

# A General Transformation Class of Semiparametric Cure Rate Frailty Models

Guoqing Diao · Guosheng Yin

Received: date / Accepted: date

**Abstract** We consider a class of cure rate frailty models for multivariate failure time data with a survival fraction. This class is formulated through a transformation on the unknown population survival function. It incorporates random effects to account for the underlying correlation, and includes the mixture cure model and the proportional hazards cure model as two special cases. We develop efficient likelihood-based estimation and inference procedures. We show that the nonparametric maximum likelihood estimators for the parameters of these models are consistent and asymptotically normal, and that the limiting variances achieve the semiparametric efficiency bounds. Simulation studies demonstrate that the proposed methods perform well in finite samples. We provide an application of the proposed methods to the data of the age at onset of alcohol dependence, from the Collaborative Study on the Genetics of Alcoholism.

**Keywords** Box-Cox transformation · Cure fraction · Empirical process · Mixture cure model · NPMLE · Proportional hazards cure model · Semiparametric efficiency

---

G. Diao  
Department of Statistics, George Mason University MS 4A7, 4400 University Drive, Fairfax, VA 22030, USA  
E-mail: gdiao@gmu.edu

G. Yin  
Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong  
E-mail: gyin@hku.hk

## 1 Introduction

Cure rate models, which are used for modeling time-to-event data incorporating a survival fraction, have become increasingly important in biomedical and genetic studies. The commonly used cure rate models include the mixture cure model and the proportional hazards cure model. To introduce these models, we let  $S_{\text{pop}}(t|\mathbf{Z}_i, \mathbf{X}_i)$  be an improper population survival function,  $\lim_{t \rightarrow \infty} S_{\text{pop}}(t|\mathbf{Z}_i, \mathbf{X}_i) > 0$ , and  $S(t|\mathbf{X}_i)$  be a proper survival function,  $\lim_{t \rightarrow \infty} S(t|\mathbf{X}_i) = 0$ , where  $\mathbf{Z}_i$  and  $\mathbf{X}_i$  are two covariate vectors for subject  $i$  ( $i = 1, \dots, n$ ). Note that  $\mathbf{Z}_i$  includes 1 and may share common components with  $\mathbf{X}_i$ . The mixture cure model (Berkson and Gage, 1952) is composed of a certain fraction of the population that will be cured or that is not susceptible to the event of interest,  $1 - \theta(\mathbf{Z}_i)$ , and a remaining proportion that will not be cured or that is susceptible to the event of interest,  $\theta(\mathbf{Z}_i)$ , such that

$$S_{\text{pop}}(t|\mathbf{Z}_i, \mathbf{X}_i) = 1 - \theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i)S(t|\mathbf{X}_i), \quad (1)$$

where  $S(t|\mathbf{X}_i)$  is the survival function for the uncured/susceptible population. A logistic regression formulation is usually assumed for  $\theta(\mathbf{Z}_i)$  so that

$$\theta(\mathbf{Z}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}.$$

The mixture cure model (1) has been extensively studied in the literature, which includes the work of Gray and Tsiatis (1989), Kuk and Chen (1992), Taylor (1995), Sy and Taylor (2000), Peng and Dear (2000), and Betensky and Schoenfeld (2001), among others. A comprehensive discussion of the mixture cure model is given by Maller and Zhou (1996).

The proportional hazards cure model, an alternative definition of a cure rate model, has been proposed and investigated by Yakovlev et al. (1993), Tsodikov (1998), Chen et al. (1999) and Tsodikov et al. (2003), among others. Its population survival function is given by

$$S_{\text{pop}}(t|\mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)F(t)\}, \quad (2)$$

where  $F(t)$  is a distribution function and  $\theta(\mathbf{Z}_i) = \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)$ . The hazard function corresponding to (2) is  $\lambda_{\text{pop}}(t|\mathbf{Z}_i) = \theta(\mathbf{Z}_i)f(t)$ , where  $f(t) = dF(t)/dt$ . The cure rate for subject  $i$  under model (2) is

---

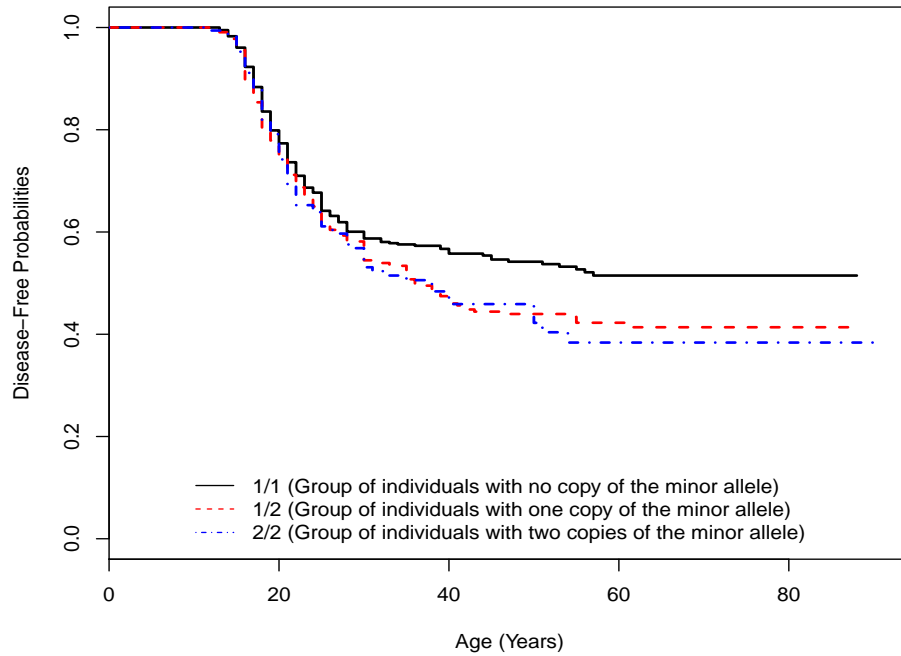
$\lim_{t \rightarrow \infty} S_{\text{pop}}(t|\mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)\}$ . In more general settings, Yin and Ibrahim (2005), Zeng et al. (2006), and Cooner et al. (2007) proposed several general classes of cure rate models that include the mixture, the proportional hazards, and the proportional odds cure models as special cases.

A critical assumption common to the aforementioned methods is the independence of the survival times; however, this may not be true in practical applications. In biomedical research, we often encounter multivariate failure time data for which the correlation may be induced by natural or artificial clustering effects. Such examples include experiments with litter-matched mice, family studies of genetic diseases, or ophthalmologic research. In these cases, the underlying correlation needs to be adjusted to ensure valid estimation and inference. A natural way to account for the underlying dependence is to introduce a frailty that is specific to each cluster. For subject  $j$  in cluster  $i$  with a covariate vector  $\mathbf{X}_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, n_i$ ), the usual Cox-type shared frailty model is given by

$$\lambda(t|\mathbf{X}_{ij}, W_i) = \lambda(t)W_i \exp(\boldsymbol{\beta}^T \mathbf{X}_{ij}), \quad (3)$$

where  $\lambda(t)$  is the unknown and unspecified baseline hazard function and  $W_i$  is the unobservable frailty induced by the  $i$ th cluster. Conditional on  $W_i$ , the failure times in cluster  $i$  are assumed to be independent. Typically, the  $W_i$ 's are assumed to be independent and identically distributed (i.i.d.) from a gamma distribution with mean one (Clayton, 1978). If a positive stable distribution is assumed for  $W_i$ , the proportional hazards structure would be preserved after integrating out  $W_i$ . Recently, Zeng and Lin (2007) and Zeng et al. (2008) proposed a general class of semiparametric transformation models with random effects for clustered failure time data. This class includes proportional hazards and proportional odds models as special cases and accommodates a variety of random-effects distributions.

Limited research has been conducted on cure rate models with multivariate failure time data. Chatterjee and Shih (2001) proposed a marginal approach using bivariate copula models. Yau and Ng (2001) and Price and Manatunga (2001) imposed frailty to account for correlation under parametric model assumptions. Chen et al. (2002) proposed a multivariate proportional hazards cure model using a positive stable frailty in the Bayesian framework. More recently, Peng et al. (2007) and Yu and Peng (2008) proposed a marginal regression



**Fig. 1** Kaplan-Meier estimates of the disease-free probabilities stratified by the genotype of SNP rs1972373 on chromosome 14 for the age at onset of alcohol dependence in the COGA data.

approach by using sandwich variance estimators and one-step jackknife variance estimators, respectively, without theoretical justifications. All the above methods except Chen et al. (2002) were based on the mixture cure model (1). An interesting example involves data from the Collaborative Study on the Genetics of Alcoholism (COGA, Begleiter et al., 1995). COGA is a nine-site national collaboration with the goal of identifying genetic factors that affect susceptibility to alcohol dependence and characterizing the related phenotypes. The data provided at the Genetic Analysis Workshop 14 (Bailey-Wilson et al., 2005), represented a total of 1,614 individuals in 143 multi-generation families, with family sizes ranging from 5 to 32. As it is known that alcoholism has a genetic component, the observations from members of the same family were considered to be dependent. In the application of our model, we examined the age at onset of alcohol dependence, and set the ages at interview to be the censoring times for the unaffected individuals. Among the 1,614 individuals in the study, 643 were affected with alcoholism, 626 of whom had known ages at onset.

Thus, the final data set for our analysis consisted of 1,371 individuals without missing genotype data: 626 who were affected with alcoholism and 745 who were unaffected. Of the 626 affected individuals, 424 were males; and of the 745 unaffected individuals, 229 were males. Figure 1 depicts the Kaplan-Meier survival curves for the alcoholism age-at-onset data for each genotype group at SNP rs1972373 on chromosome 14. We can see that a stable plateau is reached at the tail of each survival curve, which indicates that a cure rate model would be suitable for the data.

To enhance model flexibility, we consider a general class of cure rate frailty models that contains the mixture and proportional hazards cure models as two special cases. This class of models is built through a transformation (Box and Cox, 1964) on the population survival function. The Box-Cox transformation in linear models improves the normality of the errors, under which the transformed response variable is defined as  $Y^{(\alpha)} = (Y^\alpha - 1)/\alpha$  if  $\alpha \neq 0$ , and  $\log(Y)$  if  $\alpha = 0$ . In our case, by adding an extra transformation parameter, the two main formulations of cure rate models are unified together and the resulting model structure allows for a much richer class of cure rate structures. In the Bayesian paradigm, Yin (2008) investigated a different class of cure frailty models which involves the proportional odds model as a special case, but not the popular mixture cure structure.

The rest of this article is organized as follows. In Section 2, we introduce the notation and the class of cure rate frailty models based on the transformed population survival function. In Section 3, we describe the model assumptions, formulate the likelihood function and derive the asymptotic theories. We present simulation studies to examine the finite sample properties of the proposed method in Section 4, and illustrate the proposed methodology through the analysis of COGA data in Section 5. We conclude with a brief discussion in Section 6.

## 2 A Class of Cure Rate Frailty Models

Suppose that there are  $n$  independent clusters, and  $n_i$  subjects within cluster  $i$ . For  $i = 1, \dots, n$ , and  $j = 1, \dots, n_i$ , we observe  $\{Y_{ij} = \min(T_{ij}, C_{ij}), \Delta_{ij} = I(T_{ij} \leq C_{ij}), \mathbf{Z}_{ij}, \mathbf{X}_{ij}\}$ , where  $T_{ij}$  is the failure time for member  $j$  in cluster  $i$ ,  $C_{ij}$  is the censoring variable,  $\mathbf{Z}_{ij}$  and  $\mathbf{X}_{ij}$  are  $d_1$ - and  $d_2$ -dimensional vectors

of bounded covariates, respectively. The first component of  $\mathbf{Z}_{ij}$  is 1, and  $\mathbf{Z}_{ij}$  and  $\mathbf{X}_{ij}$  may share common components. The right-censoring time  $C_{ij}$  is assumed to be conditionally independent of  $T_{ij}$  given  $\mathbf{Z}_{ij}$  and  $\mathbf{X}_{ij}$  and has a finite hazard rate almost everywhere.

Define

$$G_\alpha(x) = \begin{cases} (1 + \alpha x)^{1/\alpha} & \text{if } \alpha > 0 \\ e^x & \text{if } \alpha = 0, \end{cases}$$

and let the  $W_i$ 's be i.i.d. gamma random variables with mean one and variance  $\eta$ . We consider a class of cure rate frailty models

$$S_{\text{pop}}(t|\mathbf{Z}_{ij}, \mathbf{X}_{ij}, W_i) = G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)F(t|\mathbf{X}_{ij}, W_i)). \quad (4)$$

When  $\alpha \rightarrow 0$ , model (4) becomes  $\log\{S_{\text{pop}}(t|\mathbf{Z}_{ij}, \mathbf{X}_{ij}, W_i)\} = -\theta(\mathbf{Z}_{ij}|0, W_i)F(t|\mathbf{X}_{ij}, W_i)$ , and thus reduces to the proportional hazards cure model; and when  $\alpha = 1$ , it has the mixture modeling structure,

$$S_{\text{pop}}(t|\mathbf{Z}_{ij}, \mathbf{X}_{ij}, W_i) = 1 - \theta(\mathbf{Z}_{ij}|1, W_i)F(t|\mathbf{X}_{ij}, W_i),$$

where  $F(t|\mathbf{X}_{ij}, W_i) = 1 - S(t|\mathbf{X}_{ij}, W_i)$ . The corresponding cure rate for subject  $j$  in cluster  $i$  is  $\lim_{t \rightarrow \infty} S_{\text{pop}}(t|\mathbf{Z}_{ij}, \mathbf{X}_{ij}, W_i) = G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i))$ . Thus, we can model a broad class of improper survival functions with a variety of cure fractions based on different values of  $\alpha$ . In (4), we need the constraint

$$0 \leq \alpha\theta(\mathbf{Z}_{ij}|\alpha, W_i)F(t|\mathbf{X}_{ij}, W_i) \leq 1$$

to be satisfied for all  $i, j$  and  $t$ , which can be further simplified to

$$0 \leq \alpha\theta(\mathbf{Z}_{ij}|\alpha, W_i) \leq 1,$$

since  $0 \leq F(t|\mathbf{X}_{ij}, W_i) \leq 1$ . Constrained parameter problems often make computation and analysis much more complicated. It is noteworthy that the improper population survival function in (4) is defined at the cluster-specific level.

