ V_n(h) - aG_n(h) \} - t$ . We need to show that  $\hat{h}_n$  is bounded in probability. Thus for large  $n$ ,  $\hat{h}_n$  will be in  $[-K, K]$  for some  $0 < K < \infty$  in probability, and so the desired result is true on  $\mathcal{R}$ . For this, we only need to show  $n^{1/3}d(\hat{h}_n, \hat{h}) = O_p(1)$  for some distance  $d(\cdot, \cdot)$ . The method is similar to that in the proof of Lemma 1 and is omitted.



**Fig. 1.** Nonparametric estimates and proportional odds cure model-fitted survival functions of time to IOL calcification. The model-fitted survival function is calculated as the empirical average of the predicted survival functions for each gender group.

**Table 1**

Summary Statistics for the Maximum Likelihood Estimators

Model	$n$	Parameter	True value	Estimate	SE	SEE	CP(%)
$\gamma = 0$	200	$\beta_1$	1	1.047	0.436	0.427	94.7
		$\beta_2$	-0.5	-0.538	0.249	0.245	94.9
	400	$\beta_1$	1	1.029	0.302	0.294	94.1
		$\beta_2$	-0.5	-0.522	0.168	0.169	94.9
$\gamma = 0.25$	200	$\beta_1$	1	1.048	0.481	0.471	95.0
		$\beta_2$	-0.5	-0.540	0.275	0.270	93.3
	400	$\beta_1$	1	1.031	0.334	0.325	94.5
		$\beta_2$	-0.5	-0.523	0.183	0.187	95.2
$\gamma = 0.5$	200	$\beta_1$	1	1.044	0.525	0.511	94.4
		$\beta_2$	-0.5	-0.54	0.298	0.293	94.3
	400	$\beta_1$	1	1.026	0.367	0.353	94.0
		$\beta_2$	-0.5	-0.524	0.199	0.203	95.7
$\gamma = 0.75$	200	$\beta_1$	1	1.048	0.569	0.550	93.8
		$\beta_2$	-0.5	-0.541	0.321	0.315	95.4
	400	$\beta_1$	1	1.030	0.395	0.379	94.1
		$\beta_2$	-0.5	-0.525	0.215	0.218	95.0
$\gamma = 1$	200	$\beta_1$	1	1.054	0.602	0.587	94.2
		$\beta_2$	-0.5	-0.547	0.344	0.336	95.1
	400	$\beta_1$	1	1.032	0.417	0.404	94.4
		$\beta_2$	-0.5	-0.529	0.234	0.232	94.3

**Table 2**

Simulation Results Under Misspecified Transformation

Model	Parameter	True value	Estimate	SE	SEE	CP(%)
True transformation: $G(x) = e^{-x}$						
Proportional odds model	$\beta_1$	1	1.396	0.415	0.405	84.6
	$\beta_2$	-0.5	-0.706	0.228	0.231	86.7
True transformation: $G(x) = (1+x)^{-1}$						
Proportional hazards model	$\beta_1$	1	0.806	0.325	0.316	89.7
	$\beta_2$	-0.5	-0.414	0.185	0.183	92.9
True transformation: $G(x) = (1+x/2)^{-2}$						
Proportional hazards model	$\beta_1$	1	0.889	0.317	0.305	92.1
	$\beta_2$	-0.5	-0.455	0.174	0.176	95.2
Proportional odds model	$\beta_1$	1	1.168	0.420	0.402	92.6
	$\beta_2$	-0.5	-0.595	0.226	0.230	93.9

**Table 3**

Maximum Likelihood Estimates of Regression Coefficients in the Proportional Odds Cure Model for the Calcification Data

Covariate	Estimate	SE	Est/SE	<i>p</i> -value
Intercept	0.354	0.075	4.689	<0.001
Incision length	0.447	0.260	1.719	0.086
Gender (male)	-0.310	0.343	-0.906	0.365
Age/10	-0.136	0.216	-0.626	0.531

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript