

# Mixture cure survival models with dependent censoring

Yi Li,

*Harvard School of Public Health and Dana–Farber Cancer Institute, Boston, USA*

Ram C. Tiwari

*National Cancer Institute, Bethesda, USA*

and Subharup Guha

*Harvard School of Public Health, Boston, USA*

[Received July 2005. Final revision November 2006]

**Summary.** The paper is motivated by cure detection among the prostate cancer patients in the National Institutes of Health surveillance epidemiology and end results programme, wherein the main end point (e.g. deaths from prostate cancer) and the censoring causes (e.g. deaths from heart diseases) may be dependent. Although many researchers have studied the mixture survival model to analyse survival data with non-negligible cure fractions, none has studied the mixture cure model in the presence of dependent censoring. To account for such dependence, we propose a more general cure model that allows for dependent censoring. We derive the cure models from the perspective of competing risks and model the dependence between the censoring time and the survival time by using a class of Archimedean copula models. Within this framework, we consider the parameter estimation, the cure detection and the two-sample comparison of latency distributions in the presence of dependent censoring when a proportion of patients is deemed cured. Large sample results by using martingale theory are obtained. We examine the finite sample performance of the proposed methods via simulation and apply them to analyse the surveillance epidemiology and end results prostate cancer data.

**Keywords:** Archimedean copula model; Asymptotic normality; Consistency; Cure model; Dependent censoring; Latency distribution; Martingale processes; Testing cure fraction

## 1. Introduction

With the development of modern medicine and effective therapies, the curability of many early staged cancers (e.g. prostate, breast, head and neck cancers) is becoming a reality. In the prostate cancer study of the National Institutes of Health surveillance epidemiology and end results (SEER) programme, which motivated this paper, a large proportion of patients were deemed cured of prostate cancer, meaning that an individual will have little or no risk of experiencing the event of interest, e.g. death from prostate cancer. Survival models incorporating a cure fraction, which are often termed cure rate models, have emerged recently as a powerful statistical tool for analysing such cancer studies; see for example Kuk and Chen (1992), Maller and Zhou (1996), Peng and Dear (2000) and Sy and Taylor (2000), among others. Most of the current work stems from the mixture model that was originally proposed by Boag (1949) and Berkson

*Address for correspondence:* Yi Li, Department of Biostatistics, Harvard School of Public Health and Dana–Farber Cancer Institute, LW211, 44 Binney Street, Boston, MA 02115, USA.  
E-mail: yili@jimmy.harvard.edu

and Gage (1952). A key assumption in these classical cure models is that the survival time and the potential censoring time are independent.

However, such an independence assumption is often violated in observational studies. Indeed, a non-negligible portion of patients who were diagnosed with prostate cancer in our motivating example were found to have died from heart or cardiovascular diseases. A recent study (<http://www.thewbalchannel.com/healtharchive/4161401/detail.html>) has revealed that prostate cancer and cardiovascular disease may be linked through a common risk factor, namely high cholesterol levels. Therefore, it would seem implausible to assume independence between the main end point (e.g. deaths from prostate cancer) and the censoring causes (e.g. deaths from heart diseases). Statistical analysis is often hampered by dependent censoring as the classical models will not be valid, and ignoring dependent censoring will typically lead to foreseeable biases.

In view of scarce literature that deals with survival data with cure fractions in the presence of dependent censoring, we propose in this paper a more general cure model that allows for dependent censoring. In particular we derive the mixture cure model from the perspective of competing risks and model the dependence between the censoring time and the survival time by using a class of Archimedean copula models. Within this framework, we focus on cure detection, and a comparison of latency distributions in the presence of dependent censoring when a proportion of patients is deemed cured. To our knowledge, the methods proposed in this paper are the first attempt at cure modelling in the presence of dependent censoring.

Our paper is indeed motivated by Rivest and Wells (2001), which considered estimating marginal survival in the absence of cures. However, our focus is on unobserved long-term survivors who are obscured by censoring. Hence, our setting is more general and technically more delicate. The rest of the paper is structured as follows. In Section 2, we introduce a mixture cure model with the dependence of the censoring and survival times modelled by a class of Archimedean copula models. We also derive an estimator for estimating the survival function and the cure fraction with dependent censoring. In Section 3, we study the large sample properties of the estimators proposed and present the results of consistency and asymptotic normality. We conduct hypothesis tests in a two-sample comparison setting in Section 4. We illustrate the proposed methods with applications to the SEER data in Section 5 and assess the finite sample performance via simulation in Section 6. We conclude this paper with discussion and future work in Section 7. We relegate all the technical proofs and additional lemmas to Appendix A.

## 2. Mixture cure model with Archimedean dependence

Suppose that  $T$  is the survival time, e.g. the time from the diagnosis of prostate cancer, and  $U$  is the potential random censoring time, e.g. the duration of study or death from other causes (e.g. cardiac failure), with only  $X = \min(T, U)$  and censoring indicator  $\delta = I(X = T)$  observed in practice. Denote by  $F_T(t) = P(T \leq t)$  and  $F_U(t) = P(U \leq t)$  the cumulative distribution functions, and  $S_T(t) = P(T > t)$  and  $S_U(t) = P(U > t)$  the survival functions, for  $T$  and  $U$  respectively. Scientific research often centres on discerning  $F_T(t)$  while treating  $F_U(t)$  as a nuisance.

The mixture cure model assumes  $F_T$  to be an improper distribution over the entire real line and specifies

$$F_T(t) = p F_0(t) \tag{1}$$

or, equivalently,

$$S_T(t) = 1 - p + p S_0(t), \tag{2}$$

where  $0 < p < 1$ ,  $S_0(t) = 1 - F_0(t)$  and  $F_0(t)$  is a proper distribution such that  $\lim_{t \rightarrow \infty} \{F_0(t)\} = 1$ . Models (1) and (2) consider the study population as an unobservable mixture of patients who are deemed to be susceptible (non-cured) and non-susceptible (cured).  $1 - p$  corresponds to the fraction of cured, i.e. the point mass that  $T$  puts on  $\infty$ , and  $F_0(t)$  is the distribution for the non-cured patients, which is often termed the *latency* distribution.

We complete the model by specifying the dependence of the failure time  $T$  and censoring time  $U$  via a *strict* Archimedean copula model,

$$C(t, u) \stackrel{\text{def}}{=} P(T > t, U > u) = \phi^{-1}[\phi\{S_T(t)\} + \phi\{S_U(u)\}], \tag{3}$$

where  $\phi: [0, 1] \rightarrow [0, \infty]$  is a non-increasing function such that  $\phi(1) = 0$  and  $\phi(0) = \infty$ . Examples of  $\Phi$  include  $\phi(t) = -\log(t)$ , corresponding to independent censoring, the family of Clayton’s models with  $\phi(t) = (t^{-a} - 1)/a$  (for  $a > 0$ ), and the Frank family with

$$\phi(t) = -\log\left\{\frac{1 - \exp(-at)}{1 - \exp(-a)}\right\}$$

(for  $a > 0$ ). We adopt the Archimedean copula formulation to emphasize the functional independence of the parameterizations of the marginal distribution functions, which are governed by  $S_T$  and  $S_U$ , and the dependence structure, which is governed by a class of copula functions  $\phi$ . This formulation facilitates a derivation of the estimator for  $S_T$ , our main interest.

Suppose that we observe  $n$  independent and identically distributed data,  $(X_i, \delta_i)$ ,  $i = 1, \dots, n$ , and consider the counting processes  $N_i(t) = I(X_i \leq t, \delta_i = 1)$  and the at-risk processes  $Y_i(t) = I(X_i \geq t)$ . Denote by  $N(t) = \sum N_i(t)$  and  $Y(t) = \sum Y_i(t)$ . Introduce the filtration  $\mathcal{F}_t^n = \sigma\{N_i(s), Y_i(s+), 0 \leq s \leq t, i = 1, \dots, n\}$ , which contains the survival information up to time  $t$  for all  $n$  subjects and to which all the ensuing martingales and stopping times adapt. We denote the survival function for the observed times  $X_i$  by  $\pi(t) = P(X_i > t) = C(t, t)$ .

The following remarks heuristically discuss an estimator based on equation (3), whose large sample properties will be considered in the next section. Denote by  $\hat{S}_T$ , which will be defined in equation (5), and  $\hat{S}_U$  the estimates for  $S_T$  and  $S_U$  respectively, which are right continuous and piecewise constant functions with jumps only at the observed failures and censorings respectively. Denote by  $\hat{\pi}(t)$  the empirical estimate of  $\pi(t)$ , which is  $\hat{\pi}(t) = \sum_i I(X_i > t)/n = Y(t+)/n$ .

By equation (3), at each observed time point  $X_i$ ,  $i = 1, \dots, n$ ,

$$\phi\{\hat{S}_T(X_i)\} + \phi\{\hat{S}_U(X_i)\} = \phi\{\hat{\pi}(X_i)\}.$$

Assume that  $P(T = U) = 0$  (i.e. censoring and failure cannot occur at the same time almost surely). Then, at each observed failure time point  $X_i$  (such that  $\delta_i = 1$ ), we have  $\hat{S}_U(X_i^-) = \hat{S}_U(X_i)$  and

$$\begin{aligned} \phi\{\hat{S}_T(X_i)\} - \phi\{\hat{S}_T(X_i^-)\} &= \phi\{\hat{\pi}(X_i)\} - \phi\{\hat{\pi}(X_i^-)\} \\ &= \phi\left\{\frac{Y(X_i)}{n} - \frac{1}{n}\right\} - \phi\left\{\frac{Y(X_i)}{n}\right\}. \end{aligned} \tag{4}$$

Applying equation (4) recursively, the estimator  $\hat{S}_T$  can be written by using the form of counting processes

$$\hat{S}_T(t) = \phi^{-1}\left(\int_0^t I\{Y(s) > 0\} \left[\phi\left\{\frac{Y(s)}{n} - \frac{1}{n}\right\} - \phi\left\{\frac{Y(s)}{n}\right\}\right] dN(s)\right), \tag{5}$$

which corresponds to the estimator that was derived by Rivest and Wells (2001) in the absence of a cure fraction. When computing equation (5), we invoke the convention that  $0/0 = 0$  if necessary.

It is obvious that  $\hat{S}_T(t)$  is non-increasing and is a constant when  $t \geq \max_{\delta_i=1} \{X_i\} = X^{n*}$ , the largest observed failure time. In addition this constant is non-zero when the largest value among all the observed times  $(X_1, \dots, X_n)$ , which is denoted by

$$X^n = \sup_t \{t : Y(t) > 0\}, \tag{6}$$

is censored. We also introduce the right extremes

$$\tau_{F_0} = \sup_t \{t : F_0(t) < 1\}, \tag{7}$$

$$\tau_U = \sup_t \{t : F_U(t) < 1\}, \tag{8}$$

$$\tau_X = \sup_t \{t : \pi(t) > 0\}. \tag{9}$$

From equation (1), it follows that  $\tau_{F_0} = \sup_t \{t : F_T(t) < p\} = \sup_t \{t : S_T(t) > 1 - p\}$ . Throughout, denote by  $a \wedge b = \min(a, b)$  and  $a \vee b = \max(a, b)$  for two real numbers  $a$  and  $b$ .