To accommodate various model structures, we propose a general form of the covariates,

$$\theta(\mathbf{Z}_{ij}|\alpha, W_i) = \frac{W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}{1 + \alpha W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}. \quad (5)$$

When  $\alpha = 0$ , (5) has the exponential form as in the proportional hazards cure model, i.e.,  $\theta(\mathbf{Z}_{ij}|0, W_i) = W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$ ; and when  $\alpha = 1$ , it has the logistic structure of the mixture cure model, i.e.,  $\theta(\mathbf{Z}_{ij}|1, W_i) = W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) / \{1 + W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\}$ . More importantly, the parameter constraints are automatically satisfied with (5), and thus it reduces to an unconstrained parameter problem since

$$0 \leq \frac{\alpha W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}{1 + \alpha W_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})} \leq 1, \quad i = 1, \dots, n; \quad j = 1, \dots, n_i.$$

This covariate structure completely eliminates the burden due to the parameter constraints, and nicely facilitates the estimation procedure.

Based on the Cox-type shared frailty model, we incorporate the covariates  $\mathbf{X}_{ij}$  through

$$F(t|\mathbf{X}_{ij}, W_i) = 1 - S(t)^{W_i \exp(\boldsymbol{\gamma}^T \mathbf{X}_{ij})}, \quad (6)$$

where  $S(t) = 1 - F(t)$  is the baseline survival function. Let  $\boldsymbol{\phi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta, F)$  and  $\mathbf{W} = (W_1, \dots, W_n)^T$ . Then the conditional likelihood function given  $\mathbf{W}$  is

$$\begin{aligned} & \prod_{i=1}^n \prod_{j=1}^{n_i} \left[ \left\{ G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \theta(\mathbf{Z}_{ij}|\alpha, W_i) f(Y_{ij}|\mathbf{X}_{ij}, W_i) \right\}^{\Delta_{ij}} \right. \\ & \quad \times \left. \left\{ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right\}^{1-\Delta_{ij}} \right]^{I(Y_{ij} < \infty)} \left[ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)) \right]^{I(Y_{ij} = \infty)}, \end{aligned}$$

where  $f(t|\mathbf{X}_{ij}, W_i)$  is the first derivative of  $F(t|\mathbf{X}_{ij}, W_i)$  with respect to  $t$ , and  $G'_\alpha(x)$  is that of  $G_\alpha(x)$ . We can obtain the observed-data likelihood for  $\boldsymbol{\phi}$  by integrating out the frailty in the complete-data likelihood,

$$\begin{aligned} & \prod_{i=1}^n \int_0^\infty \left( \prod_{j=1}^{n_i} \left[ \left\{ G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \theta(\mathbf{Z}_{ij}|\alpha, W_i) f(Y_{ij}|\mathbf{X}_{ij}, W_i) \right\}^{\Delta_{ij}} \right. \right. \\ & \quad \times \left. \left. \left\{ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right\}^{1-\Delta_{ij}} \right]^{I(Y_{ij} < \infty)} \right. \\ & \quad \times \left. \left. \left[ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)) \right]^{I(Y_{ij} = \infty)} \right) \psi(W_i|\eta) dW_i \end{aligned}$$

where  $\psi(W_i|\eta)$  is a gamma density function with mean 1 and variance  $\eta$ .

In order to estimate the unknown parameters, we need to maximize the observed-data likelihood. However, this maximum does not exist because one can always choose  $f(Y_{ij}) = \infty$  for some  $Y_{ij}$  with  $\Delta_{ij} = 1$ , where  $f(t) = F'(t)$  is the baseline density function. Thus, we take a nonparametric maximum likelihood approach, in which  $F$  is allowed to be a right-continuous function. Specifically, we replace  $f(Y_{ij})$  with  $F\{Y_{ij}\}$ , the jump

size of  $F(Y_{ij})$  at  $Y_{ij}$ , and the jump size of  $F(t|\mathbf{X}_{ij}, W_i)$  at  $Y_{ij}$  is  $F\{Y_{ij}|\mathbf{X}_{ij}, W_i\}$ . The resultant nonparametric likelihood function is denoted by  $L_n(\phi)$ .

We maximize the logarithm of  $L_n(\phi)$ , denoted by  $l_n(\phi)$ , to obtain the nonparametric maximum likelihood estimators (NPMLE) of  $\phi$ ,  $\hat{\phi}_n \equiv (\hat{\beta}_n, \hat{\gamma}_n, \hat{\eta}_n, \hat{F}_n)$ . It is easy to show that  $\hat{F}_n$  must be a step function with jumps only at the observed failure time points, and that  $\hat{\phi}_n$  depends on the  $Y_{ij}$ 's only through their ranks. Let  $Y_{(1)} < \dots < Y_{(m)}$  denote the ordered distinct observed failure times. We can show that  $\hat{F}_n$  must satisfy the constraint

$$\sum_{k=1}^m \hat{F}_n\{Y_{(k)}\} = 1,$$

where  $\hat{F}_n\{Y_{(k)}\}$  denotes the jump size of  $\hat{F}_n$  at  $Y_{(k)}$ . This constraint presents both numerical and theoretical challenges in terms of maximizing the nonparametric likelihood and deriving the asymptotic properties of the NPMLE. We reparameterize the unknown parameters and consider the baseline cumulative hazard function  $\Lambda(t) = -\log(1 - F(t))$ . Therefore, we estimate the jump sizes of  $\Lambda(t)$ , denoted by  $\Lambda\{t\}$ , at  $t = Y_{(k)}, k = 1, \dots, m-1$ . The jump size of  $\Lambda(t)$  at  $t = Y_{(m)}$  is fixed at  $\infty$ . Let  $\hat{\Lambda}_n$  denote the NPMLE of  $\Lambda$ , then we can estimate  $F$  by  $\hat{F}_n(t) = 1 - \exp(-\hat{\Lambda}_n(t))$  for  $t < Y_{(m)}$  and  $\hat{F}_n(t) = 1$  for  $t \geq Y_{(m)}$ .

Following the argument of Murphy and van der Vaart (2000), we can estimate the covariance matrix of  $(\hat{\beta}_n, \hat{\gamma}_n, \hat{\eta}_n)$  by using the profile likelihood function for  $(\beta, \gamma, \eta)$ , which is defined as the maximum likelihood of  $L_n(\phi)$  for any fixed  $(\beta, \gamma, \eta)$ . Alternatively, we may estimate the asymptotic covariance by simply inverting the observed information matrix for all the parameters including  $\beta, \gamma, \eta$ , and the jump sizes of  $\hat{F}_n$  or  $\hat{\Lambda}_n$ . With this approach, we can estimate the asymptotic variance for  $\hat{F}_n$  as well. Our simulation studies indicated that both approaches work very well in practical situations.

### 3 Asymptotic Properties

We first impose the following assumptions:

- (C1) Covariates  $\mathbf{X}_{ij}$  and  $\mathbf{Z}_{ij}$  are bounded with probability one. Furthermore, if there exist constant vectors  $\beta$  and  $\gamma$  such that

$$\beta^T \mathbf{Z}_{ij} = 0 \quad \text{and} \quad \gamma^T \mathbf{X}_{ij} = 0,$$

almost surely, then  $\boldsymbol{\beta} = \mathbf{0}$  and  $\boldsymbol{\gamma} = \mathbf{0}$ .

(C2) Conditional on  $\mathbf{X}_{ij}$  and  $\mathbf{Z}_{ij}$ , the right-censoring time  $C_{ij}$  is independent of  $T_{ij}$  and random effect  $W_i$ , and  $P(C_{ij} = \infty \text{ and } T_{ij} = \infty | \mathbf{X}_{ij}, \mathbf{Z}_{ij}) > 0$ .

(C3) The true values of  $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta)$ , denoted by  $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \eta_0)$ , belong to the interior of a known compact set,

$$\mathcal{B}_0 = \{(\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta) : |\boldsymbol{\beta}| \leq B \text{ and } |\boldsymbol{\gamma}| \leq B \text{ for some constant } B, \eta \text{ is bounded away from } 0 \text{ and } \infty\}.$$

(C4) The true baseline cumulative distribution function for the uncured population  $F_0$  is differentiable with  $f_0(t) \equiv F_0'(t) > 0$  for all  $t \in R^+$ .

(C5) The cluster size is completely random. In addition, there exists a positive integer  $n_0$  such that  $1 \leq n_i \leq n_0$  and  $P(n_i \geq 2) > 0$ .

Condition (C1) is equivalent to the linear independence of  $\mathbf{X}_{ij}$  and the linear independence of  $\mathbf{Z}_{ij}$ . The condition  $P(C_{ij} = \infty \text{ and } T_{ij} = \infty | \mathbf{X}_{ij}, \mathbf{Z}_{ij}) > 0$  in (C2) ensures that at least some subjects are cured and are not right-censored. If (C2) is not true, one would need to impose one of the following three assumptions as noted by Li et al. (2001): (1) regression parameters excluding intercept in  $\boldsymbol{\beta}$  cannot all be zero; (2) there exists a constant  $\tau$  such that  $P(C_{ij} = \tau | \mathbf{X}_{ij}, \mathbf{Z}_{ij}) = P(C_{ij} > \tau | \mathbf{X}_{ij}, \mathbf{Z}_{ij}) > 0$  and  $F(\tau) = 1$ ; (3)  $F(t)$  has a parametric form. The first assumption implies that the model is not identifiable if there are no covariate effects. The second assumption is essentially the same as (C2) and  $\tau$  can be treated as  $\infty$ . The third assumption does not apply in our case since  $F(t)$  is unspecified in the proposed semiparametric model. Condition (C5) implies that the cluster size is bounded and some clusters have at least two subjects. Conditions (C1), (C2) and (C5) ensure the identifiability of the unknown parameters  $\boldsymbol{\phi}$ .

We first show that the maximizers of the nonparametric likelihood function  $L_n(\boldsymbol{\phi})$  exist. For any  $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta, F)$  in the parameter space,  $L_n(\boldsymbol{\phi})$  is bounded by

$$\prod_{i=1}^n \max_{\mathbf{x}_{ij}, \mathbf{z}_{ij}, (\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta) \in \mathcal{B}_0} \int_0^\infty \left\{ \prod_{j=1}^{n_i} (1 + \alpha W_i \exp(\boldsymbol{\beta}^T \mathbf{z}_{ij}))^{\Delta_{ij}} \right\} \psi(W_i | \eta) dW_i < \infty,$$

where the inequality follows from the boundedness of  $\mathbf{X}_{ij}$  and the compactness of  $\mathcal{B}_0$ .

The following statements provide the asymptotic properties of the proposed estimators.

**Theorem 1.** Under conditions (C1)-(C5),  $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| \rightarrow 0$ ,  $\|\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\| \rightarrow 0$ ,  $|\widehat{\eta}_n - \eta_0| \rightarrow 0$ , and  $\sup_{t \in R^+} |\widehat{F}_n(t) - F_0(t)| \rightarrow 0$  almost surely, where  $\|\cdot\|$  is the Euclidean norm.

**Theorem 2.** Under conditions (C1)-(C5),  $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0, \widehat{\eta}_n - \eta_0, \widehat{F}_n - F_0)$  converges weakly to a zero-mean Gaussian process in the metric space  $l^\infty(\mathcal{H})$ , where

$$\mathcal{H} = \{(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) : \mathbf{h}_1 \in R^{d_1}, \mathbf{h}_2 \in R^{d_2}, h_3 \in R,$$

$$h_4(\cdot) \text{ is a function on } [0, \infty); \|\mathbf{h}_1\| \leq 1, \|\mathbf{h}_2\| \leq 1, |h_3| \leq 1, \|h_4\|_V \leq 1\}$$

and  $\|h_4\|_V$  denotes the total variation of  $h_4(\cdot)$  in  $[0, \infty)$ . Furthermore,  $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, \widehat{\eta}_n)$  are asymptotically efficient.

Theorem 1 establishes the consistency of the NPMLEs. The basic idea to prove Theorem 1 is as follows. Suppose that  $\widehat{\boldsymbol{\phi}}_n$  converges to  $\boldsymbol{\phi}^*$ . We construct a distribution function  $\widetilde{F}_n$  converging to  $F_0$ . Then, because  $l_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, \widehat{\eta}_n, \widehat{F}_n)/n - l_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \eta_0, \widetilde{F}_n)/n \geq 0$ , this difference diverges to the negative Kullback-Leibler divergence between  $\boldsymbol{\phi}^*$  and  $\boldsymbol{\phi}_0$ . Next we prove that  $\boldsymbol{\phi}$  is identifiable and then the identifiability result gives  $\boldsymbol{\phi}^* = \boldsymbol{\phi}_0$ . Once the consistency of the NPMLEs is established, the asymptotic distribution of the NPMLEs stated in Theorem 2 can be derived by verifying the four conditions in Theorem 3.3.1 of van der Vaart and Wellner (1996). The key steps in the proof are verifying some Donsker classes and proving the invertibility of the information operator. The proofs of Theorems 1 and 2 are given in the Appendix.

