

Mixture Cure Survival Models with Dependent Censoring

Yi Li* Ram C. Tiwari†
Subharup Guha‡

*Harvard University and Dana Farber Cancer Institute, yili@jimmy.harvard.edu

†tiwarir@mail.nih.gov

‡Harvard University, sguha@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper26>

Copyright ©2005 by the authors.

Mixture Cure Survival Models with Dependent Censoring

Yi Li, Ram C. Tiwari, and Subharup Guha

Abstract

A number of authors have studied the mixture survival model to analyze survival data with nonnegligible cure fractions. A key assumption made by these authors is the independence between the survival time and the censoring time. To our knowledge, no one has studied the mixture cure model in the presence of dependent censoring. To account for such dependence, we propose a more general cure model which allows for dependent censoring. In particular, we derive the cure models from the perspective of competing risks and model the dependence between the censoring time and the survival time 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 distribution in the presence of dependent censoring when a proportion of patients is deemed cured. Large sample results using the martingale theory are obtained. We applied the proposed methodologies to the SEER prostate cancer data.

Mixture Cure Survival Models with Dependent Censoring

BY YI LI

*Department of Biostatistics
Harvard School of Public Health and Dana-Farber Cancer Institute
Boston, MA 02115, U.S.A.*

Ram C. Tiwari

*Statistical Research and Applications Branch
National Cancer Institute
Bethesda, MD 20892, U.S.A.*

Subharup Guha

*Department of Biostatistics
Department of Biostatistics Harvard School of Public Health
Boston, MA 02115, U.S.A.*

Abstract

A number of authors have studied the mixture survival model to analyze survival data with nonnegligible cure fractions. A key assumption made by these authors is the independence between the survival time and the censoring time. To our knowledge, no one has studied the mixture cure model in the presence of dependent censoring. To account for such dependence, we propose a more general cure model which allows for dependent censoring. In particular we derive the cure models from the perspective of competing risks and model the dependence between the censoring time and the survival time 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 using the martingale theory are obtained. We applied the proposed methodologies to the SEER prostate cancer data.

KEY WORDS: Cure Model; Dependent Censoring; Martingale Processes; Archimedean Copula Model; Latency Distribution; Consistency; Asymptotic Normality; Testing Cure Fraction.

RUNNING TITLE: Cure Modeling with Dependent Censoring.

1 Introduction

Survival models incorporating a cure fraction, often termed as cure rate models, have emerged as a powerful statistical tool for analyzing cancer studies. Applications have been found in modeling time-to-event data for a variety of cancers, including prostate cancer, breast cancer, non-Hodgkins lymphoma, leukemia, melanoma, and head and neck cancer, where for these diseases, a significant proportion of patients are “cured” after therapies. By cure it is meant that an individual will have little or no risk of experiencing the event of interest, e.g. death from breast cancer. Recent years have seen a spurt in statistical literature that deals with survival data with a nonnegligible cure fraction; see e.g. Kuk and Chen (1992), Maller and Zhou (1996), Peng and Dear (2000), Sy and Taylor (2000), among others. Most of the current work stems from the mixture model originally proposed by Boag (1949) and Berkson and Gage (1952), which is formulated as follows.

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

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

$$F_T(t) = pF_0(t), \tag{1}$$

or, equivalently,

$$S_T(t) = 1 - p + pS_0(t), \tag{2}$$

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 deemed susceptible (non-cured) and non-susceptible (cured). Note that $(1-p)$ corresponds to the fraction of cured, that is, the point mass that T puts on ∞ and $F_0(t)$ is the distribution for the non-cured patients, often termed as the *latency* distribution.

A key assumption made by the current literature is the independence between the survival time T and the censoring time U . To our knowledge, none has studied the mixture cure model in the presence of dependent censoring, which are commonly observed in biological studies. For example, in the prostate cancer data set of the NIH Surveillance Epidemiology and End Results (SEER) program, a proportion of patients diagnosed with this type of cancer died from heart/cardiovascular diseases. A recent study (see <http://www.thewbalchannel.com/healtharchive/4161401/detail.html>) has revealed that the prostate cancer and the cardiovascular disease may be linked through a common risk factor, high cholesterol. Therefore, it would seem implausible to assume independence between the main endpoint (e.g. deaths from prostate cancer) and the censoring causes (e.g. deaths from heart diseases). In this paper, we propose a more general cure model which allows for dependent censoring. In particular we derive the mixture cure model from perspectives of competing risks and model the dependence between the censoring time and the survival time using a class of Archimedean copula models. Within this framework, we focus on the cure detection, and the comparison of latency distributions in the presence of dependent censoring when a proportion of patients is deemed cured.

The rest of the article is structured as follows. In Section 2 we introduce a mixture cure model with the dependence of the censoring and survival times modeled 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 show the consistency of the estimator, and in Section 4 we test for sufficient follow up, a sufficient condition for consistently estimating the cure fraction. We prove the asymptotic normality in Section 5 and conduct the hypothesis testing in Section 6. We conclude this article with discussion and future work in Section 7. We defer all the proofs to the Appendix.

2 Mixture Cure Model with Archimedean Dependence

Throughout we assume that the survival time T follows the mixture model (1), or, equivalently, (2). Denote the joint survival of the failure and censoring times by $C(t, u) = P(T > t, U > u)$.

An Archimedean copula model links the marginal survivals of T and U through

$$C(t, u) = \phi^{-1}[\phi\{S_T(t)\} + \phi\{S_U(u)\}], \quad (3)$$

where a nonincreasing function $\phi : [0, 1] \rightarrow [0, \infty]$ is specified such that $\phi(1) = 0$ and $\phi(0) = \infty$. Examples 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 $\phi(t) = -\log((1 - \exp(-at))/(1 - \exp(-a)))$ (for $a > 0$). We adopt the Archimedean copula formulation to emphasize the functional independence of the parameterizations of the marginal distribution functions, governed by S_T and S_U , and the dependence structure, governed by a class of copula functions ϕ . This formulation also facilitates a derivation of the estimator for S_T , our main interest.