For applications in the SEER prostate cancer data, we shall use  $\hat{S}_T(t)$  as defined in equation (5) to describe the survival pattern for the entire population of prostate cancer patients, use  $1 - \hat{S}_T(X^n)$  to estimate the cure fraction and use  $\hat{F}_T(t)/\hat{F}_T(X^n)$  to depict the latency distribution for the non-cured prostate cancer patients. To draw valid inference based on these estimates, we shall consider the following asymptotic results under some regularity conditions (which are listed in Appendix A.1).

- (a) The cure fraction can be consistently estimated on the basis of  $\hat{S}_T$  (as defined in equation (5)). Specifically,  $\hat{S}_T(X^n) \xrightarrow{\text{pr}} 1 - p$  or, equivalently,  $\hat{F}_T(X^n) \xrightarrow{\text{pr}} p$ .
- (b) The estimate of the cure fraction is asymptotically normally distributed, i.e.  $n^{1/2} \{ \hat{F}_T(X^n) - p \}$  or, equivalently,  $n^{1/2} \{ \hat{S}_T(X^n) - (1 - p) \}$  converges in distribution to a mean 0 normal random variable with a finite variance.
- (c) The estimate of the latency distribution is uniformly consistent and asymptotically normal. More specifically,

$$\sup_{t \in [0, \tau_X]} \left| \frac{\hat{F}_T(t)}{\hat{F}_T(X^n)} - F_0(t) \right| \xrightarrow{\text{pr}} 0,$$

and

$$n^{1/2} \left\{ \frac{\hat{F}_T(t \wedge X^n)}{\hat{F}_T(X^n)} - F_0(t \wedge X^n) \right\}$$

converges weakly to a tight Gaussian process on the Skorohod space  $D[0, \tau_X]$ .

To study the racial disparities in prostate cancer, we shall further consider a comparison of cure fractions between racial groups and propose a class of tests for testing the equality of the latency distribution  $F_0$  across racial groups along the lines of Li and Feng (2005).

### 3. Estimation and large sample results

Introduce the crude hazard function which is defined by

$$\begin{aligned} d\tilde{\Lambda}(t) &= \tilde{\lambda}(t) dt \\ &= P(t < T \leq t + dt | T > t, U > t), \end{aligned} \tag{10}$$

along with the martingale processes

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) d\tilde{\Lambda}(t).$$

Our later development relies heavily on the fact that  $M_i(t)$  are square integrable martingales with respect to filtration  $\mathcal{F}_t^n$  even when the survival time  $T$  and the censoring time  $U$  are dependent (Fleming and Harrington (1991), theorem 1.3.1). When  $T$  and  $U$  are dependent the crude hazard  $\tilde{\lambda}(t)$  may not be equal to the conventional hazard that is defined by

$$\lambda(t) = \frac{1}{dt} P(t < T \leq t + dt | T > t).$$

For example, consider a Clayton joint survival

$$C(t, u) = \{\exp(a\lambda_1 t) + \exp(a\lambda_2 u) - 1\}^{-1/a},$$

which corresponds to the Archimedean copula model with a generator  $\phi(t) = (t^{-a} - 1)/a$  and  $S_T(t) = \exp(-\lambda_1 t)$  and  $S_U(u) = \exp(-\lambda_2 u)$ , where  $a \geq 0$ ,  $\lambda_1 > 0$  and  $\lambda_2 > 0$ . It follows that the crude hazard

$$\tilde{\lambda}(t) = \lambda_1 \frac{\exp(a\lambda_1 t)}{\exp(a\lambda_1 t) + \exp(a\lambda_2 t) - 1}$$

differs from the conventional hazard  $\lambda_1$  when  $a \neq 0$ . Other counterexamples can be found in example 1.3.1 of Fleming and Harrington (1991).

Under the regularity conditions (conditions 1–6 that are listed in Appendix A.1) on  $S_T(t)$  (or  $F_T(t)$ ),  $\pi(t)$  and the copula function  $\phi$ , we can show that  $\hat{S}_T(t)$  as defined in equation (5) is a uniformly consistent estimator to  $S_T(t)$  on  $[0, \tau_X]$  (lemma 1 in Appendix A.2). It is thus natural to use the plateau of the estimated survival curve  $\hat{S}_T(X^n)$  to estimate the cure fraction  $1 - p$ . Indeed, further investigations reveal that this approach is proper if and only if the support of the latency distribution is covered by that of the censoring distribution. In particular, we have obtained a useful result (lemma 2 in Appendix A.2) that  $\hat{S}_T(X^n) \xrightarrow{PF} 1 - p$  if and only if  $\tau_{F_0} \leq \tau_X$ .

*Remark 1.* The result of consistency holds, as indicated in the proof of lemma 1, even when  $\pi(\tau_X) = 0$ , i.e. we consider the convergence over the entire support of the distribution of  $X$ , a useful result for the cure detection in our later development and a stronger result than theorem 1 of Rivest and Wells (2001) in the absence of cure.

*Remark 2.* As  $\tau_X$  characterizes the support of  $X = T \wedge U$ , we can further show (see lemma 3 in Appendix A.2) that  $\tau_X = \tau_U$  in the presence of cure under model (3), i.e. the supports of  $X$  and  $U$  coincide under an Archimedean model when the cure fraction is non-zero. Hence, when  $p < 1$  (or when the cure fraction is non-zero), it will be consistently estimated by  $1 - \hat{S}_T(X^n)$  if  $\tau_{F_0} \leq \tau_U$ , i.e. if the right extreme of the censoring distribution  $S_U$  exceeds that of the latency distribution  $F_0$ . Even when  $p = 1$ , similar proofs will show that  $1 - \hat{S}_T(X^n)$  consistently estimates  $p = 1$  provided that  $\tau_{F_0} \leq \tau_U$ . Thus, even in the absence of a cure fraction, equation (5) provides a consistent estimate for  $p$  and we shall not be misled by using equation (5) as long as  $\tau_{F_0} \leq \tau_U$ , reflecting a sufficient follow-up. Testing sufficient follow-up has been considered in detail by Maller and Zhou (1994).

We now consider the asymptotic normality of the proposed estimator for cure fractions. Define the stopped process

$$Z_n(t) = n^{1/2}[\phi\{\hat{S}_T(t \wedge X^n)\} - \phi\{S_T(t \wedge X^n)\}] \tag{11}$$

and the covariance function

$$\begin{aligned} \mathcal{C}(t_1, t_2) &= \int_0^{t_1 \wedge t_2} \pi(s) \phi' \{ \pi(s) \}^2 d\tilde{\Lambda}(s) \\ &+ 2 \int_0^{t_1 \wedge t_2} \int_0^s \pi(s) \{ 1 - \pi(u) \} \psi' \{ \pi(u) \} \psi' \{ \pi(s) \} d\tilde{\Lambda}(u) d\tilde{\Lambda}(s) \\ &+ 2 \int_0^{t_1 \wedge t_2} \int_0^s \phi' \{ \pi(u) \} \pi(s) \psi' \{ \pi(s) \} d\tilde{\Lambda}(u) d\tilde{\Lambda}(s) \\ &+ \int_{t_1 \wedge t_2}^{t_1 \vee t_2} \pi(s) \psi' \{ \pi(s) \} d\tilde{\Lambda}(s) \int_0^{t_1 \wedge t_2} [\{ 1 - \pi(u) \} \psi' \{ \pi(u) \} + \phi' \{ \pi(u) \}] d\tilde{\Lambda}(u) \end{aligned} \quad (12)$$

for  $0 \leq t_1, t_2, < \tau_X$ , where

$$\psi(s) \stackrel{\text{def}}{=} -s \phi'(s).$$

From lemma 1, this covariance function can be consistently estimated by replacing  $\pi(s)$  and  $d\tilde{\Lambda}(u)$  by their empirical counterparts  $\hat{\pi}(s)$  and  $I\{Y(u) > 0\} dN(u)/Y(u)$  respectively. We denote the obtained estimator by  $\hat{\mathcal{C}}(t_1, t_2)$ . Next denote the variance function  $v_0(t) = \mathcal{C}(t, t)$ , which coincides with the variance function that was obtained by Rivest and Wells (2001), namely the numerator of their equation (12). Assume that  $\lim_{t \rightarrow \tau_X} \{v_0(t)\} = v_0^\infty < \infty$  and

$$\mathcal{C}^\infty(t) \stackrel{\text{def}}{=} \lim_{v \rightarrow \tau_X} \{ \mathcal{C}(v, t) \} < \infty$$

for every  $t \in [0, \tau_X)$ .

*Theorem 1.*  $Z_n(t)$  converges weakly to  $I[0, \tau_X) Z(t) + I\{\tau_X\} Z^\infty$  on  $D[0, \tau_X]$ , where  $Z(t)$  is a tight Gaussian process with covariance function  $\mathcal{C}(t_1, t_2)$  and  $Z^\infty$  is a normal random variable with variance  $v_0^\infty$  and  $\text{cov}\{Z^\infty, Z(t)\} = \mathcal{C}^\infty(t)$ .

In practice, we shall use the plateau of the estimated survival curve  $\hat{S}_T(X^n)$  to estimate the cure fraction  $1 - p$  and denote the resulting estimator by  $\hat{p} = 1 - \hat{S}_T(X^n)$ . We show that  $\hat{p}$  will be asymptotically normally distributed.

*Theorem 2.* Assume that  $0 < p < 1$ . Then

$$n^{1/2} \{ \phi(1 - \hat{p}) - \phi(1 - p) \} \xrightarrow{d} Z^\infty, \tag{13}$$

where  $Z^\infty$  is a mean 0 normal random variable with variance  $v_0^\infty = \lim_{t \rightarrow \tau_X} \{v_0(t)\}$ . Furthermore,

$$n^{1/2} (\hat{p} - p) \xrightarrow{d} \frac{Z^\infty}{-\phi'(1 - p)}. \tag{14}$$

With the consistently estimated cure proportion, estimate the latency distribution by

$$\hat{F}_0(t) = \frac{\hat{F}_T(t)}{\hat{p}} = \frac{1 - \hat{S}_T(t)}{\hat{p}},$$

where  $\hat{S}_T(t)$  is as defined in equation (5). The large sample properties of  $\hat{F}_0(t)$  are summarized in the following theorem.

*Theorem 3.* For  $0 < p < 1$ ,

$$\sup_{[0, \tau_X]} |\hat{F}_0(t) - F_0(t)| \xrightarrow{\text{pr}} 0.$$

Moreover,

$$n^{1/2} \{ \hat{F}_0(t \wedge X^n) - F_0(t \wedge X^n) \} \xrightarrow{w} G(t)$$

on a Skorohod space  $D[0, \tau_X]$ , where  $\tau_X = \sup\{t : \pi(t) > 0\}$  and

$$G(t) = -\frac{Z(t)}{p \phi'\{S_T(t)\}} + \frac{Z^\infty\{1 - S_T(t)\}}{p^2 \phi'(1 - p)}$$

and  $Z(t)$  and  $Z^\infty$  are defined as in theorem 1.