The inference is based on the selected model with the transformation parameter fixed. To select the best model or estimate the transformation parameter  $\alpha$ , we may minimize the Akaike information criterion (AIC), which is defined as twice the negative log-likelihood function plus twice the number of parameters.

## 4 Simulation Studies

We conducted extensive simulation studies to examine the finite-sample performance of our proposed methodology. In the first set of simulation studies, we investigated the properties of the proposed NPMLEs and their variance estimates. We generated data from model (4), where  $Z_{ij1} = 1$ ,  $Z_{ij2}$  was a Bernoulli random variable with a success probability of 0.5,  $X_{ij1}$  was a standard normal random variable, and  $X_{ij2} = Z_{ij2}$ . The true parameter values of  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$  and  $\gamma_2$  were set to be 0.5,  $-0.5$ , 0.5 and 0.5, respectively. The gamma random

---

variable  $W_i$  had mean 1 and variance  $\eta = 1$ . The baseline survival function was from an exponential distribution with mean 1, i.e.,  $S(t) = \exp(-t)$ . We considered four different models by varying the values of  $\alpha$  from 0 to 1. The median lifetime for uncured subjects ranged from 0.24 to 0.54 and the average cure rate ranged from 0.37 to 0.50 as  $\alpha$  changed from 0 to 1. The censoring time was generated from a uniform distribution on  $(0, 20)$  and then censoring times greater than 15 were treated as  $\infty$ . Subjects who were not cured had an approximately 10% chance of being right-censored. We considered different combinations of sample sizes and cluster sizes. For each simulation set-up, we generated 1,000 data sets. As there was no closed form for the observed-data likelihood, we used the Gauss-Laguerre quadrature to numerically approximate the likelihood function. Our experience indicated that a Gauss-Laguerre quadrature with 40 abscissae would provide a very accurate approximation. We maximized the observed-data likelihood by directly using the quasi-Newton algorithm in Press et al. (1992).

In Tables 1 and 2, we summarize the estimation results for each value of  $\alpha$ : “SE” is the sampling standard error of the parameter estimator; “SEE” is the average of the estimated standard errors based on the asymptotic normal approximation; and “CP(%)” is the coverage probability of the 95% confidence interval. The biases of the proposed estimators appear to be negligible. The estimate of the standard error reflects accurately the true variation, and for moderate sample sizes and cluster sizes the confidence intervals have proper coverage probabilities. As the sample size becomes larger (either by increasing the number of clusters or the cluster size), the variances of the parameter estimates decrease, and the coverage probabilities improve. The asymptotic variances of the NPMLEs were obtained by inverting the observed information matrix. We also applied the profile likelihood approach, and obtained similar results for  $(\beta, \gamma, \eta)$ . For the nonparametric estimation of  $F(t)$ , we evaluated its estimates at  $t = 0.5$  and  $t = 1.0$ . The biases of  $\hat{F}_n(t)$  are very small, and the standard error estimates are quite close to the sampling standard errors. The coverage probability of the 95% confidence interval is slightly low for small sample sizes, however it improves substantially as the sample size increases.

The simulation results in Tables 1 and 2 were obtained by fixing the transformation parameter  $\alpha$  at the true value. In practice, we may choose the transformation that minimizes the AIC. In our situation, the

**Table 1** NPMLs under the proposed transformation cure frailty model with  $\alpha = 0$  and  $\alpha = 1$ 

$n$	$n_i$	Parameter	$\alpha = 0$				$\alpha = 1$			
			Bias	SE	SEE	CP(%)	Bias	SE	SEE	CP(%)
50	3	$\beta_1$	0.001	0.229	0.222	94.1	-0.013	0.312	0.304	92.7
		$\beta_2$	-0.009	0.172	0.168	95.1	-0.013	0.24	0.233	94.5
		$\gamma_1$	0.006	0.198	0.206	96.3	0.010	0.190	0.181	94.9
		$\gamma_2$	-0.005	0.333	0.330	94.4	0.003	0.363	0.335	93.5
		$\eta$	0.022	0.335	0.305	91.7	0.004	0.444	0.412	90.1
		$F(0.5)$	-0.006	0.098	0.096	93.4	-0.003	0.095	0.089	92.1
		$F(1.0)$	-0.002	0.121	0.114	89.4	0.003	0.111	0.104	91.0
50	4	$\beta_1$	-0.001	0.216	0.198	92.0	-0.023	0.280	0.262	92.2
		$\beta_2$	-0.006	0.143	0.140	94.9	-0.008	0.203	0.199	95.4
		$\gamma_1$	0.009	0.174	0.171	95.0	0.013	0.157	0.150	94.7
		$\gamma_2$	-0.004	0.277	0.272	93.4	-0.008	0.294	0.279	93.9
		$\eta$	0.028	0.305	0.270	90.9	-0.003	0.354	0.339	90.6
		$F(0.5)$	-0.004	0.083	0.084	94.7	-0.002	0.080	0.079	93.5
		$F(1.0)$	0.000	0.104	0.101	92.5	0.001	0.096	0.092	92.7
100	3	$\beta_1$	-0.006	0.166	0.156	93.7	-0.018	0.219	0.214	93.8
		$\beta_2$	-0.007	0.121	0.117	95.1	-0.017	0.161	0.163	94.9
		$\gamma_1$	0.014	0.150	0.143	93.4	0.006	0.130	0.125	94.4
		$\gamma_2$	-0.003	0.224	0.230	96.7	-0.012	0.242	0.233	93.8
		$\eta$	-0.007	0.222	0.211	93.4	-0.014	0.295	0.287	91.8
		$F(0.5)$	-0.006	0.069	0.068	93.5	-0.004	0.065	0.063	93.8
		$F(1.0)$	-0.004	0.084	0.083	92.6	-0.004	0.079	0.074	93.0
100	4	$\beta_1$	-0.010	0.147	0.141	94.1	-0.022	0.186	0.187	94.2
		$\beta_2$	-0.013	0.100	0.099	94.5	-0.020	0.143	0.140	94.7
		$\gamma_1$	0.013	0.124	0.120	94.3	0.011	0.105	0.105	94.4
		$\gamma_2$	0.007	0.194	0.191	95.0	-0.016	0.197	0.195	94.7
		$\eta$	0.017	0.206	0.191	94.1	0.006	0.246	0.241	93.1
		$F(0.5)$	-0.002	0.061	0.060	94.8	-0.003	0.056	0.056	93.8
		$F(1.0)$	-0.001	0.074	0.073	94.7	-0.002	0.067	0.066	93.2

numbers of unknown parameters are the same for different transformation models. Therefore, minimizing the AIC is equivalent to maximizing the log-likelihood function. Figure 2 depicts the average of the profile log-likelihood functions over 1,000 replicates of 100 clusters with size four for different transformations. As expected, the transformation that maximizes the average of the profile log-likelihood functions is close to the true transformation.

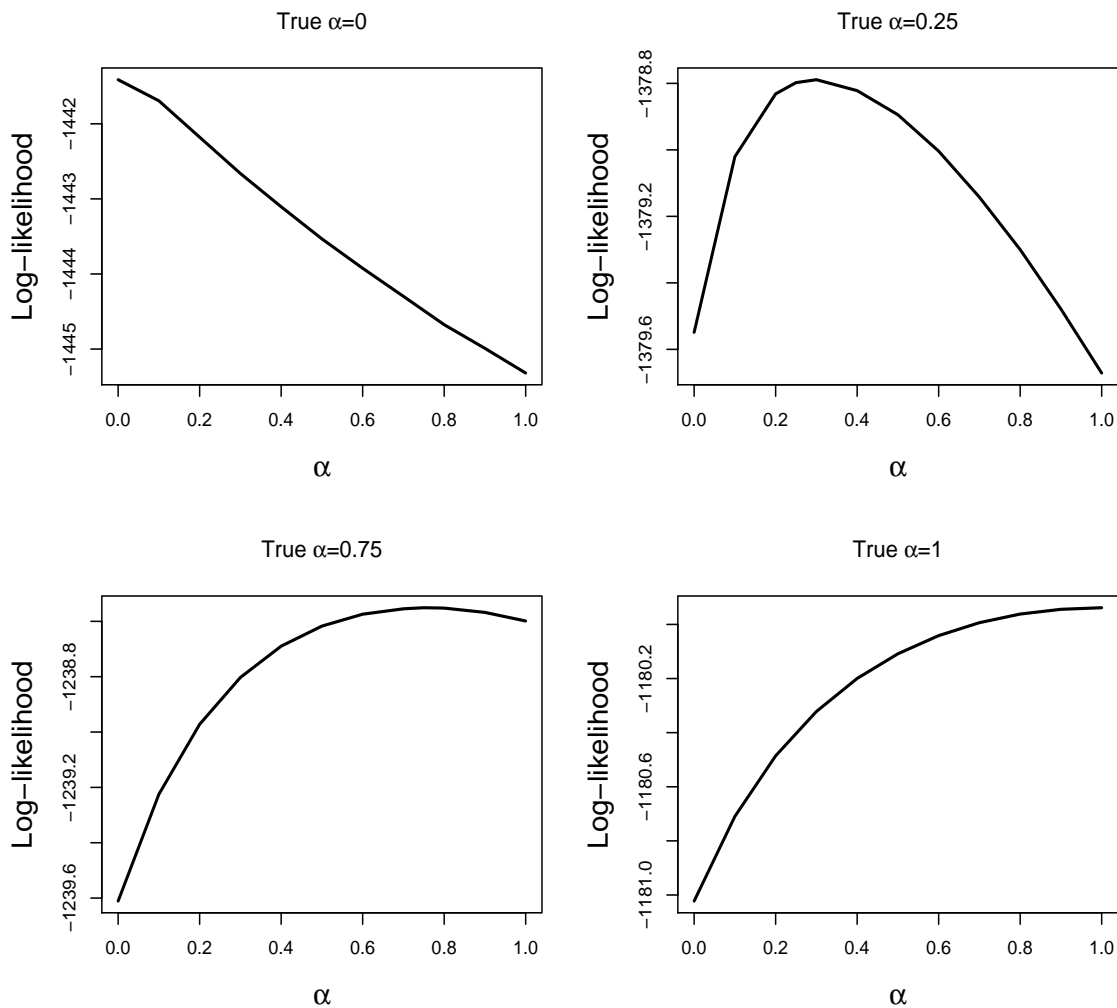
In a hypothesis testing framework, we evaluated the effect on the inference of the regression parameter if the within-cluster correlation is ignored. In particular, we were interested in testing the hypotheses  $H_0 : \beta_2 = 0$  and  $H_0 : \gamma_2 = 0$ , respectively. We considered the proportional hazards cure model, and set other parameter values to be the same as in the previous simulation studies. For making inference on  $\beta_2$ , we set  $\gamma_2 = 0.6$  and varied  $\beta_2$  from 0 to 0.3; for making inference on  $\gamma_2$ , we set  $\beta_2 = 0.3$  and varied  $\gamma_2$  from 0 to 0.6. For each set-up, we generated 1,000 data sets, each with 100 independent clusters of size four. Figure 3

**Table 2** NPMLEs under the proposed transformation cure frailty model with  $\alpha = 0.25$  and  $\alpha = 0.75$ 

$n$	$n_i$	Parameter	$\alpha = 0.25$				$\alpha = 0.75$			
			Bias	SE	SEE	CP(%)	Bias	SE	SEE	CP(%)
50	3	$\beta_1$	0.003	0.264	0.252	93.8	-0.019	0.233	0.219	92.9
		$\beta_2$	-0.008	0.200	0.190	94.5	-0.007	0.154	0.159	95.7
		$\gamma_1$	0.009	0.187	0.178	94.6	0.001	0.152	0.149	94.7
		$\gamma_2$	0.001	0.334	0.327	94.6	-0.008	0.278	0.272	95.4
		$\eta$	0.026	0.365	0.346	91.6	0.021	0.322	0.293	91.4
		$F(0.5)$	-0.008	0.094	0.090	92.5	-0.002	0.083	0.079	93.4
		$F(1.0)$	-0.001	0.111	0.108	92.2	0.002	0.100	0.095	92.1
50	4	$\beta_1$	0.002	0.306	0.286	92.1	-0.028	0.259	0.248	93.1
		$\beta_2$	-0.015	0.218	0.219	95.5	-0.026	0.195	0.185	93.9
		$\gamma_1$	-0.005	0.188	0.177	94.1	0.014	0.158	0.147	93.8
		$\gamma_2$	0.006	0.351	0.331	93.3	0.002	0.287	0.275	94.4
		$\eta$	0.003	0.412	0.386	91.6	0.007	0.353	0.325	90.3
		$F(0.5)$	-0.004	0.091	0.088	92.6	0.002	0.080	0.078	92.2
		$F(1.0)$	-0.003	0.107	0.104	92.0	0.006	0.096	0.092	91.5
100	3	$\beta_1$	0.002	0.178	0.175	94.9	-0.008	0.162	0.155	94.5
		$\beta_2$	-0.012	0.136	0.132	94.5	-0.009	0.118	0.112	93.4
		$\gamma_1$	0.011	0.129	0.123	94.7	0.013	0.105	0.103	95.6
		$\gamma_2$	0.004	0.231	0.227	95.2	-0.002	0.190	0.189	94.2
		$\eta$	0.006	0.247	0.237	92.9	-0.002	0.206	0.204	94.2
		$F(0.5)$	-0.004	0.067	0.063	92.5	-0.004	0.056	0.056	94.9
		$F(1.0)$	-0.003	0.082	0.077	92.9	-0.004	0.068	0.068	94.1
100	4	$\beta_1$	-0.002	0.214	0.202	93.7	-0.017	0.18	0.177	94.3
		$\beta_2$	-0.010	0.159	0.154	93.9	-0.022	0.131	0.131	95.5
		$\gamma_1$	0.008	0.124	0.123	94.3	0.007	0.103	0.103	94.8
		$\gamma_2$	-0.011	0.240	0.230	93.5	-0.005	0.199	0.192	94.3
		$\eta$	0.000	0.291	0.270	91.9	0.016	0.247	0.232	92.7
		$F(0.5)$	-0.004	0.063	0.063	93.5	-0.002	0.055	0.056	95.1
		$F(1.0)$	-0.005	0.076	0.075	93.6	-0.002	0.067	0.067	93.4

depicts the type I error rates and powers of the Wald tests of  $\beta_2$  and  $\gamma_2$  at the nominal significance levels of 0.05 and 0.01, respectively. The proposed test has an accurate control of the type I error rate and has substantially more power than the naive method ignoring the within-cluster correlations.