Suppose that we observe n i.i.d 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 heuristically discusses an estimator based on (3), whose large sample properties will be considered in the next section. Denote by \hat{S}_T and \hat{S}_U the estimates for S_T and S_U respectively, which are right continuous and piecewise constant functions with jumps only occurring 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 (3), at each observed time points $X_i, i=1, \dots, n$,

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

Assume that $P(T = U) = 0$ (i.e. the 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

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

Applying (4) recursively, we write the estimate \hat{S}_T in the form of counting process as follows

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

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

It is obvious that $\hat{S}_T(t)$ is nonincreasing 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 nonzero when the largest value among all the observed times (X_1, \dots, X_n) , denoted by $X^n = \sup_t\{t : Y(t) > 0\}$, is censored. Under some regularity conditions, we will explore using this constant to estimate the cure fraction and to study the asymptotic properties.

Before proceeding further, introduce the right extremes $\tau_{F_0} = \sup_t\{t : F_0(t) < 1\}$, $\tau_U = \sup_t\{t : F_U(t) < 1\}$ and $\tau_X = \sup_t\{t : \pi(t) > 0\}$. From (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 .

Our main results are as follows: under some regularity conditions (listed in Appendix A.0)

- The cure fraction can be consistently estimated based on \hat{S}_T . Specifically, $\hat{S}_T(X^n) \xrightarrow{pr} 1 - p$, or, equivalently, $\hat{F}_T(X^n) \xrightarrow{pr} p$.
- The estimate of the cure fraction is asymptotically normally distributed. That is, $\sqrt{n}\{\hat{F}_T(X^n) - p\}$ or equivalently, $\sqrt{n}\{\hat{S}_T(X^n) - (1 - p)\}$ converges in distribution to a mean zero normal random variable with a finite variance.
- The estimate of the latency distribution is uniformly consistent and asymptotically normal. More specifically, $\sup_{t \in [0, \tau_X]} \left| \frac{\hat{F}_T(t)}{\hat{F}_T(X^n)} - F_0(t) \right| \xrightarrow{pr} 0$, and $\sqrt{n} \left\{ \frac{\hat{F}_T(t \wedge X^n)}{\hat{F}_T(X^n)} - F_0(t \wedge X^n) \right\}$ converges weakly to a tight Gaussian process on the Skorohod space $D[0, \tau_X]$.

We will further consider cure detection and propose a class of tests for testing the equality of the latency distribution F_0 in a two-sample comparison setting along the lines of Li and Feng (2005).

3 Consistency

Introduce the crude hazard function defined by

$$d\tilde{\Lambda}(t) = \tilde{\lambda}(t)dt = P(t < T \leq t + dt | T > t, U > t),$$

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). Note that when T and U are dependent the crude hazard $\tilde{\lambda}(t)$ may not be equal to the conventional hazard defined by $\lambda(t) = \frac{1}{dt}P(t < T \leq t + dt | T > t)$. For example, consider a Clayton joint survival $C(t, u) = (e^{a\lambda_1 t} + e^{a\lambda_2 u} - 1)^{-\frac{1}{a}}$, which corresponds to the Archimedean copula model with $\phi(t) = (t^{-a} - 1)/a$ and $S_T(t) = e^{-\lambda_1 t}$ and $S_U(u) = e^{-\lambda_2 u}$, where $a \geq 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$. It follows that the crude hazard

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

which differs from the conventional hazard λ_1 when $a \neq 0$. Other counter-examples can be found in Example 1.3.1 of Fleming and Harrington (1991).

Under the regularity conditions ((c.1)-(c.5) listed in Appendix A.0) on $S_T(t)$ (or $F_T(t)$), $\pi(t)$ and the copula function ϕ , the following proposition states that $\hat{S}_T(t)$ is a uniformly consistent estimator to $S_T(t)$.

Proposition 1 $\phi(\hat{S}_T(t))$ 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 $[0, \tau_X]$.

Note that in this proposition we allow $\pi(\tau_X) = 0$. That is, 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.

In addition, a byproduct of this proposition is the feasibility of consistently estimating an important parameter $P(T \leq U)$, which is the probability of observing death from a particular cancer prior to censoring. Indeed,

$$\begin{aligned} P(T \leq U) &= \int_0^\infty P(T \in (t, t + dt], U > t) \\ &= \int_0^\infty \pi(t)P(T \in (t, t + dt], U > t | T > t, U > t) \\ &= \int_0^\infty \pi(t)d\tilde{\Lambda}(t). \end{aligned}$$

Proposition 1 indicates that the Nelson-Aalen estimator $\int_0^t I(Y(u) > 0)dN/Y$ consistently estimates $\tilde{\Lambda}(t)$. Hence a natural ‘plug-in’ estimator for $P(T \leq U) \stackrel{def}{=} P$ is

$$\hat{P} = \int_0^\infty \frac{Y(u)}{n} I(Y(u) > 0) \frac{dN}{Y} = \int_0^\infty I(Y(u) > 0) \frac{1}{n} dN \equiv \int_0^\infty \frac{1}{n} dN = \frac{1}{n} \sum_i^n \delta_i.$$

A simple martingale argument shows that \hat{P} consistently estimates P . Since δ_i are independent Bernoulli random variables, the central limit theorem leads to

$$\sqrt{n}(\hat{P} - P) \xrightarrow{d} N(0, P(1 - P)),$$

and one simply approximates the variance of \hat{P} by $\hat{P}(1 - \hat{P})/n$.

It is natural to use the plateau of the estimated survival curve $\hat{S}_T(X^n)$ to estimate the cure fraction $(1 - p)$. The following two propositions indicate this approach is proper if and only if the support of the latency distribution is covered by that of the censoring distribution.

Proposition 2 $\hat{S}_T(X^n) \xrightarrow{pr} (1 - p)$ if and only if $\tau_{F_0} \leq \tau_X$.

As τ_X characterizes the support of $X = T \wedge U$, we can further show that $\tau_X = \tau_U$ in the presence of cure under model (3). That is, the supports of X and U coincide under an Archimedean model when the cure fraction is non zero.

Proposition 3 When $0 < p < 1$, $\tau_X = \tau_U$ under (3).