These asymptotic results are the basis for our tests of comparing the survival patterns across various racial groups in the SEER data example. In practice, we can replace the largest observed time  $X^n$  by the largest observed failure time  $X^{n*}$ , in the estimator  $1 - \hat{p} = \hat{S}_T(X^n)$ , as  $P\{\hat{S}_T(X^n) = \hat{S}_T(X^{n*})\} = 1$  by the definition of  $\hat{S}_T$  as in equation (5).

### 4. Hypothesis testing

#### 4.1. Testing the existence of cure fraction

A natural question arising from cure modelling is whether the cure fraction exists. Hence, testing  $p < 1$  is of substantial interest. In what follows we derive a test for testing  $H_0 : p = 1$  against  $H_a : p < 1$  by extending Klebanov and Yakovlev’s test to the situation of dependent censoring. The derivations come at a small price by assuming that the underlying hazard for non-cured patients is a monotone function of time, which is a plausible assumption in most biological studies, as opposed to the restrictive non-decreasing hazard assumption that was made by Klebanov and Yakovlev (2005).

Under the mixture model (2),  $H_0$  is equivalent to  $H'_0 : \max_{0 < t < \tau_X}\{S_T(t) - S_0(t)\} = 0$ . For given data, our idea is to compute the  $1 - \alpha$  confidence interval for the difference  $\Delta = \max_t\{S_T(t) - S_0(t)\}$  and to reject  $H'_0$  at the  $\alpha$ -level. Klebanov and Yakovlev (2005) considered  $H''_0 : S(t_1) - S_0(t_1) = 0$ , where  $t_1$  is a prespecified constant. Though  $H'_0$  and  $H''_0$  are essentially equivalent, a data-driven choice of  $t_1$ , which magnifies the difference between these two survival functions, allows us to increase the power of the proposed test while controlling the significance level.

We first assume that the hazard

$$\lambda_0(t) = -\frac{d}{dt} \log\{S_0(t)\}$$

is a non-decreasing function in  $t$ , implying that  $-\log\{S_0(t)\}/t$  is a non-decreasing function. Hence, for any  $t_1 \geq t_0 > 0$ ,  $-\log\{S_0(t_1)\}/t_1 \geq -\log\{S_0(t_0)\}/t_0$  or

$$S_0(t_1) \leq S_0(t_0)^{t_1/t_0} \leq S_T(t_0)^{t_1/t_0} \tag{15}$$

and, from equation (2),  $S_T(t_1) \leq 1 - p + p S_T(t_0)^{t_1/t_0}$ . Therefore, we obtain an upper bound for  $p$ ,

$$p \leq \frac{1 - S_T(t_1)}{1 - S_T(t_0)^{t_1/t_0}}.$$

Since  $t_0$  and  $t_1$  are arbitrary,

$$p \leq \min_{0 < t_0 < t_1 < \tau_X} \left\{ \frac{1 - S_T(t_1)}{1 - S_T(t_0)^{t_1/t_0}} \right\} \stackrel{\text{def}}{=} \tilde{p}.$$

Because of the uniform consistency of  $\hat{S}_T$  that is defined in equation (5) (lemma 1) and the almost sure convergence of  $X^n$  to  $\tau_X$ ,  $\tilde{p}$  can be consistently estimated by the statistic

$$\hat{\tilde{p}} = \min \left[ \min_{0 < t_0 < t_1 < X^n} \left\{ \frac{1 - \hat{S}_T(t_1)}{1 - \hat{S}_T(t_0)^{t_1/t_0}}, 1 \right\} \right]. \tag{16}$$

From inequality (15), we have

$$S_T(t_1) - S_0(t_1) \geq S_T(t_1) - S_T(t_0)^{t_1/t_0} \stackrel{\text{def}}{=} \Delta(t_0, t_1),$$

for any fixed  $0 < t_0 < t_1$ . Our goal is to construct an asymptotic  $1 - \alpha$  confidence interval for  $\Delta(t_0, t_1)$  that is based on theorem 2 and we leave aside the question of choosing  $t_0$  and  $t_1$  for now. Let  $\hat{\Delta}(t_0, t_1) = S_T(t_1 \wedge X^n) - S_0(t_1 \wedge X^n) \geq S_T(t_1 \wedge X^n) - S_T(t_0 \wedge X^n)^{t_1/t_0}$  be the empirical counterpart of  $\Delta(t_0, t_1)$ . We show in Appendix A.6 that

$$P(\Delta \geq \mathcal{L}_n) = P\{\hat{\Delta}(t_0, t_1) \geq \mathcal{L}_n\} \geq 1 - \alpha, \tag{17}$$

where

$$\mathcal{L}_n = \hat{S}_T(t_1 \wedge X^n) - \hat{S}_T(t_0 \wedge X^n)^{t_1/t_0} - \left\{ 1 + \frac{t_1}{t_0} (1 + \varepsilon_0) S_T(t_0)^{t_1/t_0-1} \right\} \frac{D_{\alpha/2}}{n^{1/2}}, \tag{18}$$

where  $D_\alpha$  is the upper  $100\alpha$ -percentile of  $\sup_t |Z(t)/\phi'\{S_T(t)\}|$  (based on theorem 1). Since expression (17) holds true for any  $\varepsilon_0 > 0$ , we may let  $\varepsilon_0 \rightarrow 0$  in equation (18). In practice, the lower bound would be obtained by replacing the unknown  $S_T(t_0)$  in equation (18) with its consistent estimate  $\hat{S}_T(t_0 \wedge X^n)$ .

In contrast, if the hazard

$$\lambda_0(t) = -\frac{d}{dt} \log\{S_0(t)\}$$

is a non-increasing function of  $t$ , we can obtain that

$$P(\Delta \geq \mathcal{L}_n) \geq 1 - \alpha \tag{19}$$

where

$$\mathcal{L}_n = \hat{S}_T(t_0 \wedge X^n) - \hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} - \left[ \frac{D_{\alpha/2}}{n^{1/2}} + \min \left\{ \frac{t_0}{t_1} \hat{S}_T(t_1 \wedge X^n)^{t_0/t_1-1} \frac{D_{\alpha/2}}{n^{1/2}}, \left( \frac{D_{\alpha/2}}{n^{1/2}} \right)^{t_0/t_1} \right\} \right].$$

The detailed derivation has been deferred to Appendix A.7.

If the lower end of the  $1 - \alpha$  confidence interval in expression (17) or (19) (depending on whether  $\lambda_0(\cdot)$  is non-decreasing or non-increasing) is greater than 0, then hypothesis  $H_0$  would be rejected at a significance level of less than  $\alpha$ . To increase the power, the choice of  $t_0$  and  $t_1$  can be data driven. In particular, they can be chosen on the basis of equation (16) or equation (29) (again depending on the monotonicity of  $\lambda_0(\cdot)$ ) to minimize the upper bound of  $p$ . Indeed, that  $t_0$  and  $t_1$  in expression (17) or (19) are chosen by minimizing the lower bound of  $p$  (see equation (16) or equation (29)) does not affect the probabilistic arguments leading to expression (17) or (19) because the latter limit is based on the Kolmogorov distance  $n^{1/2} \sup |\hat{S}_T(\cdot \wedge X^n) - S_T(\cdot \wedge X^n)|$ , which is uniformly valid for all times  $t_1$  and  $t_0$ . Thus, the data-driven  $t_1$  and  $t_0$  will allow us to increase power while maintaining the proper level of significance.

#### 4.2. Comparisons of cure fractions and latency distributions

If the presence of a cure fraction is verified, it would also be of interest to compare the cure fractions and to study the latency distributions, for example, when evaluating racial disparities in cancer. We consider below a two-sample comparison scenario and adopt the notation that is used in the general cure model, except that we use an additional subscript  $i$  to indicate the racial groups. Specifically, we denote the time-to-event variables and censoring times by  $T_{ij}, U_{ij}$ ,  $i = 1, 2, j = 1, \dots, n_i$ , where, for example,  $i = 1$  corresponds to whites in the SEER data and  $i = 2$



to the non-whites, and  $j$  refers to the  $j$ th patient in his respective group. Let  $n = n_1 + n_2$ . We assume that  $n_1/n \rightarrow \gamma$  where  $\gamma$  is a fixed constant and  $0 < \gamma < 1$ . We further assume that the  $\{T_{ij}, U_{ij} : i = 1, 2, j = 1, \dots, n_i\}$  are all independent, but  $T_{ij}$  and  $U_{ij}$  can be dependent. Because of censoring, we observe only  $X_{ij} = T_{ij} \wedge U_{ij}$  and  $\delta_{ij} = I(T_{ij} \leq U_{ij})$ . We assume that the joint survival of  $T_{ij}$  and  $U_{ij}$  follows an Archimedean model. Define the right extremes  $\tau_{F_{0,1}}, \tau_{F_{0,2}}, \tau_{X,1}$  and  $\tau_{X,2}$ . To apply the large sample results that were obtained in the previous section, we assume that  $\tau_{F_{0,1}} \vee \tau_{F_{0,2}} \leq \tau_{X,1} \wedge \tau_{X,2} \stackrel{\text{def}}{=} \tau$ , i.e.  $[0, \tau]$  fully covers the supports of both latency distributions.

We first focus on the comparison of the cure fractions between two groups (e.g. whites *versus* non-whites) and formulate the hypotheses

$$H_0 : p_1 = p_2 (= p) \quad \text{versus} \quad H_1 : p_1 \neq p_2. \tag{20}$$

Denote by  $\hat{p}_i$  the estimate of  $p_i$  in arm  $i, i = 1, 2$ . Then under the null hypothesis in expression (20), from theorem 2, we have

$$n^{1/2}(\hat{p}_1 - \hat{p}_2) \xrightarrow{d} \frac{1}{-\phi'(1-p)} \left\{ \frac{Z_1^\infty}{\sqrt{\gamma}} - \frac{Z_2^\infty}{\sqrt{1-\gamma}} \right\},$$

where  $Z_1^\infty$  and  $Z_2^\infty$  are independent and are as defined in theorem 2 (with an added subscript for each group). Hence a Wald-type test statistic

$$(\hat{p}_1 - \hat{p}_2) / \sqrt{\left\{ \frac{(1-\gamma)\hat{v}_{0,1}^\infty + \gamma\hat{v}_{0,2}^\infty}{n\gamma(1-\gamma)\phi'(1-\hat{p})^2} \right\}} \tag{21}$$

will approximately follow a standard normal distribution. Here  $\hat{v}_{0,i}^\infty = \hat{C}_i(X_i^n, X_i^n)$  are the consistent estimates of  $v_{0,i}^\infty$  as defined in theorem 1 (with an added subscript for each group), which are readily available, and the pooled estimate  $\hat{p} = (n_1 \hat{p}_1 + n_2 \hat{p}_2)/n$ . Simulations (which, for brevity, are not reported) have verified that the distribution of statistic (21) under the null hypothesis is indeed a standard normal distribution.