Finally, we conducted simulation studies to evaluate the performance of the proposed approach when the frailty distribution was misspecified. Specifically, we used the same simulation setting as above except that the frailty was generated from a log-normal distribution with mean 1 and variance 1. Simulation results are summarized in Table 3. Although the variance parameter  $\eta$  is underestimated, the estimation and inference of regression parameters are relatively robust to the misspecified frailty distribution. For regression parameters, the proposed estimators appear to be unbiased; the estimated standard errors agree well with the sampling standard errors; and the coverage probabilities are accurate. We observed similar patterns as in Figure 3



**Fig. 2** Average of the profile log-likelihood functions over 1,000 replicates with  $n = 100$  and  $n_i = 4$ , for different transformations.

**Table 3** Sensitivity analysis under a misspecified frailty distribution

Parameter	Bias	SE	SEE	CP(%)
$\beta_1$	0.026	0.130	0.120	92.2
$\beta_2$	0.001	0.090	0.085	94.9
$\gamma_1$	0.001	0.107	0.106	95.3
$\gamma_2$	0.009	0.180	0.181	95.1
$\eta$	0.378	0.129	0.134	24.9

for the test size/power comparison. The proposed test still controls the type I error rate accurately and is substantially more powerful than its counterpart ignoring correlations.

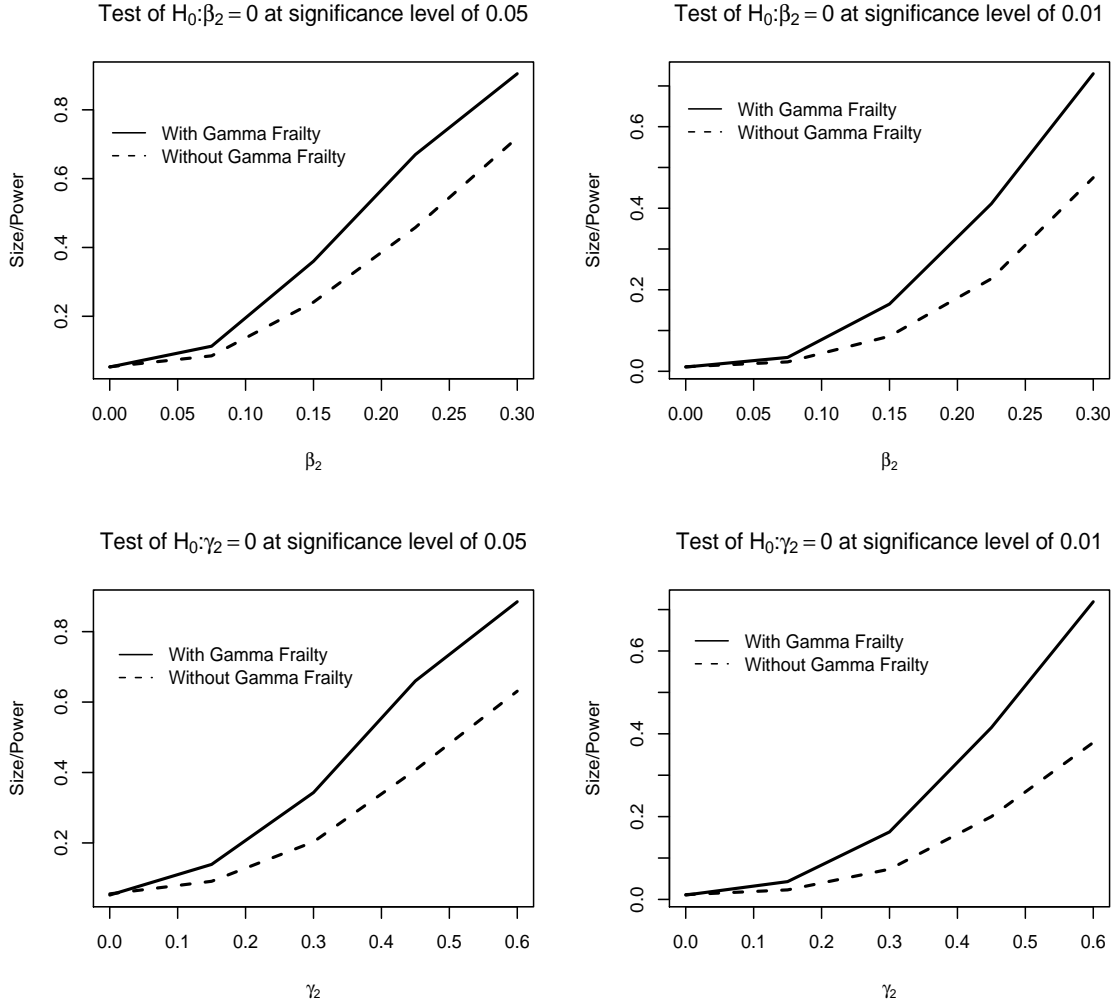
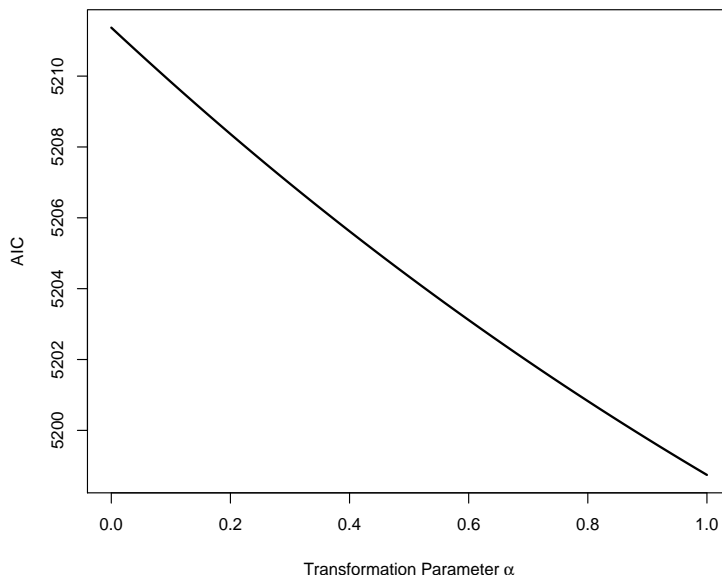


Fig. 3 Type I error rate and power of the Wald test of  $H_0 : \beta_2 = 0$  and  $H_0 : \gamma_2 = 0$ , respectively.

### 5 COGA Data

As an illustration, we applied the proposed methods to the COGA data. We considered a subject as “cured” or insusceptible to alcoholism if the subject was censored at 62 years of age or beyond. In the data set, 107 individuals had a censoring time longer than or equal to 62 years. Using a semiparametric transformation model with random effects for right-censored data, Diao and Lin (2006) found a significant association between genotype at SNP rs1972373 on chromosome 14 and the age at onset of alcoholism. In our analysis, we fit the proposed model and included the subject’s sex, the between-family genotype score and the within-family



**Fig. 4** Akaike information criterion versus transformation parameter  $\alpha$  for the COGA data.

genotype score at SNP rs1972373 as covariates, i.e.,  $\mathbf{Z}_{ij} = (1, \mathbf{X}_{ij}^T)^T$ , and  $\mathbf{X}_{ij} = (\text{Sex}_{ij}, \text{Bscore}_{ij}, \text{Wscore}_{ij})^T$ . The sex of the subject was coded as 1 for male and 0 for female, and the genotype score was coded as 0, 1, and 2 for 1/1, 1/2, and 2/2, respectively. We used the between-family genotype score to account for spurious association introduced by population admixture (Abecasis et al., 2000). We compared different models within the proposed general transformation class, including the proportional hazards cure model and the mixture cure model. According to the Akaike information criterion (see Figure 4), we selected the mixture cure model, the results of which are reported in Table 4. We can see that the sex of the subject significantly affected both the cure fraction and survival probability for the uncured subjects. Particularly, female subjects were more likely to be “cured,” that is not to have alcoholism. We also found significant associations between the SNP genotype and the cure fraction and the survival probability for uncured individuals. The subpopulation with genotype 2/2 appeared to be at a higher risk of developing alcoholism and to have a lower cure fraction. These results agree very well with the empirical observations in the Kaplan-Meier curves.

**Table 4** NPMLEs under the mixture cure frailty model for the COGA data

Parameter	Estimate	SE	Est/SE	<i>p</i> -value
$\beta_{\text{Intercept}}$	-0.719	0.092	-7.815	< 0.001
$\beta_{\text{Sex}}$	1.906	0.101	18.871	< 0.001
$\beta_{\text{Bscore}}$	0.266	0.096	2.771	0.006
$\beta_{\text{Wscore}}$	0.574	0.111	5.171	< 0.001
$\gamma_{\text{Sex}}$	0.271	0.090	3.011	0.026
$\gamma_{\text{Bscore}}$	0.150	0.079	-1.899	0.058
$\gamma_{\text{Wscore}}$	0.185	0.092	2.011	0.044
$\eta$	0.100	0.036	2.778	0.003

## 6 Discussion

We have considered a class of cure rate frailty models by imposing the Box-Cox transformation on the population survival function. Even though a nonlinear parameter constraint arises from this model formulation, our general covariate form removes the constraint completely. This class of transformation models makes the cure rate modeling scheme much more flexible and general than other methods. It nicely links the two main formulations of cure rate models, i.e., the mixture and the proportional hazards cure models. This family of cure rate frailty models has great potential in modeling multivariate survival data with a cure fraction. In contrast to the marginal models by Peng et al. (2007) and Yu and Peng (2008) that yield population-average estimates, our frailty model takes a subject-specific approach.

In this paper, we have considered the shared frailty model, i.e., all subjects from the same cluster are assumed to share a common frailty. As one referee pointed out, the shared frailty model is ideal when the cluster sizes are large and subjects within the same cluster are exchangeable. As family members are usually not exchangeable in the COGA data, it would be more desirable to use correlated frailty models. For example, Locatelli et al. (2007) studied genetic and environmental factors in susceptibility to breast cancer in a correlated frailty-mixture model. More specifically, we may accommodate family structures through

$$\lambda(t|\mathbf{X}_{ij}, R_{ij}) = \lambda(t) \exp(\boldsymbol{\gamma}^T \mathbf{X}_{ij} + R_{ij}),$$

and

$$\theta(\mathbf{Z}_{ij}|\alpha, R_{ij}) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij} + R_{ij})}{1 + \alpha \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij} + R_{ij})}$$

where  $\mathbf{R}_i \equiv (R_{i1}, \dots, R_{in_i})^T$  is a vector of normal random variables with mean values of 0 and variance-covariance matrix  $\Sigma_i$ . To take into consideration of the family structure, we can let  $\Sigma_i = \eta \mathbf{V}_i$ , where  $\mathbf{V}_i$  is the kinship coefficient matrix for the  $i$ th family. The kinship coefficient between two individuals is the probability that two genes sampled at random from each individual are identical and is often used in genetic literature to measure the relatedness between individuals. Such modeling is commonly used in the association analysis of quantitative traits in genetic family studies and can accommodate arbitrary family structures (Abecasis et al., 2000). For not too large family sizes, we can use the Gaussian-Hermite quadrature to approximate the marginal likelihood function. However, the Gaussian quadrature approximation may be computationally intensive with large families as the computation burden increases exponentially as the family size increases. Alternatively, one may consider less accurate approaches such as the Laplace approximation or the Monte Carlo method. Future research along this direction is warranted.

We have assumed gamma frailty in the proposed cure rate models, as gamma frailty is commonly used in the analysis of clustered time-to-event data. Unlike in the Cox model with gamma frailty, there is no closed form for the marginal likelihood under the cure rate gamma frailty model because of the involvement of the cure fraction. However, the gamma frailty model can be easily generalized to different frailty distributions, including the log-normal distribution or positive stable distribution. It would be worthwhile to consider a variety of frailty distributions and examine them empirically (Glidden, 2007).

We have developed free-downloading computer software for the new methods, which computes the NPMLE very fast, and can also handle large clusters. For the COGA data with 1,371 observations and the largest family size of 32, it took less than one minute to obtain the NPMLE on a Dell PowerEdge 2900 server. With cluster sizes ranging from 3 to 4, and  $n = 100$ , in the simulation studies, the computation took less than one second.