4 Testing Sufficient Follow-up

Propositions 2 and 3 indicate that when $p < 1$, it will be consistently estimated by $1 - \hat{S}_T(X^n)$ if $\tau_{F_0} < \tau_U$, that is, 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 $\tau_{F_0} < \tau_U$. Thus even in the absence of cure fraction, (5) provides a consistent estimate for p and we will not be misled by using (5) as long as $\tau_{F_0} < \tau_U$, reflecting a sufficient follow-up.

Therefore it is crucial to test the hypothesis $\tau_{F_0} < \tau_U$ for consistently estimating $(1 - p)$ by using (5). Applying the Borel-Cantelli lemma, we can show X^{n*} , the largest observed failure time, $\rightarrow \tau_{F_0} \wedge \tau_U$ almost surely, while X^n , the largest observed time (which may be censored), is arbitrarily close to τ_U , that is, $X^n \rightarrow \tau_U$ almost surely. Hence, if $\tau_{F_0} \wedge \tau_U < \infty$, $X^n - X^{n*} \rightarrow \tau_U - \tau_{F_0}$ almost surely if $\tau_{F_0} < \tau_U$, and converges to 0 almost surely if $\tau_{F_0} \geq \tau_U$. Thus a large value of $X^n - X^{n*}$ gives evidence to $H_a : \tau_{F_0} < \tau_U$ while a small value of $X^n - X^{n*}$ points to $H_0 : \tau_{F_0} \geq \tau_U$. Based on $X^n - X^{n*}$ and following Maller and Zhou (1992, 1994) we consider the test statistic

$$\alpha_n = (1 - N_n/n)^n$$

where N_n is the number of failures observed in $[2X^{n*} - X^n, X^{n*}]$. One accepts H_a when α_n is sufficiently small, e.g. $\alpha_n < 0.05$ (or N_n is sufficiently large), while accepting H_0 when α_n is sufficiently large, e.g. $\alpha_n > 0.05$ (or N_n is sufficiently small). The heuristics, along with a detailed derivation, of this test when T and U are independent is given in Maller and Zhou (1994).

Denote by $\tau^* = \tau_{F_0} \wedge \tau_U$ and define an increasing function $h(a) = \int_{\tau^*-a}^{\tau^*} \pi(t) d\tilde{\Lambda}(t)$. We impose more regularity conditions on function h . Specifically, we suppose that there exists a small $\epsilon_0 > 0$ (if $\tau_{F_0} < \tau_U$, we require $\epsilon_0 < \tau_U - \tau_{F_0}$) such that

(d.1) $h(a)$ is continuous on $[0, \epsilon_0]$.

(d.2) (dominated variation) there exists an $M > 0$ such that $0 < h(2a) < Mh(a)$ for any $a \in$

$(0, \epsilon_0]$, or, equivalently,

$$\limsup_{a \rightarrow 0^+} \frac{h(\lambda a)}{h(a)} < \infty \quad (6)$$

for all $\lambda > 0$ (see, e.g. Bingham et al. (1987)).

We first comment on the condition (d.2), which characterizes the behavior of h near 0, and show that it holds under general circumstances. In fact, when $\tau_{F_0} < \tau_U$, we have $\pi(\tau^*) = \pi(\tau_{F_0} \wedge \tau_U) > 0$. Hence, (6) reduces to

$$\limsup_{a \rightarrow 0^+} \frac{\tilde{\Lambda}(\tau^*) - \tilde{\Lambda}(\tau^* - \lambda a)}{\tilde{\Lambda}(\tau^*) - \tilde{\Lambda}(\tau^* - a)} < \infty,$$

which would be true, if we assume $\tilde{\Lambda}^{(k)}(\tau^* -) \neq 0$ for some positive integer k , by a Taylor expansion. When $\tau_{F_0} \geq \tau_U$, we can apply L'Hopital's rule on (6). Then (6) reduces to

$$\limsup_{a \rightarrow 0^+} \frac{\pi(\tau^* - \lambda a) \tilde{\Lambda}'(\tau^* - \lambda a)}{\pi(\tau^* - a) \tilde{\Lambda}'(\tau^* - a)} < \infty,$$

which again holds widely if both $\pi(t)$ and $\tilde{\Lambda}$ have finite nonzero derivatives of some order at τ^{*-} . Hence, (6) essentially requires that $\pi(t)$ and $\tilde{\Lambda}$ have 'mild' changes near τ^* and thus is expected to hold for most commonly assumed distributions.

We now show such a test is consistent when T and U are dependent through model (3) and under the regularity conditions (d.1) and (d.2). The proof is along the line of Maller and Zhou (1994) and can be found in Appendix A.4.

Proposition 4 *That $\alpha_n \rightarrow 0$ in probability if and only if $\tau_{F_0} < \tau_U$.*

One would expect that the type of test developed above in the spirit of Maller and Zhou (1994) would display a monotonic behavior. That is, the longer the duration of study, the more likely sufficient it is for making inferences about the cure rates. Klebanov and Yakovlev (2005), however, have shown that with the finite samples Maller and Zhou's test may behave unstably and non-monotonically even when the duration increases. They were less concerned with estimating the cure proportion consistently and proposed to forgo the test of $\tau_{F_0} < \tau_U$. Instead, they focused solely on testing the existence of cure fraction, which, to us, is of interest in its own merit. We will pursue this idea in the presence of dependent censoring in Section 6.

5 Asymptotic Normality

So far we have established the consistency of the estimator when the support of the latency distribution is fully covered by the censoring. To avoid technicality, we assume that $\tau_{F_0} < \tau_U$ in the ensuing developments and show that the proposed estimator for cure fraction is also asymptotically normal.