Our next interest lies in comparing two latency distributions  $F_{0,i}(t) = P(T_{ij} \leq t | T_{ij} < \infty), i = 1, 2$ . For a two-sample comparison, the statistical test is formulated as

$$H_0 : F_{0,1} = F_{0,2} \quad \text{versus} \quad H_1 : F_{0,1} \neq F_{0,2}. \tag{22}$$

On the basis of  $(X_{ij}, \delta_{ij}), j = 1, \dots, n$ , we may estimate  $F_{0,i}$  by  $\hat{F}_{0,i}(t) = \hat{p}_i^{-1} \hat{F}_i(t)$ , where  $\hat{F}_i(t)$  is the estimator for the  $F_i$  based on equation (5) and  $\hat{p}_i = \hat{F}_i(X_i^{n_i*})$  is the consistent estimator for  $p_i$ , the estimated non-cure fraction in the  $i$ th arm.

Denote the pooled conditional distribution by

$$\hat{F}_{0,\text{pool}} = \frac{n_1 \hat{p}_1 \hat{F}_{0,1} + n_2 \hat{p}_2 \hat{F}_{0,2}}{n_1 \hat{p}_1 + n_2 \hat{p}_2}.$$

To test hypothesis  $H_0$  in expression (22), we define a class of test statistics to gauge the discrepancy between the two empirical distributions  $\hat{F}_{0,1}(\cdot)$  and  $\hat{F}_{0,2}$  as follows:

$$W_n = n^{1/2} \left\{ \int_0^\infty |\hat{F}_{0,1}(t) - \hat{F}_{0,2}(t)|^r d\hat{F}_{0,\text{pool}}(t) \right\}^{1/r}, \tag{23}$$

for  $r \geq 1$ , where  $r = 2$  corresponds to the Cramér–von Mises statistic that was proposed by Li and Feng (2005) and  $r = \infty$  corresponds to the Kolmogorov–Smirnov test  $W_n = \sup_{t \in [0, \tau]} \{n^{1/2} |\hat{F}_{0,1}(t) - \hat{F}_{0,2}(t)|\}$ .

To draw inference based on  $W_n$ , we derive the asymptotic distribution of  $W_n$ , under  $H_0 : F_{0,1} = F_{0,2} (= F_0)$ , which is summarized in the following theorem.

*Theorem 4.* Assume that  $n_1/n \rightarrow \gamma$ . Then, under the null hypothesis in expression (22),

$$W_n^r \Rightarrow \mathcal{X} \stackrel{\text{def}}{=} \int_0^\infty |\tilde{G}(t)|^r dF_0(t), \quad \text{if } r < \infty,$$

and

$$W_n \Rightarrow \mathcal{X} \stackrel{\text{def}}{=} \sup_{t \in [0, \tau]} |\tilde{G}(t)| \quad \text{if } r = \infty,$$

where the Gaussian process  $\tilde{G}$  is (distributionally) uniquely defined by

$$\tilde{G}(t) = \frac{1}{\sqrt{\gamma}} G_1(t) - \frac{1}{\sqrt{1-\gamma}} G_2(t)$$

and where  $G_i(\cdot), i = 1, 2$ , are independent Gaussian processes as defined in equation (26) (with an added subscript) in Appendix A.5.

The implementation of theorem 4 requires a simulation of  $\tilde{G}(t)$ . We give an algorithm below. Let  $\hat{S}_{T,i}$  and  $\hat{p}_i$  be consistent estimates for  $S_{T,i}$  and  $p_i$  respectively (the survival and the non-cure proportion in the  $i$ th arm), and let  $\hat{C}_i(t_1, t_2)$  be a consistent estimate of the covariance function  $C_i(t_1, t_2)$  for  $i = 1, 2$ , computed by using the aforementioned estimates of the unknown quantities. Let  $W_i(t)$  be a tight Gaussian process with covariance function  $\hat{C}_i(t_1, t_2)$  and let  $W_i^\infty$  be a normal random variable with variance  $\hat{v}_{0,i}^\infty$ . Then the stochastic process that is defined as

$$\hat{G}_i(t) = -\frac{W_i(t)}{\hat{p}_i \phi' \{ \hat{S}_{T,i}(t) \}} + \frac{\{1 - \hat{S}_{T,i}(t)\} W_i^\infty}{\hat{p}_i^2 \phi' (1 - \hat{p}_i)}$$

converges weakly to the process  $G_i(\cdot)$ . For a given grid of time points (e.g. observed failure times) the finite dimensional distribution of the process  $\hat{G}_i(\cdot)$  is indeed a multivariate normal distribution with a known covariance matrix and is, thus, easy to generate. Then the combination of two independent processes, namely

$$\hat{G}(t) = \frac{1}{\sqrt{\gamma}} \hat{G}_1(t) - \frac{1}{\sqrt{1-\gamma}} \hat{G}_2(t)$$

provides an approximation to  $\tilde{G}(\cdot)$ .

### 5. Analysis of surveillance epidemiology and end results prostate cancer data

We applied the developed methods to analyse the prostate cancer data in the SEER database, which was released in April 2004. We have focused on male prostate cancer patients in Connecticut and Detroit metropolitan area who were diagnosed between year 1973 and 2001 and during the early stages of the disease, excluding the cases where the cancer had spread to remote parts of the body. There were 91873 such cases, of which 75615 people were white. The analysis consisted of estimating the survival fractions, survival curves and latency distributions for the white and non-white subpopulations, targeting on racial disparities, which were one main end point of the SEER study. As pointed out by a referee, since the SEER study was not a randomized study, care must be taken to ensure that the demographics of the study population were comparable among the comparison groups. Indeed, our descriptive analysis revealed that important outcome-related variables, such as the age of diagnostics, grade of the tumour and types of treatment received, were evenly distributed between these two subpopulations (Table 1). Hence, the differences detected in cures and latency distributions could be attributed to racial disparities.

**Table 1.** Comparison of important demographics across whites and non-whites

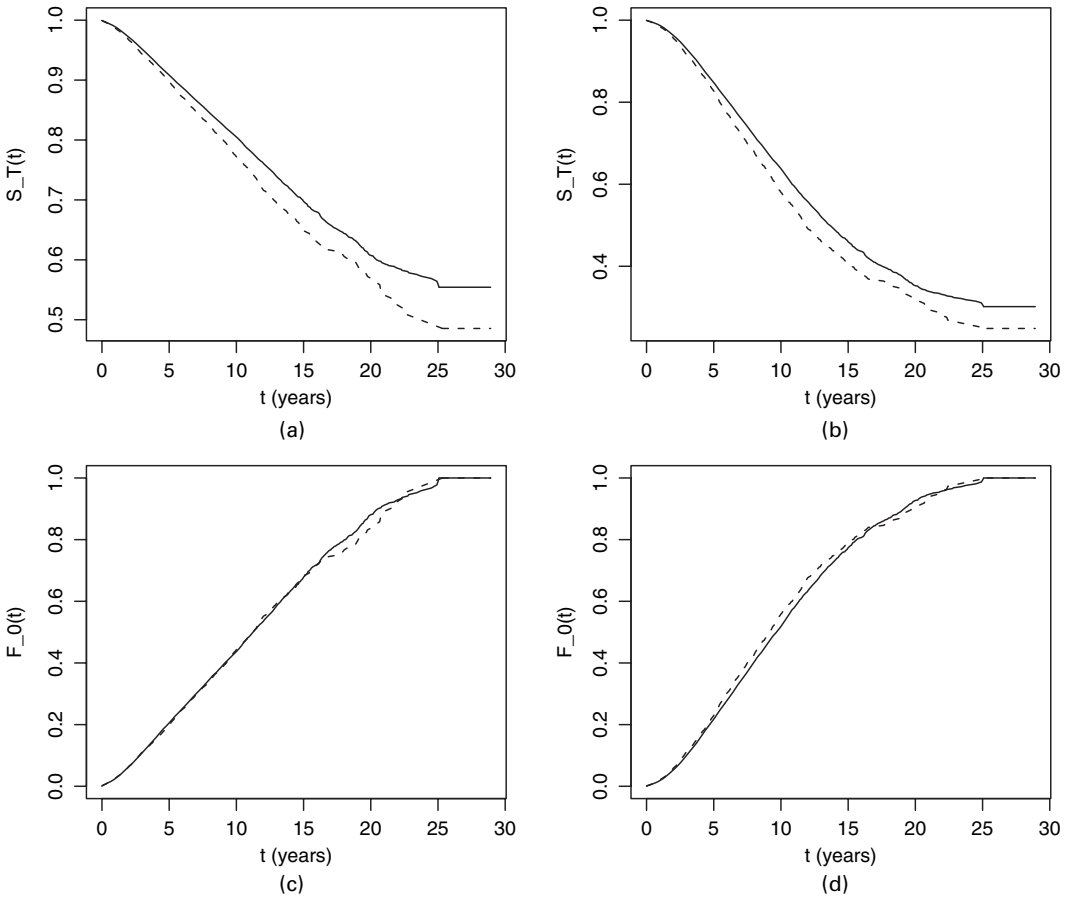
	<i>Results for whites (N = 75615)</i>	<i>Results for non-whites (N = 16258)</i>
<i>Treatment</i>		
No radiation and/or cancer-directed surgery	91.35%	90.53%
Radiation before surgery	0.19%	0.17%
Radiation after surgery	7.90%	8.29%
Radiation before and after surgery	0.017%	0.012%
Intraoperative radiation	0.007%	0.006%
Intraoperative radiation with other radiation before or after surgery	0.003%	0.006%
Sequence unknown, but both were given	0.52%	0.97%
<i>Tumour grade</i>		
1	18.29%	16.60%
2	53.63%	52.98%
3	15.52%	17.93%
4	0.62%	0.69%
Unknown	11.92%	11.78%
<i>Age at diagnosis (years)</i>		
Mean	70.1	68.2
Standard deviation	9.0	9.0
Median	71.0	68.0
25th percentile	64.0	62.0
75th percentile	76.0	74.0

About 37% of the censored observations were due to death from other causes, with cardiovascular disease (CVD) being the major cause of these deaths. As prostate cancer and CVD share common risk factors, e.g. a high intake of fat, we assumed various strengths of correlation between time to prostate cancer death and the censoring time. For illustration, we considered both Frank's and Clayton's families of Archimedean copulas, with the correlation parameter chosen such that Kendall's  $\tau$  ranged from 0 to 0.47. As expected, the point estimates of the cure fraction varied as the strength of the dependence varied—the weaker the dependence is, the larger the estimate of the cure fraction is. This indeed has some important implications in the evaluation of the progress made in cancer. With the mortality rate for CVD having a decreasing trend, fewer censorings would be due to CVD. Assuming a positive dependence between CVD and prostate cancer, we might see that the overall dependence among the prostate cancer deaths and censoring would become weaker as fewer censorings are due to CVD. Hence, the data would yield a trend of higher rates of cancer cure, indicating overall progress against cancer. In contrast, if the dependence increases, more deaths that could have resulted from CVD are transferred to cancer. As a result, we would see a faster decrease in the cure fraction estimates, thus artificially indicating that we are not making decent progress in curing cancer, though in reality there might be a higher true cure rate. Theoretical justifications for this phenomenon will be given in Section 7.