The proposed class of models is fundamentally different from that of Zeng et al. (2006). The latter does not include the mixture cure model, which, however, is one of the most commonly used cure models in the literature. In addition, their method requires independent observations and thus is not appropriate for the COGA family genetic data. We used reparameterization of the constrained parameter  $F(t)$ , whereas

Zeng et al. (2006) applied the Lagrange multiplier method. Our approach greatly facilitates the theoretical development of the asymptotic properties of the NPMLE.

Our numerical experience indicates that with a small sample size, the likelihood function under the proposed model tends to be quite flat for the transformation parameter  $\alpha$ . Moreover, the computation may not be stable if we simultaneously estimate all the unknown parameters including  $\alpha$ . Instead, we suggest to use the grid search technique to choose the transformation parameter according to the AIC. The proposed inference procedure is based on model selection after fitting each model with the transformation parameter fixed. It would be desirable to account for the additional variation that is introduced by the data-based selection of  $\alpha$ . Future investigation is warranted along this direction.

## 7 Appendix

### 7.1 Proof of Theorem 1

We first prove that under conditions (C1)-(C5), the parameters  $\phi = (\beta, \gamma, \eta, F)$  are identifiable. Suppose that two sets of parameters  $\phi$  and  $\phi^*$  give the same likelihood function for the observed data, then we claim that  $\phi = \phi^*$ . Let  $\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)$ ,  $F^*(Y_{ij}|\mathbf{X}_{ij}, W_i)$ , and  $f^*(Y_{ij}|\mathbf{X}_{ij}, W_i)$  have the same expressions as  $\theta(\mathbf{Z}_{ij}|\alpha, W_i)$ ,  $F(Y_{ij}|\mathbf{X}_{ij}, W_i)$ , and  $f(Y_{ij}|\mathbf{X}_{ij}, W_i)$ , respectively, but with  $\phi$  replaced by  $\phi^*$ . Since

$$\begin{aligned}
& \int_0^\infty \left( \prod_{j=1}^{n_i} \left[ \left\{ G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)F(Y_{ij}|\mathbf{X}_{ij}, W_i))\theta(\mathbf{Z}_{ij}|\alpha, W_i)f(Y_{ij}|\mathbf{X}_{ij}, W_i) \right\}^{\Delta_{ij}} \right. \right. \\
& \quad \times \left. \left. \left\{ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right\}^{1-\Delta_{ij}} \right]^{I(Y_{ij}<\infty)} \right. \\
& \quad \times \left. \left. \left[ G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)) \right]^{I(Y_{ij}=\infty)} \right) \psi(W_i|\eta) dW_i \right. \\
& = \int_0^\infty \left( \prod_{j=1}^{n_i} \left[ \left\{ G'_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)F^*(Y_{ij}|\mathbf{X}_{ij}, W_i))\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)f^*(Y_{ij}|\mathbf{X}_{ij}, W_i) \right\}^{\Delta_{ij}} \right. \right. \\
& \quad \times \left. \left. \left\{ G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)F^*(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right\}^{1-\Delta_{ij}} \right]^{I(Y_{ij}<\infty)} \right. \\
& \quad \times \left. \left. \left[ G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)) \right]^{I(Y_{ij}=\infty)} \right) \psi(W_i|\eta^*) dW_i \right. \tag{7}
\end{aligned}$$

for an integer  $k$  such that  $1 \leq k \leq n_i$ , according to condition (C2) we let  $Y_{ij} = \infty$  for  $j = 1, \dots, k$ ; for those  $j > k$ , we perform the following action on the  $j$ th term on both sides of (7). If  $\Delta_{ij} = 0$ , then we replace  $Y_{ij}$  with  $\infty$ ; if  $\Delta_{ij} = 1$ , then we integrate  $Y_{ij}$  from 0 to  $\infty$ . Then we obtain

$$\begin{aligned} & \int_0^\infty \left\{ \prod_{j=1}^k G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)) \right\} \left\{ \prod_{j=k+1}^{n_i} (G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)))^{1-\Delta_{ij}} \right. \\ & \quad \left. \times (1 - G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)))^{\Delta_{ij}} \right\} \psi(W_i|\eta) dW_i \\ &= \int_0^\infty \left\{ \prod_{j=1}^k G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)) \right\} \left\{ \prod_{j=k+1}^{n_i} (G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)))^{1-\Delta_{ij}} \right. \\ & \quad \left. \times (1 - G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)))^{\Delta_{ij}} \right\} \psi(W_i|\eta^*) dW_i. \end{aligned}$$

We sum the above equalities over all possible  $\{\Delta_{ij} : j = k+1, \dots, n_i\}$ . Thus, it holds that

$$\begin{aligned} & \int_0^\infty \left\{ \prod_{j=1}^k G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i)) \right\} \psi(W_i|\eta) dW_i \\ &= \int_0^\infty \left\{ \prod_{j=1}^k G_\alpha(-\theta^*(\mathbf{Z}_{ij}|\alpha, W_i)) \right\} \psi(W_i|\eta^*) dW_i. \end{aligned}$$

When  $\alpha = 0$ , we obtain

$$\left( \frac{1}{1 + \eta e^{\sum_{j=1}^k \beta^T \mathbf{Z}_{ij}}} \right)^{1/\eta} = \left( \frac{1}{1 + \eta^* e^{\sum_{j=1}^k \beta^{*T} \mathbf{Z}_{ij}}} \right)^{1/\eta^*} \quad (8)$$

When  $\alpha > 0$ , we obtain

$$\begin{aligned} & \int_0^\infty \left\{ \prod_{j=1}^k (1 + \alpha W_i e^{\beta^T \mathbf{Z}_{ij}})^{-1/\alpha} \right\} \psi(W_i|\eta) dW_i \\ &= \int_0^\infty \left\{ \prod_{j=1}^k (1 + \alpha W_i e^{\beta^{*T} \mathbf{Z}_{ij}})^{-1/\alpha} \right\} \psi(W_i|\eta^*) dW_i \end{aligned}$$

Note that the left-hand side of (8) is a strictly monotone function of  $\sum_{j=1}^k \beta^T \mathbf{Z}_{ij}$ . Therefore there exists a unique function  $q(\cdot)$  such that  $\sum_{j=1}^k \beta^T \mathbf{Z}_{ij} = q(\sum_{j=1}^k \beta^{*T} \mathbf{Z}_{ij}|\eta, \eta^*)$ . Since (8) holds for arbitrary  $1 \leq k \leq n_i$  and arbitrary  $\mathbf{Z}_{ij}, j = 1, \dots, k$ ,  $q$  must satisfy  $q(x+y) = q(x) + q(y)$  for arbitrary  $x, y \in R$ . Therefore  $q$  takes the form  $q(x) = cx$ , where  $c$  is a constant. It can be show that (8) cannot hold for all possible values of  $\beta^T \mathbf{Z}_{ij}$  if  $c \neq 1$ . It follows that  $\beta^T \mathbf{Z}_{ij} = \beta^{*T} \mathbf{Z}_{ij}$ . By condition (C1), we prove that  $\beta = \beta^*$ . Immediately, we can obtain  $\eta = \eta^*$ . We can prove this result in a similar way when  $\alpha > 0$ .

Next, we let  $\Delta_{ij} = 1, Y_{ij} = 0$  for  $j \leq k$ , and perform the same action as shown previously for  $j > k$  to obtain

$$\begin{aligned} & \int_0^\infty \left\{ \prod_{j=1}^k \theta(\mathbf{Z}_{ij} | \alpha, W_i) W_i e^{\gamma^T \mathbf{X}_{ij}} f(0) \right\} \psi(W_i | \eta) dW_i \\ &= \int_0^\infty \left\{ \prod_{j=1}^k \theta^*(\mathbf{Z}_{ij} | \alpha, W_i) W_i e^{\gamma^{*T} \mathbf{X}_{ij}} f^*(0) \right\} \psi(W_i | \eta^*) dW_i. \end{aligned} \quad (9)$$

The index set  $\{1, \dots, k\}$  in (9) can be replaced by any subset of  $\{1, \dots, n_i\}$ . Furthermore, by assumption (C4),  $f(0) > 0$ . Thus, it is easy to derive from (9) that  $\gamma = \gamma^*$  and  $f(0) = f^*(0)$ .

To show that  $F = F^*$ , we let  $\Delta_{ij} = 1$  in (7) and integrate  $Y_{ij}$  from 0 to  $y$ ; we perform the same action as shown previously for  $j > k$  to obtain

$$\begin{aligned} & \int_0^\infty \left\{ \prod_{j=1}^{n_i} (1 - G_\alpha(-\theta(\mathbf{Z}_{ij} | \alpha, W_i) F(y | X_{ij}, W_i))) \right\} \psi(W_i | \eta) dW_i \\ &= \int_0^\infty \left\{ \prod_{j=1}^{n_i} (1 - G_\alpha(-\theta^*(\mathbf{Z}_{ij} | \alpha, W_i) F^*(y | X_{ij}, W_i))) \right\} \psi(W_i | \eta^*) dW_i. \end{aligned}$$

As both sides of the foregoing equation are strictly monotone in  $F(y)$  and  $F^*(y)$ , we have  $F(y) = F^*(y)$  for any  $y$ .

We next prove that  $\widehat{F}_n(t)$  is bounded in  $[0, 1]$  with probability 1. By differentiating  $l_n(\phi)$  with respect to  $\Lambda\{Y_{ij}\}$  for  $Y_{ij} < Y_{(m)}$  and setting the derivative to 0, we see that  $\widehat{\Lambda}_n$  satisfies the following equation:

$$\begin{aligned} \frac{\Delta_{ij}}{\Lambda\{Y_{ij}\}} &= \sum_{k=1}^n \int_0^\infty R_{1k}(\widehat{\beta}_n, \widehat{\gamma}_n, F, W_k) R_{2k}(Y_{ij}, \widehat{\beta}_n, \widehat{\gamma}_n, F, W_k) \psi(W_k | \widehat{\eta}_n) dW_k \\ &\quad \times \left\{ \int_0^\infty R_{1k}(\widehat{\beta}_n, \widehat{\gamma}_n, F, W_k) \psi(W_k | \widehat{\eta}_n) dW_k \right\}^{-1} \end{aligned}$$

where

$$\begin{aligned} & R_{1k}(\beta, \gamma, F, W_k) \\ &= \prod_{l=1}^{n_k} \left[ \left\{ G'_\alpha(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k)) \theta(\mathbf{Z}_{kl} | \alpha, W_k) f(Y_{kl} | \mathbf{X}_{kl}, W_k) \right\}^{\Delta_{kl}} \right. \\ &\quad \times \left. \left\{ G_\alpha(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k)) \right\}^{1-\Delta_{kl}} \right]^{I(Y_{kl} < \infty)} \\ &\quad \times \left[ G_\alpha(-\theta(\mathbf{Z}_{kl} | \alpha, W_k)) \right]^{I(Y_{ij} = \infty)}, \end{aligned}$$

and

$$\begin{aligned}
R_{2k}(t, \beta, \gamma, F, W_k) &= \sum_{l=1}^{n_k} W_k e^{\gamma^T \mathbf{X}_{kl}} I(t \leq Y_{kl} < \infty) \left[ \Delta_{kl} + \theta(\mathbf{Z}_{kl} | \alpha, W_k) e^{-W_k e^{\gamma^T \mathbf{X}_{kl}} \Lambda(Y_{kl})} \right. \\
&\quad \times \left\{ \Delta_{kl} \frac{G''_{\alpha}(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k))}{G'_{\alpha}(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k))} \right. \\
&\quad \left. \left. + (1 - \Delta_{kl}) \frac{G'_{\alpha}(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k))}{G_{\alpha}(-\theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k))} \right\} \right] \\
&= \sum_{l=1}^{n_k} W_k e^{\gamma^T \mathbf{X}_{kl}} I(t \leq Y_{kl} < \infty) \\
&\quad \times \frac{\Delta_{kl}(1 - \alpha \theta(\mathbf{Z}_{kl} | \alpha, W_k)) + \theta(\mathbf{Z}_{kl} | \alpha, W_k)(1 - F(Y_{kl} | \mathbf{X}_{kl}, W_k))}{1 - \alpha \theta(\mathbf{Z}_{kl} | \alpha, W_k) F(Y_{kl} | \mathbf{X}_{kl}, W_k)}.
\end{aligned}$$

It is easy to show that  $\widehat{\Lambda}_n\{Y_{(l)}\} > 0$  for  $l < m$  and  $\alpha \geq 0$ . Therefore,  $\widehat{F}_n(Y_{(l)}) = 1 - \exp(-\widehat{\Lambda}_n(Y_{(l)}))$  is bounded with probability one. Thus, by choosing a subsequence, still indexed by  $\{n\}$ , we assume that  $\widehat{F}_n$  converges pointwise to a cumulative distribution function  $F^*$  in  $[0, \infty)$ . Because  $\widehat{\beta}_n, \widehat{\gamma}_n, \widehat{\eta}_n$  belong to a compact set, by choosing a further subsequence, we can assume that  $\widehat{\beta} \rightarrow \beta^*, \widehat{\gamma} \rightarrow \gamma^*, \widehat{\eta} \rightarrow \eta^*$ .