Define the stopped process

$$Z_n(t) = \sqrt{n}\{\phi(\hat{S}_T(t \wedge X^n)) - \phi(S_T(t \wedge X^n))\}. \quad (7)$$

and the covariance function

$$\begin{aligned} \mathcal{C}(t_1, t_2) &= \int_0^{t_1 \wedge t_2} \pi(s)(\phi'(\pi(s)))^2 d\tilde{\Lambda}(s) + 2 \int_0^{t_1 \wedge t_2} \int_0^s \pi(s)(1 - \pi(u))\psi'(\pi(u))\psi'(\pi(s))d\tilde{\Lambda}(u)d\tilde{\Lambda}(s) \\ &+ 2 \int_0^{t_1 \wedge t_2} \int_0^s \phi'(\pi(u))\pi(s)\psi'(\pi(s))d\tilde{\Lambda}(u)d\tilde{\Lambda}(s) \\ &+ \int_{t_1 \wedge t_2}^{t_1 \vee t_2} \pi(s)\psi'(\pi(s))d\tilde{\Lambda}(s) \int_0^{t_1 \wedge t_2} [\{1 - \pi(u)\}\psi'\{\pi(u)\} + \phi'\{\pi(u)\}] d\tilde{\Lambda}(u) \end{aligned} \quad (8)$$

for $0 \leq t_1, t_2, < \tau_X$, where $\psi(s) \stackrel{def}{=} -s\phi'(s)$. From Proposition 1, this covariance function can be consistently estimated by replacing $\pi(s)$ and $d\tilde{\Lambda}(u)$ with their empirical counterparts, $\hat{\pi}(s)$ and $I(Y(u) > 0)dN(u)/Y(u)$ respectively. Denote the variance function $v_0(t) = \mathcal{C}(t, t)$, which coincides with the variance function obtained by Rivest and Wells (note there are two typographic errors in their formula). Assume that $\lim_{t \rightarrow \tau_X} v_0(t) = v_0^\infty < \infty$ and $\mathcal{C}^\infty(t) \stackrel{def}{=} \lim_{v \rightarrow \tau_X} \mathcal{C}(v, t) < \infty$ for every $t \in [0, \tau_X]$.

Proposition 5 $Z_n(t)$ 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 the covariance function $\mathcal{C}(t_1, t_2)$ and Z^∞ is a normal random variable with the variance v_0^∞ and $cov\{Z^\infty, Z(t)\} = \mathcal{C}^\infty(t)$.

Denote by $\hat{p} = 1 - \hat{S}_T(X^n)$. We are now ready to show the asymptotic normality of \hat{p} , the estimator for the cure fraction.

Proposition 6 Assume that $0 < p < 1$ and

$$\lim_{t \rightarrow \tau_{F_0}} \frac{1 - F_0(t)}{\pi(t)} < 1. \quad (9)$$

Then we have

$$\sqrt{n}\{\phi(1 - \hat{p}) - \phi(1 - p)\} \xrightarrow{d} Z^\infty, \quad (10)$$

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

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} \frac{Z^\infty}{-\phi'(1 - p)}. \quad (11)$$

Once we have identified the cure proportion we shall be able to compute the estimate for the latency distribution as follows

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

The following two propositions concerns the large sample results for this estimator.

Proposition 7 For $0 < p < 1$,

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

Proposition 8 Let $\tau_X = \sup\{t : \pi(t) > 0\}$. Then

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

on a Skorohod space $D[0, \tau_X]$, where $G(t) = -\frac{Z(t)}{p\phi'(S_T(t))} + \frac{Z^\infty\{1 - S_T(t)\}}{p^2\phi'(1-p)}$ and $Z(t)$ and Z^∞ are defined as in Proposition 5.

Note that if we replace the largest observed time X^n by the largest observed failure time X^{n*} , in the estimator $(1 - \hat{p}) = \hat{S}_T(X^n)$, all the large sample results hold. This follows as $P\{\hat{S}_T(X^n) = \hat{S}_T(X^{n*})\} = 1$ by the definition of \hat{S}_T .

6 Hypothesis Testing

6.1 Testing the Existence of Cure Fraction

A natural question arising from cure modeling is whether the cure fraction exists. Hence, testing $p < 1$ is of substantial interest. In the following 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 the underlying hazard for non-cured patients

is a monotone function of time, a plausible assumption in most biological studies, as opposed to the restrictive non-decreasing hazard assumption 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 a 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 α level. Klebanov and Yakovlev 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 $\lambda_0(t) = -d/dt \log S_0(t)$ is a non-decreasing function in t , implying that $-\log S_0(t)/t$ is a non-decreasing function. Hence, for any $t_1 \geq t_0 > 0$,

$$\frac{-\log S_0(t_1)}{t_1} \geq \frac{-\log S_0(t_0)}{t_0}$$

or,

$$S_0(t_1) \leq (S_0(t_0))^{t_1/t_0} \leq (S_T(t_0))^{t_1/t_0} \tag{12}$$

and from (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 is arbitrary,

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

Because of the uniform consistency of \hat{S}_T (Proposition 1) and the almost sure convergence of X^n to τ_X , \tilde{p} can be consistently estimated by the statistic

$$\hat{\tilde{p}} = \min\left\{ \min_{0 < t_0 < t_1 < X^n} \frac{1 - \hat{S}_T(t_1)}{1 - \{\hat{S}_T(t_0)\}^{t_1/t_0}}, 1 \right\}. \tag{13}$$

From (12), we have

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

for any fixed $0 < t_0 < t_1$. Our goal is to construct an asymptotic $(1 - \alpha)$ confidence interval for $\Delta(t_0, t_1)$ based on Proposition 5 and leave aside the question of choosing t_0, t_1 for now.

To apply proposition 5, we consider the truncated version of S_T and its estimator \hat{S}_T . Since $X^n \rightarrow \tau_X$ almost surely, the following inequality holds with probability 1 for any $0 < t_0 < t_1 < \tau_X$

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

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

Consider

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

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

$$|(\hat{S}_T(t_1 \wedge X^n))^{t_1/t_0} - (S_T(t_1 \wedge X^n))^{t_1/t_0}| \leq (1 + \epsilon_0) \frac{t_1}{t_0} S_T(t_0)^{t_1/t_0 - 1} |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|, \quad (14)$$

where ϵ_0 is any fixed positive number.