Table 2 presents the estimated cure fractions for the two subpopulations for an analysis based on Frank's family. Assuming that the censoring mechanisms correspond to approximately equal values of Kendall's  $\tau$ , the cure fractions for whites are uniformly higher than those for non-whites.

**Table 2.** Cure fractions based on Frank's family of Archimedean copulas

Kendall's $\tau$	Results for whites		Results for non-whites	
	Estimate	Standard error	Estimate	Standard error
0	0.55	0.015	0.49	0.039
0.01	0.55	0.015	0.48	0.039
0.03	0.50	0.015	0.43	0.039
0.09	0.50	0.015	0.43	0.039
0.12	0.44	0.015	0.37	0.038
0.16	0.39	0.015	0.32	0.035
0.19	0.39	0.015	0.32	0.035
0.32	0.30	0.013	0.25	0.030
0.42	0.22	0.010	0.18	0.023
0.47	0.20	0.010	0.16	0.021



**Fig. 1.** Summary of prostate cancer results for Archimedean copulas based on Frank's family (—, whites; ----, non-whites): (a) survival curves for Kendall's  $\tau = 0.00$ ; (b) survival curves for Kendall's  $\tau = 0.32$ ; (c) latency distributions for Kendall's  $\tau = 0.00$ ; (d) latency distributions for Kendall's  $\tau = 0.32$

**Table 3.** Results based on Frank’s family of Archimedean copulas

<i>Kendall’s <math>\tau</math></i>	<i>Cramér–von Mises test</i>	<i>p-value</i>	<i>Kolmogorov–Smirnov test</i>	<i>p-value</i>	<i>Wald z</i>	<i>p-value (<math>\times 10^{-3}</math>)</i>	<i>95% credible interval for <math>\Delta</math></i>
0	27.79	0.26	15.97	0.06	−3.04	2.365	(0.195,1)
0.01	27.67	0.25	15.94	0.06	−3.04	2.365	(0.195,1)
0.03	23.18	0.26	14.28	0.07	−2.93	3.389	(0.198,1)
0.09	23.18	0.26	14.28	0.07	−2.93	3.389	(0.198,1)
0.12	24.51	0.19	12.53	0.08	−2.94	3.282	(0.198,1)
0.16	30.74	0.09	11.44	0.07	−3.05	2.288	(0.196,1)
0.19	30.74	0.10	11.44	0.09	−3.05	2.288	(0.194,1)
0.32	48.01	0.01	13.04	0.01	−3.47	0.520	(0.177,1)
0.42	66.24	<0.01	13.47	<0.01	−4.31	0.016	(0.148,1)
0.47	69.82	<0.01	13.42	<0.01	−4.59	0.004	(0.137,1)

Table 2 also indicates that naïvely assuming an independent model may give a misleading result and that drawing inference does need to take the dependence between the survival and informative censoring into account. Fig. 1 plots the survival curves and the latency distributions for the two subpopulations. The graphs indicate that the prostate cancer survival rates are higher for whites, irrespective of the degree of dependence in the censoring mechanism.

Table 3 displays the results for dependent censoring under Frank’s family for various values of Kendall’s  $\tau$ . The second to fifth columns test the null hypothesis (22) that the latency distribution for whites ( $F_{0,1}$ ) equals that for non-whites ( $F_{0,2}$ ). The second and third columns display the Cramer–von Mises test statistics that are defined in equation (23) with the  $p$ -values estimated by using theorem 4. The fourth and fifth columns present the results for the Kolmogorov–Smirnov test. For  $\tau \leq 0.19$ , there is no strong evidence at the 1% level of significance that the latency distributions of whites and non-whites are different. In contrast, there is strong evidence of a difference in the latency distributions if the dependence between the survival time and the censoring is large (e.g. when  $\tau \geq 0.32$ ).

The sixth and seventh columns present the results for testing whether whites and non-whites have the same cure fractions. The theory is developed in Section 4.2 and the test statistic is defined in expression (21). There is strong evidence that the cure fractions are different for the two subpopulations. Using the theoretical results that were derived in Section 4.1, we tested whether or not a cure fraction exists for the entire population. For this, we computed a 95% one-sided confidence interval for  $\Delta = \max_t \{S_T(t) - S_0(t)\}$ , using all 91 873 cases in the data set. Expression (17) was used to compute the bounds since the estimated hazard  $\hat{\lambda}_0(\cdot)$  was found to be non-decreasing. Expression (16) was used to find suitable choices for  $t_0 < t_1$  by a stochastic search. These intervals are specified in the last column of Table 3. The lower bounds of the intervals are all positive, implying that there is significant evidence at the 5% level that a cure fraction exists for the entire population. Analyses assuming dependent censoring under Clayton’s family of copulas were similar to those obtained under Frank’s family and lead to the same conclusions.

### 6. Simulation study

We investigated by simulation the finite sample behaviour of the cure fraction estimator, i.e.  $\hat{p} = 1 - \hat{S}_T(X^n)$ , where  $\hat{S}_T$  is defined in equation (5). Theorem 2 shows that this estimator is consistent and asymptotically normal, with expression (14) specifying its asymptotic variance.

We simulated the survival data by generating independent censoring times  $U_i$ , where  $i = 1, \dots, n$ , from the exponential distribution with hazard rate  $r$  (mean  $1/r$ ). Conditional on the censoring times, the failure times  $X_i$  were generated for dependent censoring under Frank's copula model with various correlation parameters  $a = 0, 2.1, 5.7$ , corresponding roughly to Kendall's  $\tau = 0, 0.2, 0.47$ . The latency distribution of the failure times was exponential with mean 1 and truncated at  $\tau_{F_0} = 2$ . The true cure fraction of the failures was  $p = 0.3$ . The rate  $r$  in the censoring distribution was chosen to be 1, 0.5 and 0.2, resulting in roughly 60%, 40% and 20% censoring among 'non-cured' patients. For each simulated data, the estimate  $\hat{p} = 1 - \hat{S}_T(X^n)$  was then computed and its asymptotic variance was computed using expression (14).

These steps were repeated for 3000 replications to obtain estimates of  $p$  based on the 3000 different data sets. The empirical variance of  $\hat{p}$  was computed and compared with the average asymptotic variance. Table 4 presents the results for various combinations of sample size, Frank family parameter  $a$  and censoring rate  $r$ .

For any given  $(a, r)$  pair, estimate  $\hat{p}$  is found to approach the true value as the sample size  $n$  increases. Additionally, the difference between the empirical and asymptotic standard errors tends to 0, and the empirical coverage probabilities of the 95% confidence intervals approach the nominal value of the confidence level. These results verify the validity of theorem 2. For a given value of dependence parameter  $a$  and sample size  $n$ , the standard errors decrease with

**Table 4.** Summary of simulation results investigating the asymptotic behaviour of estimator  $\hat{p} = 1 - \hat{S}_T(X^n)$ , where  $\hat{S}_T(\cdot)$  is defined in equation (5) and  $X^n$  are the largest observed times†

$a$	$r$	$n$	$\hat{p}$	Empirical standard error	Asymptotic standard error	95% credible interval coverage
0	1	50	0.311792	0.1140858	0.0981302	0.865
0	1	100	0.3035328	0.08593271	0.07637336	0.903
0	1	500	0.3013318	0.03856888	0.03702753	0.936
2.1	1	50	0.3086511	0.1049240	0.09002797	0.878
2.1	1	100	0.3049256	0.07592004	0.06883757	0.908
2.1	1	500	0.3008560	0.03328404	0.0322699	0.939
5.7	1	50	0.3154014	0.1001547	0.08272287	0.882
5.7	1	100	0.309741	0.06823369	0.06206997	0.910
5.7	1	500	0.3025891	0.02880224	0.02869691	0.949
0	0.5	50	0.2999443	0.08688357	0.08017418	0.908
0	0.5	100	0.2985752	0.06125795	0.05896112	0.935
0	0.5	500	0.2998565	0.02743993	0.02707487	0.942
2.1	0.5	50	0.3032127	0.07910919	0.07369643	0.919
2.1	0.5	100	0.2989564	0.0569155	0.05332041	0.924
2.1	0.5	500	0.3007000	0.0242311	0.02439310	0.953
5.7	0.5	50	0.3027497	0.06907557	0.06668318	0.938
5.7	0.5	100	0.3011927	0.04985439	0.04809085	0.935
5.7	0.5	500	0.3003604	0.0215772	0.02184662	0.950
0	0.2	50	0.3001357	0.07285168	0.06916037	0.927
0	0.2	100	0.3006226	0.0506231	0.0501023	0.942
0	0.2	500	0.2998538	0.02258296	0.02277172	0.952
2.1	0.2	50	0.2996179	0.06919427	0.06588285	0.932
2.1	0.2	100	0.2985278	0.04925562	0.04750322	0.936
2.1	0.2	500	0.2995015	0.02154124	0.02160814	0.946
5.7	0.2	50	0.3028895	0.06584249	0.06286433	0.930
5.7	0.2	100	0.2988817	0.04722508	0.04532081	0.931
5.7	0.2	500	0.2993998	0.02109654	0.02064824	0.943

†The true cure fraction was assumed to be  $p = 0.3$ .

**Table 5.** Summary of simulation results verifying the covariance expression (12) for  $\text{cov}\{\hat{S}_T(1), \hat{S}_T(2)\}$ , where  $\hat{S}_T(\cdot)$  is defined in equation (5)<sup>†</sup>

$a$	$r$	$n$	$\hat{S}_T(t_1=1)$	$\hat{S}_T(t_2=2)$	Empirical coverage	Asymptotic coverage
0	1	50	0.4970387	0.3155806	0.004653271	0.005108764
0	1	100	0.4872239	0.2907100	0.001973786	0.002541864
0	1	500	0.4904664	0.3000590	0.000481298	0.000543318
2.1	1	50	0.4914768	0.2936169	0.004572677	0.004072854
2.1	1	100	0.4904447	0.3107039	0.001998270	0.002214692
2.1	1	500	0.4949633	0.3044195	0.000530253	0.000436514
5.7	1	50	0.5033339	0.3184996	0.002683341	0.003344544
5.7	1	100	0.4854107	0.3039750	0.002001305	0.001674436
5.7	1	500	0.4882282	0.3029674	0.000385107	0.000323667
0	0.5	50	0.4856533	0.2982809	0.003532756	0.003786317
0	0.5	100	0.4885342	0.3023957	0.001884913	0.001975122
0	0.5	500	0.4871333	0.3004378	0.000432969	0.000403036
2.1	0.5	50	0.4933665	0.3090195	0.003304896	0.003379181
2.1	0.5	100	0.4914137	0.2998028	0.001809866	0.001710108
2.1	0.5	500	0.4852423	0.3003169	0.000367073	0.000354900
5.7	0.5	50	0.4997257	0.3082358	0.003449204	0.002844627
5.7	0.5	100	0.4870287	0.3060905	0.001472245	0.001519335
5.7	0.5	500	0.5007257	0.3022358	0.000324920	0.000294462
0	0.2	50	0.4902883	0.3038794	0.003792067	0.003259159
0	0.2	100	0.4833	0.2938062	0.001939279	0.001637402
0	0.2	500	0.4873104	0.2995824	0.000296443	0.000340456
2.1	0.2	50	0.4880728	0.2976328	0.003246520	0.00301053
2.1	0.2	100	0.4860294	0.2979742	0.001349038	0.001563168
2.1	0.2	500	0.4853753	0.2998419	0.000429316	0.000322165
5.7	0.2	50	0.4862388	0.2987405	0.003476443	0.002784509
5.7	0.2	100	0.4871271	0.3011671	0.001435091	0.001484796
5.7	0.2	500	0.4897375	0.3021698	0.000309732	0.000308627

<sup>†</sup>The true values are  $S_T(1) = 0.4882$  and  $S_T(2) = 0.3$ .

censoring rate  $r$ . This is reasonable because a smaller rate of censoring implies stochastically greater censoring times and a smaller proportion of censored observations, resulting in a more precise estimate of the cure fraction.