Next, we show that  $\beta^* = \beta_0, \gamma^* = \gamma_0, \eta^* = \eta_0$ , and  $F^*(t) = F_0(t)$ . To do so, we construct another function  $\widetilde{\Lambda}_n$  which only has jumps at  $Y_{ij}$  such that  $\Delta_{ij} = 1$  and satisfies

$$\begin{aligned}
\frac{\Delta_{ij}}{\widetilde{\Lambda}_n\{Y_{ij}\}} &= \sum_{k=1}^n \int_0^{\infty} R_{1k}(\beta_0, \gamma_0, F_0, W_k) R_{2k}(Y_{ij}, \gamma_0, \beta_0, F_0, W_k) \psi(W_k | \eta_0) dW_k \\
&\quad \times \left\{ \int_0^{\infty} R_{1k}(\beta_0, \gamma_0, F_0, W_k) \psi(W_k | \eta_0) dW_k \right\}^{-1},
\end{aligned}$$

for  $Y_{ij} < Y_{(m)}$ . Then we define  $\widetilde{\Lambda}_n(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} I(Y_{ij} \leq t) \widetilde{\Lambda}_n\{Y_{ij}\}$  and  $\widetilde{F}_n(t) = 1 - \exp(-\widetilde{\Lambda}_n(t))$  for  $t < Y_{(m)}$  and  $\widetilde{F}_n(t) = 1$  for  $t \geq Y_{(m)}$ .

We show that  $\widetilde{F}_n(t)$  converges to  $F_0(t)$  uniformly in  $t \in [0, \infty)$  with probability 1. To this end, we wish to show that the following two classes

$$\begin{aligned}
\mathcal{C}_1 &= \left\{ \int_0^{\infty} R_{1k}(\beta, \gamma, F, W_k) \psi(W_k | \eta) dW_k : \right. \\
&\quad \left. (\beta, \gamma, \eta) \in \mathcal{B}_0, F \text{ is a cumulative distribution function in } [0, \infty) \right\}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{C}_2 &= \left\{ \int_0^{\infty} R_{1k}(\beta, \gamma, F, W_k) R_{2k}(y, \beta, \gamma, F, W_k) \psi(W_k | \eta) dW_k : \right. \\
&\quad \left. (\beta, \gamma, \eta) \in \mathcal{B}_0, F \text{ is a cumulative distribution function in } [0, \infty), y \in [0, \infty) \right\}
\end{aligned}$$

are P-Donsker. For any  $(\beta_1, \gamma_1, \eta_1)$  and  $(\beta_2, \gamma_2, \eta_2) \in \mathcal{B}_0$  and any cumulative distribution functions  $F_1$  and  $F_2$  in  $[0, \infty)$ , we have

$$\begin{aligned} & |R_{1k}(\beta_1, \gamma_1, F_1, W_k) - R_{1k}(\beta_2, \gamma_2, F_2, W_k)| \\ & \leq O(1) \left\{ \|\beta_1 - \beta_2\| + \|\gamma_1 - \gamma_2\| + \sum_{l=1}^{n_k} |F_1(Y_{kl}) - F_2(Y_{kl})| \right\} (B_{01} + W_k^{n_k}) \end{aligned}$$

and

$$|\psi(W_k|\eta_1) - \psi(W_k|\eta_2)| \leq O(1) W_k^{B_{02}-1} e^{-\frac{W_k}{B_{03}}} |\eta_1 - \eta_2|,$$

where  $B_{0p}, p = 1, 2, 3$ , are positive constants. It follows that

$$\begin{aligned} & \left| \int_0^\infty R_{1k}(\beta_1, \gamma_1, F_1, W_k) \psi(W_k|\eta_1) dW_k - \int_0^\infty R_{1k}(\beta_2, \gamma_2, F_2, W_k) \psi(W_k|\eta_2) dW_k \right| \\ & \leq O(1) \left\{ \|\beta_1 - \beta_2\| + \|\gamma_1 - \gamma_2\| + |\eta_1 - \eta_2| + \sum_{l=1}^{n_k} |F_1(Y_{kl}) - F_2(Y_{kl})| \right\}. \end{aligned}$$

Since the following class

$$\mathcal{B}_0 \times \{F : F \text{ is a cumulative distribution function in } [0, \infty)\}$$

is a Donsker class, we conclude that  $\mathcal{C}_1$  is P-Donsker. Note that  $I(y \leq Y_{kl} < \infty)$  is P-Donsker. By the same argument for  $\mathcal{C}_1$ , we can show that  $\mathcal{C}_2$  is also P-Donsker. It is easy to verify that  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are bounded from below and above. Thus,  $\{\log g : g \in \mathcal{C}_1\}$ ,  $\{\log g : g \in \mathcal{C}_2\}$ , and  $\{g_2/g_1 : g_1 \in \mathcal{C}_1, g_2 \in \mathcal{C}_2\}$  are all P-Donsker. Since a P-Donsker class is also a Glivenko-Cantelli class, by the Glivenko-Cantelli theorem (van der Vaart and Wellner, 1996, p.122),  $\tilde{A}_n(t)$  converges almost surely to  $E\{\sum_{j=1}^{n_i} I(Y_{ij} \leq t) \Delta_{ij} / \mu(Y_{ij})\}$ , where

$$\begin{aligned} \mu(y) &= E \left[ \int_0^\infty R_{1k}(\beta_0, \gamma_0, F_0, W_k) R_{2k}(y, \gamma_0, \beta_0, F_0, W_k) \psi(W_k|\eta_0) dW_k \right. \\ & \quad \left. \times \left\{ \int_0^\infty R_{1k}(\beta_0, \gamma_0, F_0, W_k) \psi(W_k|\eta_0) dW_k \right\}^{-1} \right] \\ &= E \left[ \sum_{l=1}^{n_k} W_k e^{\gamma_0^T \mathbf{X}_{kl}} E \left\{ I(y \leq Y_{kl} < \infty) \right. \right. \\ & \quad \left. \left. \times \frac{\Delta_{kl}(1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)) + \theta_0(\mathbf{Z}_{kl}|\alpha, W_k)(1 - F_0(Y_{kl}|\mathbf{X}_{kl}, W_k))}{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(Y_{kl}|\mathbf{X}_{kl}, W_k)} \right| n_k \right\} \right], \end{aligned}$$

and  $\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)$  and  $F_0(Y_{kl}|\mathbf{X}_{kl}, W_k)$  are evaluated at  $\phi_0$ .

Denoting by  $S_C(\cdot|\mathbf{X}_{kl}, \mathbf{Z}_{kl})$  the survival function of  $C_{kl}$  given  $(\mathbf{X}_{kl}, \mathbf{Z}_{kl})$ , we have

$$\begin{aligned}
& E \left\{ I(y \leq Y_{kl} < \infty) \frac{\Delta_{kl}(1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)) + \theta_0(\mathbf{Z}_{kl}|\alpha, W_k)(1 - F_0(Y_{kl}|\mathbf{X}_{kl}, W_k))}{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(Y_{kl}|\mathbf{X}_{kl}, W_k)} \Big| n_k \right\} \\
&= E \left\{ \int_y^\infty \frac{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k) + \theta_0(\mathbf{Z}_{kl}|\alpha, W_k)(1 - F_0(t|\mathbf{X}_{kl}, W_k))}{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(t|\mathbf{X}_{kl}, W_k)} \right. \\
&\quad \times G'_\alpha(-\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(t|\mathbf{X}_{kl}, W_k))\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)f_0(t|\mathbf{X}_{kl}, W_k)S_C(t|\mathbf{X}_{kl}, \mathbf{Z}_{kl})dt \Big| n_k \left. \right\} \\
&\quad - E \left\{ \int_y^\infty \frac{\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)(1 - F_0(t|\mathbf{X}_{kl}, W_k))}{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(t|\mathbf{X}_{kl}, W_k)} \right. \\
&\quad \times G_\alpha(-\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(t|\mathbf{X}_{kl}, W_k))dS_C(t|\mathbf{X}_{kl}, \mathbf{Z}_{kl}) \Big| n_k \left. \right\} \\
&= E \left\{ \frac{\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)(1 - F_0(y|\mathbf{X}_{kl}, W_k))}{1 - \alpha\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(y|\mathbf{X}_{kl}, W_k)} \right. \\
&\quad \times G_\alpha(-\theta_0(\mathbf{Z}_{kl}|\alpha, W_k)F_0(y|\mathbf{X}_{kl}, W_k))S_C(y|\mathbf{X}_{kl}, \mathbf{Z}_{kl}) \Big| n_k \left. \right\},
\end{aligned}$$

where  $f_0(t|\mathbf{X}_{kl}, W_k)$  is the first derivative of  $F_0(t|\mathbf{X}_{kl}, W_k)$  with respect to  $t$  and the second equality follows from integration by part. Thus,

$$\begin{aligned}
& E \left\{ \sum_{j=1}^{n_i} \frac{I(Y_{ij} \leq t)\Delta_{ij}}{\mu(Y_{ij})} \right\} \\
&= E \left[ \sum_{j=1}^{n_i} E \left\{ \int_0^t \frac{S_C(y|\mathbf{X}_{ij}, \mathbf{Z}_{ij})}{\mu(y)} G'_\alpha(-\theta_0(\mathbf{Z}_{ij}|\alpha, W_i)F_0(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right. \right. \\
&\quad \times \theta_0(\mathbf{Z}_{ij}|\alpha, W_i)W_i e^{\gamma_0^T \mathbf{X}_{ij}} (1 - F_0(Y_{ij}|\mathbf{X}_{ij}, W_i))d\Lambda_0(y) \Big| n_i \left. \left. \right\} \right] \\
&= \int_0^t d\Lambda_0(y) = \Lambda_0(t).
\end{aligned}$$

Consequently,  $\tilde{F}_n(t)$  uniformly converges to  $F_0(t)$ ,  $t \in [0, \infty)$ .

As the observed log-likelihood function at  $(\hat{\beta}_n, \hat{\gamma}_n, \hat{\eta}_n, \hat{F}_n)$  is larger or equal to the observed log-likelihood function at  $(\beta_0, \gamma_0, \eta_0, \tilde{F}_n)$ , we have

$$\frac{1}{n} l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{\eta}_n, \hat{F}_n) \geq \frac{1}{n} l_n(\beta_0, \gamma_0, \eta_0, \tilde{F}_n).$$

By taking limits on both sides, we obtain  $-K((\beta^*, \gamma^*, \eta^*, F^*), (\beta_0, \gamma_0, \eta_0, F_0)) \geq 0$ , where  $K(\cdot, \cdot)$  denotes the Kullback-Leibler information of  $(\beta^*, \gamma^*, \eta^*, F^*)$  with respect to the true parameters. Therefore, the following equality

$$\int_0^\infty R_{1k}(\beta_0, \gamma_0, F_0, W_i)\psi(W_k|\eta_0)dW_k = \int_0^\infty R_{1k}(\beta^*, \gamma^*, F^*, W_i)\psi(W_k|\eta^*)dW_k,$$

holds almost surely. According to the identifiability of the model, we obtain  $\beta^* = \beta_0, \gamma^* = \gamma_0, \eta^* = \eta_0$ , and  $F^* = F_0$ . Because every subsequence of  $n$  contains a further subsequence for  $(\hat{\beta}_n, \hat{\gamma}_n, \hat{\eta}_n, \hat{F}_n)$  which converges uniformly to  $(\beta_0, \gamma_0, \eta_0, F_0)$ , we have the convergence of the entire sequence. Hence, with probability one,  $\hat{\beta}_n \rightarrow \beta_0, \hat{\gamma}_n \rightarrow \gamma_0, \hat{\eta}_n \rightarrow \eta_0$  and  $\hat{F}_n(t) \rightarrow F_0(t)$  for every  $t \in [0, \infty)$ . Particularly, we have  $\sup_{t \in [0, \infty)} |\hat{F}_n(t) - F_0(t)| \rightarrow 0$  due to the continuity of  $F_0$ .  $\square$

## 7.2 Proof of Theorem 2

We first define a neighborhood of  $\phi_0$ , denoted by  $\mathcal{U}$ , as

$$\mathcal{U} = \{(\beta, \gamma, \eta, F) : \|\beta - \beta_0\| + \|\gamma - \gamma_0\| + |\eta - \eta_0| + \sup_{t \in [0, \infty)} |F(t) - F_0(t)| < \epsilon_0\}$$

for a very small constant  $\epsilon_0$ . Based on the consistency theorem,  $\hat{\phi}_n$  belongs to  $\mathcal{U}$  with probability close to 1 when the sample size  $n$  is large enough. We define a sequence of maps  $U_n$  mapping  $\mathcal{U}$  into  $l^\infty(\mathcal{H})$  as

$$\begin{aligned} U_n(\phi)[\mathbf{h}_1, \mathbf{h}_2, h_3, h_4] & \\ & \equiv n^{-1} \frac{d}{d\epsilon} l_n \left( \beta + \epsilon \mathbf{h}_1, \gamma + \epsilon \mathbf{h}_2, \eta + \epsilon h_3, F(t) + \epsilon \int_0^t Q_F[h_4](s) dF(s) \right) \Big|_{\epsilon=0} \\ & \equiv A_{n1}[\mathbf{h}_1] + A_{n2}[\mathbf{h}_2] + A_{n3}[h_3] + A_{n4}[h_4], \end{aligned}$$

where  $Q_F[h_4](t) = h_4(t) - \int_0^\infty h_4(s) dF(s)$ , and  $A_{np}, p = 1, \dots, 4$  are linear functionals on  $R^{d_1}, R^{d_2}, R$ , and  $BV[0, \infty)$ , which is the space of functions with finite total variation in  $[0, \infty)$ . Let  $l_\beta(\phi), l_\gamma(\phi), l_\eta(\phi)$ , and  $l_F(\phi)[\int Q_F[h_4] dF]$  denote the score functions for  $\beta, \gamma, \eta$ , and the score for  $F$  along the path  $F(t) + \epsilon \int_0^t Q_F[h_4](s) dF(s)$ , respectively, for a single cluster. Then

$$U_n(\phi)[\mathbf{h}_1, \mathbf{h}_2, h_3, h_4] = \mathcal{P}_n \left\{ \mathbf{h}_1^T l_\beta(\phi) + \mathbf{h}_2^T l_\gamma(\phi) + h_3 l_\eta(\phi) + l_F(\phi) \left[ \int Q_F[h_4] dF \right] \right\},$$

where  $\mathcal{P}_n$  denotes the empirical measure based on  $n$  independent clusters. Define

$$\begin{aligned} Q_{ij}(\beta, \gamma, F, W_i) &= I(Y_{ij} = \infty) \frac{G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i))}{G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i))} \\ &+ I(Y_{ij} < \infty) \left\{ \Delta_{ij} \frac{G''_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i))}{G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i))} \right. \\ &+ (1 - \Delta_{ij}) \frac{G'_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i))}{G_\alpha(-\theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i))} \left. \right\}. \end{aligned}$$