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

$$P(\sqrt{n} \sup_t |(\hat{S}_T(t \wedge X^n) - S_T(t \wedge X^n))| \leq D_\alpha) \rightarrow 1 - \alpha, \quad (15)$$

where D_α is the upper $\alpha \times 100$ percentile of $\sup_t |Z(t)/\phi'(S_T(t))|$ (based on Proposition 5). 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 (when n is sufficiently large)

$$P(\Delta \geq \mathcal{L}_n) = P(\hat{\Delta}(t_0, t_1) \geq \mathcal{L}_n) \geq 1 - \alpha, \quad (16)$$

where

$$\mathcal{L}_n = \hat{S}_T(t_1 \wedge X^n) - (\hat{S}_T(t_0 \wedge X^n))^{t_1/t_0} - (1 + \frac{t_1}{t_0}(1 + \epsilon_0)(S_T(t_0))^{t_1/t_0-1})D_{\alpha/2}/\sqrt{n} \quad (17)$$

Since (16) holds true for any $\epsilon_0 > 0$, we may let $\epsilon_0 \rightarrow 0$ in (17). In practice, the lower bound would be obtained by replacing the unknown $S_T(t_0)$ in (17) with its consistent estimate $\hat{S}_T(t_0 \wedge X^n)$.

If the hazard $\lambda_0(t) = -d/dt \log S_0(t)$ 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

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

which can be consistently estimated by the statistic

$$\hat{\tilde{p}} = \min\left\{\min_{0 < t_0 < t_1 < X^n} \frac{1 - \hat{S}_T(t_0)}{1 - \{\hat{S}_T(t_1)\}^{t_0/t_1}}, 1\right\}. \quad (18)$$

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

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

and redefine

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

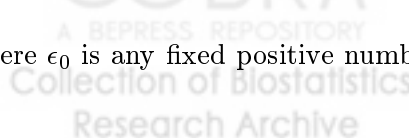
which can be written

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

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

$$|(\hat{S}_T(t_1 \wedge X^n))^{t_0/t_1} - (S_T(t_1 \wedge X^n))^{t_0/t_1}| < (1 + \epsilon_0) \frac{t_0}{t_1} S_T(t_1)^{t_0/t_1-1} |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|, \quad (19)$$

where ϵ_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, notice, 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

$$|(\hat{S}_T(t_1 \wedge X^n))^{t_0/t_1} - (S_T(t_1 \wedge X^n))^{t_0/t_1}| \leq |\hat{S}_T(t_1 \wedge X^n) - S_T(t_1 \wedge X^n)|^{t_0/t_1}, \quad (20)$$

as $0 < t_0/t_1 < 1$. Therefore, combining (19) and (20) gives

$$|(\hat{S}_T(t_1 \wedge X^n))^{t_0/t_1} - (S_T(t_1 \wedge X^n))^{t_0/t_1}| \leq \min\{(1 + \epsilon_0) \frac{t_0}{t_1} S_T(t_1)^{t_0/t_1-1} O_n, O_n^{t_0/t_1}\},$$

where $O_n = \sup |\hat{S}_T(t \wedge X^n) - S_T(t \wedge X^n)|$.

Hence, using (15) and let $\epsilon_0 \rightarrow 0$, we obtain that

$$P(\Delta \geq \mathcal{L}_n) \geq P(\hat{\Delta}(t_0, t_1) \geq \mathcal{L}_n) \geq 1 - \alpha \quad (21)$$

where

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

If the lower end of the $(1 - \alpha)$ confidence interval in (16) or (21) (depending on whether $\lambda_0(\cdot)$ is non-decreasing or non-increasing) is greater than 0, then H_0 would be rejected at a significant level of less than α . To increase the power, the choice of t_0 and t_1 can be data driven. In particular, they can be chosen based on (13) or (18) (again depending on the monotonicity of $\lambda_0(\cdot)$) to minimize the upper bound of p . Indeed, that t_0 and t_1 in (16) or (21) are chosen by minizing the lower bound of p (c.f. (13) or (18)) does not affect the probabilistic arguments leading to (16) or (21) because the latter limit is based on the Kolmogorov distance $\sqrt{n} \sup |\hat{S}_T(\cdot \wedge X^n) - S_T(\cdot \wedge X^n)|$, which is uniformly valid for all times t_1, t_0 . Thus, the data driven t_1, t_0 will allow us to increase power while maintaining the proper significant level.

6.2 Comparisons of Cure Fractions and Latency Distributions

If the presense of cure fraction is verified, it would also be of interest to compare the cure fractions and study the latency distributions, for example, when evaluating the efficacy of treatments. We consider below a two-treatment comparison scenario and adopt the notation used in the general cure model, except that we use an additional subscript i to indicate the treatments. 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 the control arm and $i = 2$ to the experimental arm, and j refers to the j -th patient in his respective treatment arm. Let $n = n_1 + n_2$. We assume that $n_1/n \rightarrow \gamma$ where γ 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}, U_{ij} can be dependent. Because of censoring, we only observe $X_{ij} = T_{ij} \wedge U_{ij}$ and $\delta_{ij} = I(T_{ij} \leq U_{ij})$. We assume that the joint survival of T_{ij}, U_{ij} follows an Archimedean model. Define the right extremes $\tau_{F_{0,1}}, \tau_{F_{0,2}}, \tau_{X,1}, \tau_{X,2}$. To apply the obtained large sample results in the previous section, we assume that $\tau_{F_{0,1}} \vee \tau_{F_{0,2}} \leq \tau_{X,1} \wedge \tau_{X,2} \stackrel{def}{=} \tau$. That is, $[0, \tau]$ fully covers the supports of both latency distributions.

We first focus on the comparison of the cure fractions between the two treatment arms and formulate the following hypotheses

$$H_0 : p_1 = p_2 (= p) \text{ vs } H_1 : p_1 \neq p_2. \quad (22)$$

Denote by \hat{p}_i the estimate of p_i in arm $i, i = 1, 2$. Then under the null hypothesis in (22), from Proposition 6, we have

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

where Z_1^∞ and Z_2^∞ are independent and are as defined in Proposition 5 (with an added subscript for each treatment arm). Hence a Wald-type test statistic

$$(\hat{p}_1 - \hat{p}_2) / \sqrt{\frac{(1-\gamma)\hat{v}_{0,1}^\infty + \gamma\hat{v}_{0,2}^\infty}{n\gamma(1-\gamma)\{\phi'(1-\hat{p})\}^2}} \quad (23)$$

will approximately follow a standard normal distribution. Here $\hat{v}_{0,i}^\infty$ are the consistent estimates of $v_{0,i}^\infty$ as defined in Proposition 4 (with an added subscript for each treatment arm) and the pooled estimate $\hat{p} = (n_1\hat{p}_1 + n_2\hat{p}_2)/n$.