Simulations were also performed to verify the covariance structure that was derived in expression (12). In particular, Table 5 uses the simulated data to verify expression (12) for the covariance between survival function estimates at  $t_1 = 1$  and  $t_2 = 2$ . For any given  $(a, r)$  pair, the empirical covariance matches well the asymptotic value, especially as  $n$  grows.

### 7. Discussion and future work

This paper proposes a mixture cure model which allows dependent censoring. In particular, we have considered the parameter estimation, the cure detection and the comparison of latency distributions in the presence of dependent censoring when a proportion of patients is deemed cured. The dependence between the survival time and its potential censoring time is modelled by using a class of Archimedean copula models with a known  $\phi$ -function. This allows us to vary the degree of dependence smoothly and to investigate the effect of the degree of such dependence on cure detections and comparisons. In practice, however, selecting a right  $\phi$ -function in the copula is often hampered by the fact that, with the current data on  $(X, \delta)$ , the copula model is not identifiable (Tsiatis, 1978). In some applications where both the censoring and the failure

times were observed for a subsample (Bartholomew, 1957), a suitable copula function could be identified with this additional information. Indeed, recent literature has focused on imposing additional assumptions and models on the dependence structure between the failure time and the dependent censoring variable to make the problem identifiable. For example, Moeschberger and Klein (1995) proposed models for dependent competing risks, whereas Hogan and Laird (1997) and Little (1995) proposed joint likelihood methods that simultaneously model the outcome of interest and the missing data process with both longitudinal and failure time outcomes. Scharfstein *et al.* (2001) and Scharfstein and Robins (2002) conducted sensitivity analyses based on unmeasured factors, and Siannis *et al.* (2005) considered sensitivity analysis for informative censoring in parametric survival models. In contrast, Kalbfleisch and Prentice (2002), page 262, noted that time-dependent covariates (e.g. the information about the time of prostate cancer progression and heart failure) will help in estimating the interrelation between the censoring and failure times. None have considered cure models. Hence, model diagnostics and choices of copulas in the presence of cures is a promising research area which may be pursued in the framework that has been established in this paper and along with the lines of these aforementioned references

Additionally, on the basis of the established analytical framework in this paper, it would be feasible to conduct bias analysis when the dependence structure between survival and censoring times is misspecified. In particular, we can quantify the biases in the estimates of cure fractions for such misspecifications.

Using the same argument in lemma 1, we can show that for any  $\phi$ -function (which satisfies the regularity conditions 1–5), the estimate based on equation (5) converges uniformly to

$$S_T^*(t) = \phi^{-1} \left[ - \int_0^t \phi' \{ \pi(s) \} \pi(s) d\tilde{\Lambda}(s) \right].$$

When  $\phi$  is misspecified,  $S_T^*$  may not be equal to  $S_T$ , the true survival function. Hence the estimate of cure converges to  $1 - p^* = \lim_{t \rightarrow \tau_X} \{ S_T^*(t) \}$ , which may not be equal to the true cure fraction  $1 - p = \lim_{t \rightarrow \tau_X} \{ S_T(t) \}$ . Analogous to corollary 6.1 of Zheng and Klein (1996) and proposition 2 of Rivest and Wells (2001), we can characterize the asymptotic effect of changing the level of dependence between  $T$  and  $U$  on estimating the cure fractions. Specifically, let  $\phi_1$  and  $\phi_2$  be two functions used in equation (5). If  $\phi_1'(t)/\phi_2'(t)$  increases in  $t$  then the asymptotic limit of cure fraction  $1 - p_1^* \geq 1 - p_2^*$ , which follows from proposition 2 of Rivest and Wells (2001) by taking  $t \rightarrow \tau_X$ . Genest and MacKay (1986) showed that  $\phi_1'(t)/\phi_2'(t) \uparrow t$  implies that  $\phi_1$  corresponds to less dependence between  $T$  and  $U$  than  $\phi_2$  under equation (3). This result is of substantial interest as it reveals that, under undetected positive dependence between  $T$  and  $U$ , failing to account for such dependence (e.g. the Kaplan–Meier estimate of cure fraction that was proposed by Maller and Zhou (1992)) will tend to overestimate the true cure fraction. In contrast, if there is negative dependence between  $T$  and  $U$ , a naïve Kaplan–Meier estimate will underestimate the true cure fraction.

We have focused on the strict Archimedean copula model family because it encompasses the well-known bivariate families, including the Clayton and the Frank families. It remains an open question whether our results will be valid for non-strict Archimedean copula models, wherein  $\phi(0) < \infty$ . Our future work also involves generalizing the proposed work beyond the Archimedean family to any copula model along the line of Rivest and Wells (2001). Specifically, suppose that  $P(T > t, U > u) = \mathcal{C}\{S_T(t), S_U(u)\}$ , where the copula  $\mathcal{C}$  does not necessarily belong to the Archimedean family. Define  $\phi_{\mathcal{C}}(\cdot)$  satisfying

$$\mathcal{C}_{10}[s, S_U\{S_T^{-1}(s)\}] \phi_{\mathcal{C}}(\mathcal{C}[s, S_U\{S_T^{-1}(s)\}]) = \phi_{\mathcal{C}}(s)$$



with boundary condition  $\phi_C(1) = 0$ . It can be shown that the estimator (5) that is calculated by using such  $\phi_C$  will be a consistent estimator for  $S_T(t)$  (see Rivest and Wells (2001)). However, in general, no closed form solutions exist for  $\phi_C$  if  $\mathcal{C}$  does not belong to the Archimedean family, and numerical methods are needed.

**Acknowledgements**

The authors thank the Joint Editor, an Associate Editor and two referees for their insightful suggestions.

**Appendix A: Technical details**

*A.1. Regularity conditions*

We impose the following regularity conditions on  $S_T(t)$  (or  $F_T(t)$ ),  $\pi(t)$  and the copula function  $\phi$ .

- (a)  $\phi$  is strictly decreasing on  $(0, 1]$  and is sufficiently smooth in the following sense: the first two derivatives of  $\phi(s)$  and  $\psi(s) \stackrel{\text{def}}{=} -s \phi'(s)$  are bounded for  $s \in [\varepsilon, 1]$  where  $\varepsilon > 0$  is arbitrary. In addition, the first derivative of  $\phi(s)$  is bounded away from 0 on  $[0, 1]$  (condition 1).
- (b)  $0 < \int_0^{\tau_X} \psi\{\pi(s)\}^k d\tilde{\Lambda}(s) < \infty$  for  $k = 0, 1, 2$ , where  $\tilde{\Lambda}(\cdot)$  is defined in equation (10) (condition 2).
- (c)  $\int_0^{\tau_X} |\psi'\{\pi(s)\}| d\tilde{\Lambda}(s) < \infty$  (condition 3).
- (d)  $\lim \sup_{t \rightarrow \tau_X} \int_t^{\tau_X} [\psi\{\pi(s)\}^2 / \pi(s)] d\tilde{\Lambda}(s) = 0$  (condition 4).
- (e)  $S_T(t)$  and  $\hat{S}_0(t)$  are continuous over  $[0, \tau_X]$  if  $\tau_X < \infty$ . Otherwise, define  $S_T(\infty) = \lim_{t \rightarrow \infty} \{S_T(t)\}$  (condition 5).
- (f)  $\lim_{t \rightarrow \tau_{F_0}} [1 - F_0(t)] / \pi(t) < 1$  (condition 6).

Condition 1 ensures that estimator (5) is well defined. Note that, when  $k = 0$ , condition 2 reduces to that cumulative crude hazard  $\tilde{\Lambda}(\tau_X)$  is finite, ensuring a non-zero cure proportion. Moreover, condition 2 when  $k = 1, 2$  coupled with condition 3 are technical conditions characterizing the bounds of the first and second derivatives of the copula function  $\phi$ , facilitating the consistency proof. Condition 4 ensures that a sufficiently large proportion of subjects are at risk at the end of follow-up, which facilitates the proof of asymptotic normality. Condition 5 precludes the situation that positive point masses are put on discrete time points (except on time  $\infty$ ), which is needed for cure detection. Condition 6 ensures that the tail of the observed survival times is heavier than that of the latency distribution. Hence, enough information will be available at the tail for cure estimation. Tedious computations show that the Clayton and Frank copulas would satisfy these conditions.

*A.2. Lemmas and proofs*

*Lemma 1.* Let  $\hat{S}_T(t)$  be defined in equation (5) and let  $\tau_X$  be defined in equation (9). Then  $\phi\{\hat{S}_T(t)\}$  converges to  $\phi\{S_T(t)\}$  uniformly on  $[0, \tau_X]$ . Moreover,  $\hat{S}_T(t)$  converges to  $S_T(t)$  uniformly on  $[0, \tau_X]$  and the Nelson–Aalen estimator

$$\int_0^t I\{Y(s) > 0\} \frac{dN(s)}{Y(s)}$$

converges to  $\tilde{\Lambda}(t)$  in probability uniformly on  $[0, \tau_X]$ .

*Proof.* We give the main steps here and the detailed proof can be found in Li *et al.* (2006). We first show that, for any fixed  $t_0$  such that  $\pi(t_0) > 0$ ,

$$\sup_{t \in [0, t_0]} |\phi\{\hat{S}_T(t)\} - \phi\{S_T(t)\}| \xrightarrow{\text{pr}} 0,$$

which involves using Lengart’s inequality (Fleming and Harrington, 1991). We then show that

$$\sup_{0 \leq t \leq \tau_X} |\phi\{\hat{S}_T(t)\} - \phi\{S_T(t)\}| \xrightarrow{\text{pr}} 0,$$

which requires examining the behaviour of the involved random quantities near the boundary ( $\tau_X$ ) and, hence, needs more delicate arguments. Finally, we demonstrate the uniform convergence of

$$\int_0^t I\{Y(s) > 0\} \frac{dN(s)}{Y(s)}$$

to  $\tilde{\Lambda}(t)$  on  $[0, \tau_X]$  by observing that

$$\int_0^t I\{Y(s) > 0\} \frac{dN(s)}{Y(s)} - \int_0^t d\tilde{\Lambda}(s) = \int_0^t I\{Y(s) > 0\} \frac{dM(s)}{Y(s)} - \int_0^t I\{Y(s) = 0\} d\tilde{\Lambda}(s).$$

*Lemma 2.* Let  $\hat{S}_T, X^n, \tau_{F_0}$  and  $\tau_X$  be as defined in equations (5), (6), (7) and (9) respectively. Then

$$\hat{S}_T(X^n) \xrightarrow{PF} 1 - p$$

if and only if  $\tau_{F_0} \leq \tau_X$ .