The explicit expression of the functionals  $A_{np}, p = 1, \dots, 4$  are given as follows

$$\begin{aligned}
A_{n1}[\mathbf{h}_1] &= n^{-1} \sum_{i=1}^n \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) R_{\boldsymbol{\beta}i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) [\mathbf{h}_1] \psi(W_i|\eta) dW_i \right\} \\
&\quad \times \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \psi(W_i|\eta) dW_i \right\}^{-1}, \\
A_{n2}[\mathbf{h}_2] &= n^{-1} \sum_{i=1}^n \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) R_{\boldsymbol{\gamma}i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) [\mathbf{h}_2] \psi(W_i|\eta) dW_i \right\} \\
&\quad \times \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \psi(W_i|\eta) dW_i \right\}^{-1}, \\
A_{n3}[h_3] &= n^{-1} \sum_{i=1}^n \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) R_{\eta i}(\eta, W_i) h_3 \psi(W_i|\eta) dW_i \right\} \\
&\quad \times \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \psi(W_i|\eta) dW_i \right\}^{-1},
\end{aligned}$$

and

$$\begin{aligned}
A_{n4}[h_4] &= n^{-1} \sum_{i=1}^n \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) R_{Fi}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \left[ \int Q_F[h_4] dF \right] \psi(W_i|\eta) dW_i \right\} \\
&\quad \times \left\{ \int_0^\infty R_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \psi(W_i|\eta) dW_i \right\}^{-1},
\end{aligned}$$

where

$$\begin{aligned}
&R_{\boldsymbol{\beta}i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) [\mathbf{h}_1] \\
&= \sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \mathbf{h}_1 (1 - \alpha \theta(\mathbf{Z}_{ij}|\alpha, W_i)) \\
&\quad \times \left\{ I(Y_{ij} < \infty) \Delta_{ij} - Q_{ij}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \theta(\mathbf{Z}_{ij}|\alpha, W_i) F(Y_{ij}|\mathbf{X}_{ij}, W_i) \right\}, \\
&R_{\boldsymbol{\gamma}i}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) [\mathbf{h}_2] \\
&= \sum_{j=1}^{n_i} \mathbf{X}_{ij}^T \mathbf{h}_2 I(Y_{ij} < \infty) \left\{ \Delta_{ij} (1 + \log(1 - F(Y_{ij}|\mathbf{X}_{ij}, W_i))) \right. \\
&\quad \left. + Q_{ij}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \theta(\mathbf{Z}_{ij}|\alpha, W_i) (1 - F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right. \\
&\quad \left. \times \log(1 - F(Y_{ij}|\mathbf{X}_{ij}, W_i)) \right\}, \\
&R_{Fi}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \left[ \int Q_F[h_4] dF \right] \\
&= \sum_{j=1}^{n_i} \left\{ -Q_{ij}(\boldsymbol{\beta}, \boldsymbol{\gamma}, F, W_i) \theta(\mathbf{Z}_{ij}|\alpha, W_i) W_i e^{\boldsymbol{\gamma}^T \mathbf{X}_{ij}} S(Y_{ij})^{W_i e^{\boldsymbol{\gamma}^T \mathbf{X}_{ij}} - 1} \right. \\
&\quad \left. + I(Y_{ij} < \infty) \Delta_{ij} (W_i e^{\boldsymbol{\gamma}^T \mathbf{X}_{ij}} - 1) \right\} \\
&\quad \times \int_0^{Y_{ij}} Q_F[h_4](y) dF(y) + I(Y_{ij} < \infty) \Delta_{ij} Q_F[h_4](Y_{ij}),
\end{aligned}$$

and  $R_{\eta i}(\eta, W_i)$  is the first derivative of  $\log \psi(W_i|\eta)$  with respect to  $\eta$ . Correspondingly, we define the limit map  $U : \mathcal{U} \rightarrow l^\infty(\mathcal{H})$  as

$$\begin{aligned} U(\boldsymbol{\phi})[\mathbf{h}_1, \mathbf{h}_2, h_3, h_4] & \\ & \equiv A_1[\mathbf{h}_1] + A_2[\mathbf{h}_2] + A_3[h_3] + A_4[h_4] \\ & \equiv \mathcal{P} \left\{ \mathbf{h}_1^T l_\beta(\boldsymbol{\phi}) + \mathbf{h}_2^T l_\gamma(\boldsymbol{\phi}) + h_3 l_\eta(\boldsymbol{\phi}) + l_F \left[ \int Q_F[h_4] dF \right] \right\}, \end{aligned}$$

where  $\mathcal{P}$  denotes the expectation of the empirical measure. It is easy to see that  $U_n(\widehat{\boldsymbol{\phi}}_n) = 0$  and  $U(\boldsymbol{\phi}_0) = 0$ .

We shall prove the theorem by verifying the following four properties stated in Theorem 3.3.1 of van der Vaart and Wellner (1996):

- (P1)  $\sqrt{n}(U_n - U)(\widehat{\boldsymbol{\phi}}_n) - \sqrt{n}(U_n - U)(\boldsymbol{\phi}_0) = o_p \left( 1 + \sqrt{n} \|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \sqrt{n} \|\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\| + \sqrt{n} |\widehat{\eta}_n - \eta_0| + \sup_{y \in [0, \infty)} |\widehat{F}_n(y) - F_0(y)| \right)$ .
- (P2)  $\sqrt{n}(U_n - U)(\boldsymbol{\phi}_0)$  converges to a tight random element  $\boldsymbol{\xi}$ .
- (P3)  $U(\boldsymbol{\phi})$  is Fréchet-differentiable at  $\boldsymbol{\phi}_0$ .
- (P4) The derivative of  $U(\boldsymbol{\phi})$  at  $\boldsymbol{\phi}_0$ , denoted by  $U'(\boldsymbol{\phi}_0)$ , is continuously invertible.

To verify property (P1), we first note that based on the explicit expressions of  $l_\beta, l_\gamma, l_\eta$ , and  $l_F$ , we have

$$\begin{aligned} & \left| \left\{ \mathbf{h}_1^T l_\beta(\boldsymbol{\phi}_1) + \mathbf{h}_2^T l_\gamma(\boldsymbol{\phi}_1) + h_3 l_\eta(\boldsymbol{\phi}_1) + l_F(\boldsymbol{\phi}_1) \left[ \int Q_{F_1}[h_4] dF_1 \right] \right\} \right. \\ & \quad \left. - \left\{ \mathbf{h}_1^T l_\beta(\boldsymbol{\phi}_2) + \mathbf{h}_2^T l_\gamma(\boldsymbol{\phi}_2) + h_3 l_\eta(\boldsymbol{\phi}_2) + l_F(\boldsymbol{\phi}_2) \left[ \int Q_{F_2}[h_4] dF_2 \right] \right\} \right| \\ & \leq O(1) \left\{ \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| - \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\| + |\eta_1 - \eta_2| + \sum_{j=1}^{n_i} |F_1(Y_{ij}) - F_2(Y_{ij})| + \int_0^\infty |F_1(t) - F_2(t)| dt \right\}, \end{aligned}$$

for any pair  $\boldsymbol{\phi}_1$  and  $\boldsymbol{\phi}_2 \in \mathcal{U}$ . Therefore

$$\begin{aligned} & \sup_{(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H}} E \left[ \left\{ \mathbf{h}_1^T l_\beta(\boldsymbol{\phi}) + \mathbf{h}_2^T l_\gamma(\boldsymbol{\phi}) + h_3 l_\eta(\boldsymbol{\phi}) + l_F(\boldsymbol{\phi}) \left[ \int Q_F[h_4] dF \right] \right. \right. \\ & \quad \left. \left. - \mathbf{h}_1^T l_\beta(\boldsymbol{\phi}_0) - \mathbf{h}_2^T l_\gamma(\boldsymbol{\phi}_0) - h_3 l_\eta(\boldsymbol{\phi}_0) - l_F(\boldsymbol{\phi}_0) \left[ \int Q_{F_0}[h_4] dF_0 \right] \right\}^2 \right] \end{aligned}$$

converges to zero if  $\|\beta - \beta_0\| + \|\gamma - \gamma_0\| + |\eta - \eta_0| + \sup_{y \in [0, \infty)} |F(y) - F_0(y)| \rightarrow 0$ . In addition, by the same arguments as in the consistency proof, the class of functions

$$\left\{ \begin{aligned} & \mathbf{h}_1^T l_\beta(\phi) + \mathbf{h}_2^T l_\gamma(\phi) + h_3 l_\eta(\phi) + l_F(\phi) \left[ \int Q_F[h_4] dF \right] \\ & - \mathbf{h}_1^T l_\beta(\phi_0) - \mathbf{h}_2^T l_\gamma(\phi_0) - h_3 l_\eta(\phi_0) - l_F(\phi_0) \left[ \int Q_{F_0}[h_4] dF_0 \right] : \\ & (\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H}, \phi \in \mathcal{U} \end{aligned} \right\}$$

is P-Donsker. Therefore, according to Lemma 3.3.5 of van der Vaart and Wellner (1996), property (P1) holds.

Property (P2) holds again because of the P-Donsker property of the class

$$\left\{ \mathbf{h}_1^T l_\beta(\phi_0) + \mathbf{h}_2^T l_\gamma(\phi_0) + h_3 l_\eta(\phi_0) + l_F(\phi_0) \left[ \int Q_{F_0}[h_4] dF_0 \right] : (\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H} \right\}.$$

Furthermore, the limit random element  $\xi$  is a Gaussian process indexed by  $(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H}$  and the covariance between  $\xi(\mathbf{h}_{11}, \mathbf{h}_{21}, h_{31}, h_{41})$  and  $\xi(\mathbf{h}_{12}, \mathbf{h}_{22}, h_{32}, h_{42})$  is equal to

$$\begin{aligned} & E \left[ \left\{ \mathbf{h}_{11}^T l_\beta(\phi_0) + \mathbf{h}_{21}^T l_\gamma(\phi_0) + h_{31} l_\eta(\phi_0) + l_F(\phi_0) \left[ \int Q_{F_0}[h_{41}] dF_0 \right] \right\} \right. \\ & \quad \times \left. \left\{ \mathbf{h}_{12}^T l_\beta(\phi_0) + \mathbf{h}_{22}^T l_\gamma(\phi_0) + h_{32} l_\eta(\phi_0) + l_F(\phi_0) \left[ \int Q_{F_0}[h_{42}] dF_0 \right] \right\} \right]. \end{aligned}$$

The Fréchet differentiability can be directly verified by using the smoothness of  $U(\phi)$ .

It remains to show that the derivative of  $U$  is continuously invertible at  $\phi_0$ . For convenience, we abbreviate  $U'(\phi_0)$  as  $U'$ . We note that  $U'$  is a map from the set  $\mathcal{U}' \equiv \{(\beta - \beta_0, \gamma - \gamma_0, \eta - \eta_0, F - F_0) : (\beta, \gamma, \eta, F) \in \mathcal{U}\}$  to  $l^\infty(\mathcal{H})$ . With straightforward calculations, we obtain

$$\begin{aligned} & U'(\beta - \beta_0, \gamma - \gamma_0, \eta - \eta_0, F - F_0)[\mathbf{h}_1, \mathbf{h}_2, h_3, h_4] \\ & = (\beta - \beta_0)^T \mathcal{Q}_1(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) + (\gamma - \gamma_0)^T \mathcal{Q}_2(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \\ & \quad + (\eta - \eta_0) \mathcal{Q}_3(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) + \int_0^\infty \mathcal{Q}_4(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) d(F - F_0), \end{aligned}$$

where

$$\mathcal{Q}_1(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) = E\{l_{\beta\beta}(\phi_0)\}\mathbf{h}_1 + E\{l_{\beta\gamma}(\phi_0)\}\mathbf{h}_2 + E\{l_{\beta\eta}(\phi_0)\}h_3 + E\{l_{\beta F}(\phi_0)\} \left[ \int Q_{F_0}[h_4] dF_0 \right],$$

$$\mathcal{Q}_2(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) = E\{l_{\gamma\beta}(\phi_0)\}\mathbf{h}_1 + E\{l_{\gamma\gamma}(\phi_0)\}\mathbf{h}_2 + E\{l_{\gamma\eta}(\phi_0)\}h_3 + E\{l_{\gamma F}(\phi_0)\} \left[ \int Q_{F_0}[h_4] dF_0 \right],$$

$$\mathcal{Q}_3(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) = E\{l_{\eta\beta}(\phi_0)\}\mathbf{h}_1 + E\{l_{\eta\gamma}(\phi_0)\}\mathbf{h}_2 + E\{l_{\eta\eta}(\phi_0)\}h_3 + E\{l_{\eta F}(\phi_0)\} \left[ \int Q_{F_0}[h_4] dF_0 \right],$$