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

$$H_0 : F_{0,1} = F_{0,2} \text{ vs } H_1 : F_{0,1} \neq F_{0,2}. \quad (24)$$

Based on $(X_{ij}, \delta_{ij}), j = 1, \dots, n$, 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 (5) and $\hat{p}_i = \hat{F}_i(X^{n_{i^*}})$ is the consistent estimator for p_i , the estimated non-cure fraction in the i -th arm.

Denote the pooled conditional distribution by

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

To test H_0 in (24), 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

$$W_n = \sqrt{n} \left[\int_0^\infty |\hat{F}_{0,1}(t) - \hat{F}_{0,2}(t)|^r d\hat{F}_{0,pool}(t) \right]^{\frac{1}{r}}, \quad (25)$$

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

The following proposition gives the asymptotic distribution of W_n , under $H_0 : F_{0,1} = F_{0,2} (= F_0)$.

Proposition 9 *Assume that $n_1/n \rightarrow \gamma$. Then under the null hypothesis in (24),*

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

if $r < \infty$, and

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

if $r = \infty$, where the Gaussian process 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 (39) (with an added subscript).

For gaining additional insight into the limiting distribution of \mathcal{X} , we take the case of $r = 2$, and consider a Loève type expansion in terms of principal components under H_0 . Specifically, we represent the distribution of the random variable \mathcal{X} in Proposition 9 as a mixture of noncentral chi-squares, which would facilitate the numerical realizations.

By exploiting the independence of $G_1(\cdot)$ and $G_2(\cdot)$, we can compute the covariance function $K(s, t)$ of the Gaussian process $G(\cdot)$. Specifically,

$$K(s, t) = E\{G(s)G(t)\} = \gamma^{-1}\{a_1(s, t)\mathcal{C}_1(s, t) + b_1(s, t)v_{0,1}^\infty - d_1(s, t)\mathcal{C}_1^\infty(t) - d_1(t, s)\mathcal{C}_1^\infty(s)\} \\ + (1 - \gamma)^{-1}\{a_2(s, t)\mathcal{C}_2(s, t) + b_2(s, t)v_{0,2}^\infty - d_2(s, t)\mathcal{C}_2^\infty(t) - d_2(t, s)\mathcal{C}_2^\infty(s)\}, \quad (26)$$

where

$$a_i(s, t) = p_i^{-2}\{\phi'(1 - p_i F_0(t))\phi'(1 - p_i F_0(s))\}^{-1}, \\ b_i(s, t) = (p_i\phi'(1 - p_i))^{-2}F_0(t)F_0(s), \\ d_i(s, t) = -(p_i^2\phi'(1 - p_i F_0(t))\phi'(1 - p_i))^{-1}F_0(s),$$

and $\mathcal{C}_i(s, t)$, $\mathcal{C}_i^\infty(s)$ are as defined in Proposition 5 (with an added subscript) for $i = 1, 2$. The following proposition presents the result of the Loève principal component decomposition of \mathcal{X} .

Proposition 10 *The distribution for the limiting random variable \mathcal{X} can be represented as the following mixture of noncentral chi-squares*

$$\mathcal{X} \stackrel{D}{=} \sum_{k=1}^{\infty} \lambda_k Z_k^2, \quad (27)$$

where Z_k are i.i.d. standard normal random variables and λ_k are the eigenvalues of a symmetric compact positive linear operator \mathcal{T} on Hilbert space $(L^2([0, \infty]), (\cdot, \cdot))$ with inner product $(f, g) = \int_0^\tau f(s)g(s)F^*(ds)$,

$$(\mathcal{T}f)(t) = \int_0^\infty K(s, t)f(s)F^*(ds).$$

Again, without loss of generality, we may assume that the λ_k are decreasing in k to zero.

The proof is similar to Li and Feng (2005), which was derived when T and U are independent (i.e. $\phi(t) = -\log(t)$) and is thus omitted.

7 Numerical Studies

7.1 Real Data Example

We applied the developed methods to analyze the prostate cancer data in SEER Cancer Incidence Public-Use Database, released on April 2004 based on the November 2003 submission. We focused on prostate cancer cases in Connecticut and Detroit Metropolitan area diagnosed between year 1973 and 2001 and during the early stages of the disease, where the tumor was still in situ, localized or regional, excluding cases where the cancer had spread to remote parts of the body. There were 91,873 such cases, of which 75,615 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 health disparities. Using the theoretical results of the previous sections, we tested whether or not the two subpopulations were different in these respects.

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 factor, e.g. 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 the 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 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 the CVD and the prostate cancer, we might see that the overall dependence among the prostate cancer death and the censoring would become weaker as fewer censorings are due to CVD. Hence, the data would yield a trend of higher cancer cure rates, indicating overall progress against cancer. On the other hand, 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 cancer; though in reality there

might be a higher true cure rate. Theoretical justifications for this phenomenon will be given in Proposition 11 in the discussion section.

Figure 1 and Table 1 present 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 τ , the cure fractions for whites are uniformly higher than those for non-whites. Figure 2 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 2 displays the results for dependent censoring under Frank's family for different Kendall's tau. The first four columns test the null hypothesis (24) that the latency distribution for whites ($F_{0,1}$) equals that for non-whites ($F_{0,2}$). Columns 1 and 2 display the Cramer-von Mises test statistics defined in (25) with the p-values estimated using Proposition 9. The third and fourth columns present the results for the Kolmogorov-Smirnov test. For $\tau \leq 0.19$, there is no strong evidence at the 1% significance level that the latency distributions of whites and non-whites are different. On the other hand, 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 fifth and sixth columns present the results for testing whether whites and non-whites have the same cure fractions. The theory is developed in section 6.2 and the test statistics is defined in (23). There is strong evidence that the cure fractions are different for the two subpopulations. Using the theoretical results derived in Section 6.1, 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 (16) was used to compute the bounds since the estimated hazard $\hat{\lambda}_0(\cdot)$ was found to be non-decreasing. Expression (13) 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 1. 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. Analysis assuming dependent censoring under Clayton's family of copulas were similar to those obtained under Frank's family and lead to the same conclusions.