*Proof.* By the definition of  $\tau_X, X_n \rightarrow \tau_X$  almost surely. Consider

$$\begin{aligned} |\phi\{\hat{S}_T(X^n)\} - \phi\{S_T(\tau_X)\}| &\leq |\phi\{\hat{S}_T(X^n)\} - \phi\{S_T(X^n)\}| + |\phi\{S_T(X^n)\} - \phi\{S_T(\tau_X)\}| \\ &\leq \sup_{0 \leq t \leq \tau_X} |\phi\{\hat{S}_T(t)\} - \phi\{S_T(t)\}| + |\phi\{S_T(X^n)\} - \phi\{S_T(\tau_X)\}|. \end{aligned}$$

Hence, by the uniform convergence of  $\phi\{\hat{S}_T(t)\}$  and continuity of  $S_T(t)$  at  $\tau_X$ , we have  $\hat{S}_T(X^n) \xrightarrow{PF} S_T(\tau_X)$ . So  $\hat{S}_T(X^n) \xrightarrow{PF} 1 - p$  if and only if  $S_T(\tau_X) = 1 - p$ . Since  $\tau_{F_0} = \sup\{t : S_T(t) > 1 - p\}$ , it then follows that  $S_T(\tau_X) = 1 - p$  if and only if  $\tau_{F_0} \leq \tau_X$ .

*Lemma 3.* Let  $\tau_X$  and  $\tau_U$  be as defined in equations (9) and (8) respectively. When  $0 < p < 1, \tau_X = \tau_U$  under equation (3).

*Proof.* Since  $\pi(t) \leq S_U(t)$ , hence  $\{t : \pi(t) > 0\} \subset \{t : S_U(t) > 0\}$ , yielding  $\tau_X \leq \tau_U$ .

However, we can also show that  $\tau_U \leq \tau_X$ . Indeed we only need to consider the case when  $\tau_X < \infty$ . Otherwise the inequality holds trivially. Specifically, when  $\tau_X \leq \infty, \pi(\tau_X+) = 0$  and therefore  $\phi\{\pi(\tau_X+)\} = \infty$ . Under equation (3),

$$\phi\{S_T(\tau_X+)\} + \phi\{S_U(\tau_X+)\} = \phi\{\pi(\tau_X+)\},$$

and, as  $p < 1, S_T(\tau_X+) \geq S_T(\infty) = 1 - p > 0$ . So  $\phi\{S_T(\tau_X+)\} < \infty$ . Hence  $\phi\{S_U(\tau_X+)\} = \infty$ , which implies that  $S_U(\tau_X+) = 0$ . By the definition of  $\tau_U = \sup\{t : S_U(t) > 0\}$ , it follows that  $\tau_U \leq \tau_X$ .

### A.3. Sketchy proof of theorem 1

An outline of the proof is given below, whereas a complete proof is deferred to Li *et al.* (2006). Using the same argument as in Rivest and Wells (2001), up to an  $o_p(1)$  term, we have that

$$\begin{aligned} Z_n(t) &= n^{1/2} \left( -\frac{1}{n} \int_0^{t \wedge X^n} I\{Y(s) > 0\} \phi' \left\{ \frac{Y(s)}{n} \right\} dM(s) + \int_0^{t \wedge X^n} I\{Y(s) > 0\} \left[ \psi \left\{ \frac{Y(s)}{n} \right\} - \psi\{\pi(s)\} \right] d\tilde{\Lambda}(s) \right) \\ &= Z_{n,1}(t) + Z_{n,2}(t), \end{aligned} \tag{24}$$

where  $X^n$  is as defined in equation (6). Rivest and Wells (2001) showed, for any  $t_0$  such that  $\pi(t_0) > 0$ , that  $Z_n(t)$  converges weakly to  $Z(t)$  on  $D[0, t_0]$ . To show the weak convergence of  $Z_n(t)$  on  $D[0, \tau_X]$ , it is sufficient to show the tightness of  $Z_n(t)$  in a small (left) neighbourhood of  $\tau_X$  in view of theorems 13.2 and 16.8 of Billingsley (1999). Hence, it suffices to show that for any  $\varepsilon > 0$

$$\lim_{t \rightarrow \tau_X} \limsup_n \left[ P \left\{ \sup_{s \in (t, \tau_X]} |Z_n(s) - Z_n(t)| > \varepsilon \right\} \right] = 0, \tag{25}$$

which will be accomplished by using Lenglart's inequality. We next compute the covariance function for the limiting process  $Z(t)$ , which will be used in computing the asymptotic distributions of the derived test statistics in Section 4. The derivation of this covariance function is rather involved as  $Z(t)$  is not an independent increment process; the detail can also be found in Li *et al.* (2006).

**A.4. Proof of theorem 2**

Note that

$$n^{1/2}\{\phi(1 - \hat{p}) - \phi(1 - p)\} = n^{1/2}[\phi\{\hat{S}_T(X^n)\} - \phi\{S_T(X^n)\}] + n^{1/2}[\phi\{S_T(X^n)\} - \phi(1 - p)] \\ = Z_n(\tau_X) + n^{1/2}[\phi\{S_T(X^n)\} - \phi(1 - p)]$$

where  $\hat{S}_T(t)$  is defined in equation (5), and  $Z_n(\tau_X)$ , as defined in equation (11), converges weakly to  $Z^\infty$  by theorem 1.

We only need to show that

$$n^{1/2}[\phi\{S_T(X^n)\} - \phi(1 - p)] \xrightarrow{\text{pr}} 0.$$

For a fixed  $\varepsilon > 0$ , consider an increasing sequence  $a_n$  such that

$$F_0(a_n^-) \leq 1 - \frac{\varepsilon}{n^{1/2}p} \leq F_0(a_n).$$

It follows that  $a_n \rightarrow \tau_{F_0}$ , where  $\tau_{F_0}$  is the right extreme of  $F_0$ . Thus,

$$P\{n^{1/2}|p - F(X^n)| > \varepsilon\} = P\{n^{1/2}|p - p F_0(X^n)| > \varepsilon\} = P(X^n \leq a_n).$$

With regularity condition 6, it follows that  $\pi(t) \geq 1 - F_0(t)$  when  $t$  is sufficiently close to  $\tau_{F_0}$ . Consequently, when  $n$  is sufficiently large

$$P(X^n \leq a_n) = \{1 - \pi(a_n)\}^n \leq \left(1 - \frac{\varepsilon}{n^{1/2}p}\right)^n \rightarrow 0.$$

Thus,  $n^{1/2}|p - F(X^n)|$  converges to 0 in probability and so does  $n^{1/2}|\phi(1 - p) - \phi\{S_T(X^n)\}|$  to 0 by the boundedness condition on  $\phi'(\cdot)$  (the regularity condition 1). Therefore expression (13) holds, which also implies expression (14) by the Slutsky theorem. Further note that

$$\sup_{[0, \tau_X]} |\hat{F}_0(t) - F_0(t)| \leq \frac{1}{\hat{p}} \sup_{[0, \tau_X]} |\hat{S}_T(t) - S_T(t)| + \frac{1}{p\hat{p}} |\hat{p} - p|.$$

Hence the result follows as  $\hat{S}_T(t)$  converges to  $S_T(t)$  uniformly on  $[0, \tau_X]$  coupled with  $\hat{p} - p \xrightarrow{\text{pr}} 0$ .

**A.5. Proof of theorem 3**

First observe that

$$n^{1/2}\{\hat{F}_0(t \wedge X^n) - F_0(t \wedge X^n)\} = n^{1/2} \left\{ \frac{\hat{F}_T(t \wedge X^n)}{\hat{p}} - \frac{F_T(t \wedge X^n)}{p} \right\} \\ = \frac{-n^{1/2}[\phi\{\hat{S}_T(t \wedge X^n)\} - \phi\{S_T(t \wedge X^n)\}]}{p\phi'\{S_T(t \wedge X^n)\}} - \frac{1 - S_T(t \wedge X^n)}{p^2} n^{1/2}(\hat{p} - p) + o_p(1).$$

Now, since

$$n^{1/2}[\phi\{\hat{S}_T(t \wedge X^n) - \phi\{S_T(t \wedge X^n)\}] \xrightarrow{w} Z(t)$$

on  $D[0, \tau_X]$  (theorem 2) in conjunction with

$$n^{1/2}(\hat{p} - p) \xrightarrow{d} \frac{Z^\infty}{-\phi'(1 - p)}$$

and  $X^n \rightarrow \tau_X$  almost surely, it follows that

$$n^{1/2}\{\hat{F}_0(t) - F_0(t)\} \xrightarrow{w} G(t)$$

on  $D[0, \tau_X]$ , where

$$G(t) = -\frac{Z(t)}{p\phi'\{S_T(t)\}} + \frac{\{1 - S_T(t)\}Z^\infty}{p^2\phi'(1 - p)}. \tag{26}$$

**A.6. Derivation of equation (17)**

We need to apply theorem 2, and hence consider the truncated version of  $S_T$  and its estimator  $\hat{S}_T$  as defined in equation (5). Since  $X^n \rightarrow \tau_X$  almost surely, the following inequality holds with probability 1 for any  $0 < t_0 < t_1 < \tau_X$ :

$$S_T(t_1 \wedge X^n) - S_0(t_1 \wedge X^n) \geq S_T(t_1 \wedge X^n) - S_T(t_0 \wedge X^n)^{t_1/t_0} \stackrel{\text{def}}{=} \hat{\Delta}(t_0, t_1),$$

and  $\hat{\Delta}(t_0, t_1) \rightarrow \Delta(t_0, t_1)$  almost surely.