and

$$\begin{aligned} \mathcal{Q}_4(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) &= E\{l_{F\beta}(\phi_0)[F - F_0]\}\mathbf{h}_1 + E\{l_{F\gamma}(\phi_0)[F - F_0]\}\mathbf{h}_2 \\ &\quad + E\{l_{F\eta}(\phi_0)[F - F_0]\}h_3 + E\{l_{FF}(\phi_0)[\int Q_{F_0}[h_4]dF_0, F - F_0]\}, \end{aligned}$$

where  $l_{\beta\beta}(\phi)$ ,  $l_{\beta\gamma}(\phi)$ ,  $l_{\beta\eta}(\phi)$ , and  $l_{\beta F}(\phi)[\int Q_F[h_4]dF]$  denote the first derivatives of  $l_\beta(\phi)$  with respect to  $\beta$ ,  $\gamma$ ,  $\eta$ , and  $F$  along the path  $F + \epsilon \int Q_F[h_4]dF$ , respectively;  $l_{\gamma\beta}(\phi)$ ,  $l_{\gamma\gamma}(\phi)$ ,  $l_{\gamma\eta}(\phi)$ , and  $l_{\gamma F}(\phi)[\int Q_F[h_4]dF]$  denote the first derivatives of  $l_\gamma(\phi)$  with respect to  $\beta$ ,  $\gamma$ ,  $\eta$ , and  $F$  along the path  $F + \epsilon \int Q_F[h_4]dF$ , respectively;  $l_{\eta\beta}(\phi)$ ,  $l_{\eta\gamma}(\phi)$ ,  $l_{\eta\eta}(\phi)$ , and  $l_{\eta F}(\phi)[\int Q_F[h_4]dF]$  denote the first derivatives of  $l_\eta(\phi)$  with respect to  $\beta$ ,  $\gamma$ ,  $\eta$ , and  $F$  along the path  $F + \epsilon \int Q_F[h_4]dF$ , respectively; and  $l_{F\beta}(\phi)[F - F_0]$ ,  $l_{F\gamma}(\phi)[F - F_0]$ ,  $l_{F\eta}(\phi)[F - F_0]$ , and  $l_{FF}(\phi)[\int Q_F[h_4]dF, F - F_0]$  denote the first derivatives of  $l_F(\phi)[F - F_0]$  with respect to  $\beta$ ,  $\gamma$ ,  $\eta$ , and  $F$  along the path  $F + \epsilon \int Q_F[h_4]dF$ , respectively.

If we treat  $(\beta - \beta_0, \gamma - \gamma_0, \eta - \eta_0, F - F_0)$  as an element in  $l^\infty(\mathcal{H})$  by defining its value at  $(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4)$  as

$$(\beta - \beta_0)^T \mathbf{h}_1 + (\gamma - \gamma_0)^T \mathbf{h}_2 + (\eta - \eta_0)h_3 + \int Q_{F_0}[h_4]d(F - F_0),$$

then the operator  $\mathcal{Q} \equiv (\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4)$  can be considered as a sum of a continuously invertible linear operator and a compact operator from  $l^\infty(\mathcal{H})$  to  $l^\infty(\mathcal{H})$ . Thus it suffices to prove that  $\mathcal{Q}$  is a one-to-one map (Rudin, 1973, pp. 99-103); that is, there exists some  $(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H}$  such that  $\mathcal{Q}(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) = 0$ , and then we need to show that  $\mathbf{h}_1 = \mathbf{0}$ ,  $\mathbf{h}_2 = \mathbf{0}$ ,  $h_3 = 0$ , and  $Q_{F_0}[h_4] = 0$ . If so, then

$$U'(\beta - \beta_0, \gamma - \gamma_0, \eta - \eta_0, F - F_0)[\mathbf{h}_1, \mathbf{h}_2, h_3, h_4] = 0$$

for any  $\phi \in \mathcal{U}$ . In particular, we choose

$$\beta = \beta_0 + \epsilon \mathbf{h}_1, \quad \gamma = \gamma_0 + \epsilon \mathbf{h}_2,$$

$$\eta = \eta_0 + \epsilon h_3, \quad F(y) = F_0(y) + \epsilon \int_0^y Q_{F_0}[h_4](s)dF_0(s)$$

for a small constant  $\epsilon$ . By the definition of  $U'$ , we obtain

$$E \left[ \left\{ \mathbf{h}_1^T l_\beta(\phi_0) + \mathbf{h}_2^T l_\gamma(\phi_0) + h_3 l_\eta(\phi_0) + l_F(\phi_0) \left[ \int Q_{F_0}[h_4]dF_0 \right] \right\}^2 \right] = 0.$$

Thus  $\mathbf{h}_1^T l_\beta(\phi_0) + \mathbf{h}_2^T l_\gamma(\phi_0) + h_3 l_\eta(\phi_0) + l_F(\phi_0) [\int Q_{F_0}[h_4] dF_0] = 0$  almost surely. Equivalently, the following equation holds

$$\int_0^\infty R_{1i}(\beta_0, \gamma_0, F_0, W_i) \left\{ R_{\beta_i}(\beta_0, \gamma_0, F_0, W_i)[\mathbf{h}_1] + R_{\gamma_i}(\beta_0, \gamma_0, F_0, W_i)[\mathbf{h}_2] + R_{\eta_i}(\eta_0, W_i)h_3 \right. \\ \left. + R_{F_i}(\beta_0, \gamma_0, F_0, W_i) [\int Q_{F_0}[h_4] dF_0] \right\} \psi(W_i|\eta) dW_i = 0. \quad (10)$$

We will show that this equation yields that  $\mathbf{h}_1 = \mathbf{0}$ ,  $\mathbf{h}_2 = \mathbf{0}$ ,  $h_3 = 0$ , and  $Q_{F_0}[h_4] = 0$ . We follow the ideas of proving the identifiability in the proof of consistency. Specifically, we let  $Y_{ij} = \infty$  for  $j = 1, \dots, k$ , for any integer  $k$  such that  $1 \leq k \leq n_i$ ; for those  $j > k$ , we perform the following action on the  $j$ th term on the left-hand side of (10). If  $\Delta_{ij} = 0$ , then we replace  $Y_{ij}$  with  $\infty$ ; if  $\Delta_{ij} = 1$ , then we integrate  $Y_{ij}$  from 0 to  $\infty$ . Then we sum the resulting quantity over all possible  $\Delta_{ij} : j = k+1, \dots, n_i$  and we can obtain  $\mathbf{h}_1 = \mathbf{0}$  and  $h_3 = 0$ . Next, we let  $\Delta_{ij} = 1$  and  $Y_{ij} = 0$  for  $j \leq k$  and perform the same action for  $j > k$  to obtain  $\mathbf{h}_2 = \mathbf{0}$ . Finally, we let  $\Delta_{ij} = 1$  for  $j \leq k$  and integrate  $Y_{ij}$  from 0 to  $y$ ; and we perform the same action described previously for  $j > k$  to obtain  $Q_{F_0}[h_4] = 0$ . Therefore, we have shown the invertibility of  $U'(\phi_0)$ .

By Theorem 3.3.1 of van der Vaart and Wellner (1996), we conclude that  $\sqrt{n}(\widehat{\beta}_n - \beta_0, \widehat{\gamma}_n - \gamma_0, \widehat{\eta}_n - \eta_0, \widehat{F}_n - F_0)$  converges weakly to  $-U'\xi$ . Specifically, for any  $(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4) \in \mathcal{H}$ , we have

$$\sqrt{n} \left\{ (\widehat{\beta}_n - \beta_0)^T \mathbf{h}_1 + (\widehat{\gamma}_n - \gamma_0)^T \mathbf{h}_2 + (\widehat{\eta}_n - \eta_0) h_3 + \int_0^\infty Q_{F_0}[h_4] d(\widehat{F}_n - F_0) \right\} \\ = -\sqrt{n} \left\{ l_\beta(\phi_0)^T \widetilde{\mathbf{h}}_1 + l_\gamma(\phi_0)^T \widetilde{\mathbf{h}}_2 + l_\eta(\phi_0) \widetilde{h}_3 + l_F(\phi_0) [Q_{F_0}[\widetilde{h}_4]] \right\} + o_p(1), \quad (11)$$

where  $(\widetilde{\mathbf{h}}_1, \widetilde{\mathbf{h}}_2, \widetilde{h}_3, \widetilde{h}_4) = (\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4)^{-1}(\mathbf{h}_1, \mathbf{h}_2, h_3, h_4)$ . By choosing  $h_4 = 0$  in (11), we conclude that  $\widehat{\beta}_n^T \mathbf{h}_1$ ,  $\widehat{\gamma}_n^T \mathbf{h}_2$ , and  $\widehat{\eta}_n h_3$  are asymptotic linear estimators of  $\beta_0^T \mathbf{h}_1$ ,  $\gamma_0^T \mathbf{h}_2$ , and  $\eta_0 h_3$ , respectively, and the corresponding influence functions are on the linear space spanned by the score functions. This implies that  $\widehat{\beta}_n$ ,  $\widehat{\gamma}_n$ , and  $\widehat{\eta}_n$  are semiparametrically efficient by the semiparametric efficiency theory (Bickel et al., 1993,

Ch. 3). □

**Acknowledgements** The authors are grateful to the Associate Editor and the anonymous referee for their careful reading of the earlier version of the manuscript, and providing valuable comments and suggestions. The authors thank the COGA investigators and Jean W. MacCluer for providing the COGA data from GAW14, which was supported in part by the National Institutes of Health grant GM31575. The research was supported in part by National Cancer Institute grant CA150698 (Diao), and a grant from the Research Grants Council of Hong Kong (Yin).

---

## References

- Abecasis, G. R., Cardon, L. R. and Cookson, W. O. C. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics*, *66*, 279-292.
- Bailey-Wilson, J. E., Thomas, D. and MacCluer, J. W. (2005). Genetic Analysis Workshop 14: Summarizing analyses comparing microsatellite and SNP marker loci for genome-wide scans. *Genetic Epidemiology*, *29(Suppl 1)*, S1-S132.
- Begleiter, H., Reich, T., Hesselbrock, V., Porjesz, B., Li, T.K., Schuckit, M. A., Edenberg, H. J. and Rice, J. P. (1995). The collaborative study on the genetics of alcoholism. *Alcohol Health Res World*, *19*, 228-236.
- Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, *47*, 501-515.
- Betensky, R. A. and Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics*, *57*, 282-286.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B*, *26*, 211-252.
- Chatterjee, N. and Shih, J. (2001). A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics*, *57*, 779-786.
- Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, *94*, 909-919.
- Chen, M. H., Ibrahim, J. G. and Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*, *80*, 101-126.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*, 141-151.

- Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, *102*, 560-572.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, *34*, 187-220.
- Diao, G. and Lin, D. Y. (2006). Semiparametric variance-component models for linkage and association analyses of censored trait data. *Genetic Epidemiology*, *30*, 570-581.
- Glidden, D. V. (2007). Pairwise dependence diagnostics for clustered failure-time data. *Biometrika*, *94*, 371-385.
- Gray, R. J. and Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics*, *45*, 899-904.
- Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, *79*, 531-541.
- Li, C. S., Taylor, J. M. G. and Judy, P. S. (2001). Identifiability of cure models. *Statistics & Probability Letters*, *54*, 389-395.
- Locatelli, I., Rosina, A., Lichtenstein, P. and Yashin, A. (2007). A correlated frailty model with long-term survivors for estimating the heritability of breast cancer. *Statistics in Medicine*, *26*, 3722-3734.
- Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley.
- Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, *95*, 449-465.
- Peng, Y. and Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, *56*, 237-243.
- Peng, Y., Taylor, J. M. G. and Yu, B. (2007). A marginal regression model for multivariate failure time data with a surviving fraction. *Lifetime Data Analysis*, *13*, 351-369.

- 
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: the art of scientific computing*, Second Edition. Cambridge: Cambridge University Press.
- Price, D. L. and Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, 20, 1515-1527.
- Rudin, W. (1973). *Functional Analysis*. New York: McGraw-Hill.
- Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56, 227-236.
- Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51, 899-907.
- Tsodikov, A. D. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, 54, 1508-1516.
- Tsodikov, A. D., Ibrahim, J. and Yakovlev, A. Y. (2003) Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, 98, 1063-1078.
- van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
- Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A. D. (1993). A simple stochastic model of tumour recurrence and its applications to data on premenopausal breast cancer. In *Biometrie et Analyse de Dormees Spatio-Temporelles*, 12, (eds. B. Asselain, M. Boniface, C. Duby, C. Lopez, J. P. Masson and J. Tranchefort). Société Française de Biométrie, ENSA Rennes, France, pp. 66-82.
- Yau, K. K. W. and Ng, A. S. K. (2001). Long-term survivor mixture model with random effects: application to a multicentre clinical trial of carcinoma. *Statistics in Medicine*, 20, 1591-1607.
- Yin, G. (2008). Bayesian transformation cure frailty models with multivariate failure time data. *Statistics in Medicine*, 27, 5929-5940.

- Yin, G. and Ibrahim, J. (2005). Cure rate models: a unified approach. *The Canadian Journal of Statistics*, 33, 559-570.
- Yu, B. and Peng, Y. (2008). Mixture cure models for multivariate survival data. *Computational Statistics and Data Analysis*, 52, 1524-1532.
- Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *Journal of the Royal Statistical Society: Series B*, 69, 507-564.
- Zeng, D., Yin, G. and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of American Statistical Association*, 101, 670-684.