7.2 Simulation Study

We investigated by simulation the finite-sample behavior of the cure fraction estimator, i.e. $\hat{p} = 1 - \hat{S}_T(X^n)$. Proposition 6 proves that this estimator is consistent and asymptotically normal, with (11) 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 parameter $a = 0, 2.1, 5.7$, corresponding roughly to Kendall's tau $0, 0.2, 0.47$. The latency distribution of the failure times was exponential with mean one 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, 0.2, resulting 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 (11).

The above steps were repeated for 3,000 replications to obtain estimates of p based on the 3,000 different data sets. The empirical variance of \hat{p} was computed and compared with the average asymptotic variance. Table 3 presents the results for various combination 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 tend to zero, and the empirical coverage probabilities of the 95% confidence intervals approach the nominal value of the confidence level. These results verify the validity of Proposition 6. For a given value of dependence parameter a and sample size n , the standard errors decrease with censoring rate r . This is reasonable because a smaller censoring rate 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 derived in (8). In particular, Table 4 uses the simulated data to verify expression (8) 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.

8 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 modeled using a class of Archimedean copula models with a known ϕ function. In practice, however, selecting a right ϕ function in the copula is often hampered by the fact that with the current data on (X, δ) , the copula model is not identifiable (Tsiatis, 1978). In some applications where both the censoring and failure times were observed for a sub-sample (Bartholomew, 1957), a suitable copula function could be identified with this additional information. Hence, the estimation of cure models incorporating, in the framework outlined in this article, an assessment of the dependence between the censoring and failure times appears to be a promising research area; some related work can be found in Lin et al. (1996) and Wang (2003).

Additionally, based on the analytical framework we have set up it would be feasible to conduct bias analysis when the dependence structure between survival and censoring times are misspecified. In particular, we can quantify the biases in the estimates of cure fractions for such misspecifications.

Using the same argument in Proposition 1, we can show that for any ϕ function (which satisfies the regularity conditions (c.1)-(c.5)), the estimate based on (5) converges uniformly to

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

When ϕ 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 Kelin (1996) and Proposition 2 of Rivest and Wells (2001), the following proposition concerns the asymptotic impact of changing

the level of dependency between T and U on estimating the cure fractions.

Proposition 11 *Let ϕ_1 and ϕ_2 be two functions used in (5). If $\phi_1'(t)/\phi_2'(t)$ increases in t then the asymptotic limit of cure fraction $1 - p_1^* \geq 1 - p_2^*$.*

Indeed Proposition 11 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 ϕ_1 corresponds to less dependence between T and U than ϕ_2 under (3). Thus Proposition 11 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 proposed by Maller and Zhou (1992)) will tend to over-estimate the true cure fraction. On the other hand, if there exists negative dependence between T and U , a naive Kaplan-Meier estimate will under-estimate the true cure fraction.

9 Acknowledgements

The first author thanks Professors David Harrington and Robert Gray for helpful comments.

Appendix: Technical Details

A.0: Regularity Conditions

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

(c.1) ϕ is strictly decreasing on $(0, 1]$ and is sufficiently smooth. Further assume that the first two derivatives of $\phi(s)$ and $\psi(s) \stackrel{def}{=} -s\phi'(s)$ are bounded for $s \in [\epsilon, 1]$ where $\epsilon > 0$ is arbitrary.

In addition, the first derivative of $\phi(s)$ is bounded away from 0 on $[0, 1]$.

(c.2) $0 < \int_0^{\tau_X} \{\psi(\pi(s))\}^k d\tilde{\Lambda}(s) < \infty$ for $k = 0, 1, 2$

(c.3) $\int_0^{\tau_X} |\psi'(\pi(s))| d\tilde{\Lambda}(s) < \infty$

(c.4) $\limsup_{t \rightarrow \tau_X} \int_t^{\tau_X} \frac{(\psi(\pi(s)))^2}{\pi(s)} d\tilde{\Lambda}(s) = 0$

(c.5) $S_T(t)$ and $S_0(t)$ are continuous over $[0, \tau_X]$ if $\tau_X < \infty$. Otherwise, define $S_T(\infty) = \lim_{t \rightarrow \infty} S_T(t)$.

A.1: Proof of Proposition 1

First show that for any fixed t_0 such that $\pi(t_0) > 0$,

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

A Taylor expansion and the regularity condition (c.1) on ϕ gives, on $[0, t_0]$,

$$\phi(\hat{S}_T(t)) = - \int_0^t \frac{1}{n} \phi' \left\{ \frac{Y(s)}{n} \right\} dN(s) + e_n,$$

where $e_n = o_p(1)$ uniformly over $[0, t_0]$. Some algebra yields

$$\phi(S_T(t)) = - \int_0^t \phi'(\pi(s)) \pi(s) d\tilde{\Lambda}(s).$$

Hence,

$$\begin{aligned} & \phi(\hat{S}_T(t)) - \phi(S_T(t)) \\ = & - \frac{1}{n} \int_0^t I(Y(s) > 0) \phi' \left\{ \frac{Y(s)}{n} \right\} dM(s) + \int_0^t I(Y(s) > 0) \left[\psi \left\{ \frac{Y(s)}{n} \right\} - \psi(\pi(s)) \right] d\tilde{\Lambda}(s) \\ & - \int_0^t I(Y(s) = 0) \phi'(\pi(s)) \pi(s) d\tilde{\Lambda}(s) + e_n \\ = & Z_1(t) + Z_2(t) + Z_3(t) + e_n, \end{aligned}$$

where $M(s) = \sum_{i=1}^n M_i(s)$ is a martingale.