Consider

$$\hat{\Delta}(t_0, t_1) = \hat{S}_T(t_1 \wedge X^n) - \hat{S}_T(t_0 \wedge X^n)^{t_1/t_0} + S_T(t_1 \wedge X^n) - \hat{S}_T(t_1 \wedge X^n) - \{S_T(t_0 \wedge X^n)^{t_1/t_0} - \hat{S}_T(t_0 \wedge X^n)^{t_1/t_0}\}.$$

Using a Taylor series expansion, we have

$$|\hat{S}_T(t_0 \wedge X^n)^{t_1/t_0} - S_T(t_0 \wedge X^n)^{t_1/t_0}| = \frac{t_1}{t_0} \xi_n^{t_1/t_0 - 1} |\hat{S}_T(t_0 \wedge X^n) - S_T(t_0 \wedge X^n)|,$$

where  $\xi_n$  is between  $\hat{S}_T(t_0 \wedge X^n)$  and  $S_T(t_0 \wedge X^n)$ . Lemma 1 then immediately implies that  $\xi_n \xrightarrow{P} S_T(t_0)$ . Hence, with a probability going to 1,

$$|\hat{S}_T(t_1 \wedge X^n)^{t_1/t_0} - S_T(t_1 \wedge X^n)^{t_1/t_0}| \leq (1 + \varepsilon_0) \frac{t_1}{t_0} S_T(t_0)^{t_1/t_0 - 1} |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|, \tag{27}$$

where  $\varepsilon_0$  is any fixed positive number.

Also, the weak convergence of  $n^{1/2}\{\hat{S}_T(\cdot \wedge X^n) - S_T(\cdot \wedge X^n)\}$ , coupled with the continuous mapping theorem, gives

$$P(n^{1/2} \sup_t |\{\hat{S}_T(t \wedge X^n) - S_T(t \wedge X^n)\}| \leq D_\alpha) \rightarrow 1 - \alpha, \tag{28}$$

where  $D_\alpha$  is the upper  $100\alpha$ -percentile of  $\sup_t |Z(t)/\phi'\{S_T(t)\}|$  (on the basis of theorem 1). Then, we have the following asymptotic lower confidence limit for  $\Delta(t_0, t_1)$ , and hence, for  $\Delta = \sup_{0 < t < \tau_X} \{S_T(t) - S_0(t)\}$  (which is larger than  $\hat{\Delta}(t_0, t_1)$  almost surely). More specifically, some basic probabilistic arguments lead to equation (17) when  $n$  is sufficiently large.

**A.7. Derivation of equation (19)**

If the hazard

$$\lambda_0(t) = -\frac{d}{dt} \log\{S_0(t)\}$$

is a non-increasing function of  $t$ , similar arguments lead to  $S_0(t_0) \leq S_T(t_1)^{t_0/t_1}$  for any  $t_0 \leq t_1$  and that the upper bound for  $p$  is

$$\tilde{p} \stackrel{\text{def}}{=} \min_{0 < t_0 < t_1 < \tau_X} \left\{ \frac{1 - S_T(t_0)}{1 - S_T(t_1)^{t_0/t_1}} \right\}$$

which can be consistently estimated by the statistic

$$\hat{p} = \min \left[ \min_{0 < t_0 < t_1 < X^n} \left\{ \frac{1 - \hat{S}_T(t_0)}{1 - \hat{S}_T(t_1)^{t_0/t_1}}, 1 \right\} \right]. \tag{29}$$

In view of  $S_T(t_0) - S_0(t_0) \geq S_T(t_0) - S_T(t_1)^{t_0/t_1}$ , we redefine  $\Delta(t_0, t_1)$  such that

$$\Delta(t_0, t_1) \stackrel{\text{def}}{=} S_T(t_0) - S_T(t_1)^{t_0/t_1}$$

and redefine

$$\hat{\Delta}(t_0, t_1) \stackrel{\text{def}}{=} S_T(t_0 \wedge X^n) - S_T(t_1 \wedge X^n)^{t_0/t_1},$$

which can be written

$$\hat{\Delta}(t_0, t_1) = \hat{S}_T(t_0 \wedge X^n) - \hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} + S_T(t_0 \wedge X^n) - \hat{S}_T(t_0 \wedge X^n) - \{S_T(t_1 \wedge X^n)^{t_0/t_1} - \hat{S}_T(t_1 \wedge X^n)^{t_0/t_1}\}.$$

Using the Taylor series expansion, we have

$$S_T(t_1 \wedge X^n)^{t_0/t_1} - \hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} = \frac{t_0}{t_1} \xi_n^{t_0/t_1-1} \{S_T(t_1 \wedge X^n) - \hat{S}_T(t_1 \wedge X^n)\},$$

where  $\xi_n$  is between  $\hat{S}_T(t_1 \wedge X^n)$  and  $S_T(t_1 \wedge X^n)$ . Lemma 1 then directly implies that  $\xi_n \xrightarrow{P} S_T(t_1)$ . Hence, with a probability going to 1,

$$|\hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} - S_T(t_1 \wedge X^n)^{t_0/t_1}| < (1 + \varepsilon_0) \frac{t_0}{t_1} S_T(t_1)^{t_0/t_1-1} |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|, \tag{30}$$

where  $\varepsilon_0$  is any fixed positive number.

If  $S_T(t_1)$  is close to 0,  $S_T(t_1)^{t_0/t_1-1}$  may not be well bounded. However, note that, for any constants  $0 \leq x, y, a \leq 1$ , it is easy to show that  $|x^a - y^a| \leq |x - y|^a$ . It follows that

$$|\hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} - S_T(t_1 \wedge X^n)^{t_0/t_1}| \leq |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|^{t_0/t_1}, \tag{31}$$

as  $0 < t_0/t_1 < 1$ . Therefore, combining inequalities (30) and (31) gives

$$|\hat{S}_T(t_1 \wedge X^n)^{t_0/t_1} - S_T(t_1 \wedge X^n)^{t_0/t_1}| \leq \min \left\{ (1 + \varepsilon_0) \frac{t_0}{t_1} S_T(t_1)^{t_0/t_1-1} O_n, O_n^{t_0/t_1} \right\},$$

where  $O_n = \sup |\hat{S}_T(t \wedge X^n) - S_T(t \wedge X^n)|$ . Hence, using expression (28) and letting  $\varepsilon_0 \rightarrow 0$ , we obtain equation (19).

### A.8. Proof of theorem 4

Denote by  $X^n = X^{n_1} \wedge X^{n_2}$  and define residual processes

$$\varepsilon_{n_i}(t) = n_i^{1/2} \{ \hat{F}_{0,i}(t \wedge X^n) - F_0(t \wedge X^n) \}.$$

The sample paths of stochastic processes  $\varepsilon_{n_i}$  reside in the Skorohod space  $D_R[0, \tau]$ . Then it follows by theorem 2 that

$$\varepsilon_{n_i}(t) \xrightarrow{w} G_i(t).$$

Hence by the continuous mapping theorem, when  $r < \infty$

$$\begin{aligned} W_n^r &= \int_0^\tau \left| n^{1/2} \{ \hat{F}_{0,1}(t \wedge X^n) - \hat{F}_{0,2}(t \wedge X^n) \} \right|^r d\hat{F}_{\text{pool}}^*(t) \\ &\sim \int_0^\tau \left| \frac{1}{\sqrt{\gamma}} \varepsilon_{n_1}(t) - \frac{1}{\sqrt{(1-\gamma)}} \varepsilon_{n_2}(t) \right|^r d\hat{F}_{\text{pool}}^*(t) \\ &\Rightarrow \int_0^\tau |\tilde{G}(t)|^r dF_0(t) \\ &= \int_0^\infty |\tilde{G}(t)|^r dF_0(t), \end{aligned}$$

where ‘ $\sim$ ’ represents pointwise equivalence asymptotically and ‘ $\Rightarrow$ ’ denotes weak convergence.

When  $r = \infty$ ,

$$\begin{aligned} W_n &= \sup_{t \in [0, X^{n_1} \vee X^{n_2}]} |n^{1/2} \{ \hat{F}_{0,1}(t \wedge X^n) - \hat{F}_{0,2}(t \wedge X^n) \}| \\ &= \sup_{t \in [0, \tau]} |n^{1/2} \{ \hat{F}_{0,1}(t \wedge X^n) - \hat{F}_{0,2}(t \wedge X^n) \}| \\ &\sim \sup_{t \in [0, \tau]} \left| \frac{1}{\sqrt{\gamma}} \varepsilon_{n_1}(t) - \frac{1}{\sqrt{(1-\gamma)}} \varepsilon_{n_2}(t) \right| \\ &\Rightarrow \sup_{t \in [0, \tau]} |\tilde{G}(t)|. \end{aligned}$$

## References

- Bartholomew, D. J. (1957) A problem in life testing. *J. Am. Statist. Ass.*, **52**, 350–355.
- Berkson, J. and Gage, R. P. (1952) Survival curves for cancer patients following treatment. *J. Am. Statist. Ass.*, **47**, 501–515.
- Billingsley, P. (1999) *Convergence of Probability Measures*. New York: Wiley.
- Boag, J. W. (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy (with discussion). *J. R. Statist. Soc. B*, **11**, 15–53.
- Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*. New York: Wiley.
- Genest, C. and MacKay, R. J. (1986) Archimedean copulas and bivariate families with continuous marginals. *Can. J. Statist.*, **14**, 145–159.
- Hogan, J. and Laird, N. (1997) Model-based approach to analysing incomplete longitudinal and failure time data. *Statist. Med.*, **16**, 259–272.
- Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Klebanov, L. B. and Yakovlev, A. Y. (2005) A new approach to testing for sufficient follow up in cure-rate analysis. *Technical Report*. Department of Biostatistics and Computational Biology, University of Rochester, Rochester.
- Kuk, A. Y. C. and Chen, C. (1992) A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531–541.
- Li, Y. and Feng, J. (2005) A nonparametric comparison of conditional distributions with non-negligible cure fractions. *Lifetime Data Anal.*, **11**, 367–387.
- Li, Y., Tiwari, R. and Guha, S. (2006) Technical proofs for “Mixture cure survival models with dependent censoring”. *Technical Report*. Harvard School of Public Health and Dana–Farber Cancer Institute, Boston.
- Little, R. J. A. (1995) Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Ass.*, **90**, 1112–1121.
- Maller, R. A. and Zhou, S. (1992) Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, 731–739.
- Maller, R. A. and Zhou, S. (1994) Testing for sufficient follow-up and outliers in survival data. *J. Am. Statist. Ass.*, **89**, 1499–1506.
- Maller, R. A. and Zhou, S. (1996) *Survival Analysis with Long-term Survivors*. New York: Wiley.
- Moeschberger, M. L. and Klein, J. P. (1995) Statistical methods for dependent competing risks. *Lifetime Data Anal.*, **1**, 195–204.
- Peng, Y. and Dear, K. B. (2000) A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**, 237–243.
- Rivest, L. P. and Wells, M. T. (2001) A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. Multiv. Anal.*, **79**, 138–155.
- Scharfstein, D. O. and Robins, J. M. (2002) Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*, **89**, 617–634.
- Scharfstein, D. O., Robins, J. M., Eddings, W. and Rotnitzky, A. (2001) Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics*, **57**, 404–413.
- Siannis, F., Copas, J. and Lu, G. (2005) Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, **6**, 77–91.
- Sy, J. P. and Taylor, J. M. (2000) Estimation in a Cox proportional hazards cure model. *Biometrics*, **56**, 227–236.
- Tsiatis, A. A. (1978) An example of nonidentifiability in competing risks. *Scand. Act. J.*, **78**, 235–239.
- Zheng, M. and Klein, J. P. (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**, 127–138.