When $t \in [0, t_0]$,

$$\begin{aligned} 0 < Z_3(t) & \leq I(Y(t) = 0) \int_0^t \psi(\pi(s)) d\tilde{\Lambda}(s) \\ & < I(Y(t_0) = 0) \int_0^{\tau_X} \psi(\pi(s)) \pi(s) d\tilde{\Lambda}(s). \end{aligned}$$

By the strong law of large numbers $Y(t_0)/n \rightarrow \pi(t_0) (> 0)$ almost surely. Hence $Y(t_0) \rightarrow \infty$ almost surely. From this, coupled with the regularity condition (c.2), we have the uniform convergence of $Z_3(t)$ over $[0, t_0]$. It remains to demonstrate the convergence of $Z_1(t)$ and $Z_2(t)$.

Consider the variation process of $Z_1(t)$,

$$\begin{aligned} \langle Z_1, Z_1 \rangle(t) & = \int_0^t I(Y(s) > 0) \left[\phi' \left\{ \frac{Y(s)}{n} \right\} \right]^2 \frac{Y(s)}{n^2} d\tilde{\Lambda}(s) \\ & = \int_0^t \frac{I(Y(s) > 0)}{Y(s)} \left[\psi \left\{ \frac{Y(s)}{n} \right\} \right]^2 d\tilde{\Lambda}(s). \end{aligned}$$

Then it follows that $Z_1^2(t) - \langle Z_1, Z_1 \rangle(t)$ is a martingale. By Lenglar's inequality (see, e.g., Fleming and Harrington (1991))

$$\begin{aligned} & P\left(\sup_{t \in [0, t_0]} |Z_1(t)| > \epsilon\right) \\ & < \frac{\eta}{\epsilon^2} + P\left(\int_0^{t_0} \frac{I(Y(s) > 0)}{Y(s)} \left(\psi\left(\frac{Y(s)}{n}\right)\right)^2 d\tilde{\Lambda}(s) > \eta\right) \\ & < \frac{\eta}{\epsilon^2} + P\left\{\frac{1}{Y(t_0)} \int_0^{t_0} \left(\psi\left(\frac{Y(s)}{n}\right)\right)^2 d\tilde{\Lambda}(s) > \eta\right\}. \end{aligned}$$

Since the empirical process $\frac{Y(s)}{n} \rightarrow \pi(s)$ in probability uniformly on $[0, \infty)$ and because of the boundness regularity conditions on $\psi(\cdot)$ and $\psi'(\cdot)$ on $[\pi(t_0), 1]$, $\psi^2\left(\frac{Y(s)}{n}\right)$ converges to $\psi^2(\pi(s))$ uniformly on $[0, t_0]$.

Hence $\int_0^{t_0} \left(\psi\left(\frac{Y(s)}{n}\right)\right)^2 d\tilde{\Lambda}(s) \rightarrow \int_0^{t_0} \left(\psi(\pi(s))\right)^2 d\tilde{\Lambda}(s) < \infty$ (by the regularity condition (c.2)). So $\frac{1}{Y(t_0)} \int_0^{t_0} \left(\psi\left(\frac{Y(s)}{n}\right)\right)^2 d\tilde{\Lambda}(s) \xrightarrow{pr} 0$ as $Y(t_0) \xrightarrow{pr} \infty$. Hence, $P(\sup_{0 \leq t \leq t_0} |Z_1(t)| > \epsilon) \rightarrow 0$ for any $\epsilon > 0$.

Now consider

$$Z_2(t) = \int_0^t I(Y(s) > 0) \psi'(\pi(s)) \left\{ \frac{Y(s)}{n} - \pi(s) \right\} d\tilde{\Lambda}(s) + e_n$$

where $e_n = o_p(1/n)$ uniformly on $[0, t_0]$. Further note that

$$\sup |Z_2(t)| \leq \sup |e_n| + \left\{ \int_0^{t_0} |\psi'(\pi(s))| d\tilde{\Lambda}(s) \right\} \sup_{0 \leq s \leq t_0} \left| \frac{Y(s)}{n} - \pi(s) \right|,$$

which implies (under the regularity condition (c.3)) that

$$\sup_{0 \leq t \leq t_0} |Z_2(t)| \xrightarrow{pr} 0.$$

Thus we have proved that

$$\sup_{0 \leq t \leq t_0} |\phi(\hat{S}_T(t)) - \phi(S_T(t))| \xrightarrow{pr} 0$$

for any t_0 such that $\pi(t_0) > 0$.

Now we show that

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

We only consider the situation when $\tau_X < \infty$ as the proof follows similarly when $\tau_X = \infty$. Fix a small $\epsilon > 0$ and consider any $t \in [\tau_X - \epsilon, \tau_X]$. With monotonicity of S_T and ϕ , it follows that

$$|\phi(\hat{S}_T(t)) - \phi(S_T(t))|$$

$$\begin{aligned}
&< |\phi(\hat{S}_T(\tau_X)) - \phi(\hat{S}_T(\tau_X - \epsilon))| + |\phi(\hat{S}_T(\tau - \epsilon)) - \phi(S_T(\tau_X - \epsilon))| \\
&\quad + |\phi(S_T(\tau_X - \epsilon)) - \phi(S_T(\tau_X))|.
\end{aligned}$$

Also note that

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

Using the uniform convergence of $\hat{S}_T(t)$ on $[0, \tau_X - \epsilon]$ and letting $\epsilon \rightarrow 0+$ yields

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

As $\phi'(\cdot)$ is bounded away from 0 on $[0, 1]$ (condition c.1), a Taylor expansion immediately yields $\sup_{0 \leq t \leq \tau_X} |\hat{S}_T(t) - S_T(t)| \xrightarrow{pr} 0$.

Applying a similar argument, we may demonstrate the uniform convergence of $\int_0^t I(Y(s) > 0) \frac{dN}{Y}$ to $\tilde{\Lambda}(t)$ on $[0, \tau_X]$ by observing that

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

A.2: Proof of Proposition 2

By the definition of τ_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 τ_X , we have $\hat{S}_T(X^n) \xrightarrow{pr} S_T(\tau_X)$. So $\hat{S}_T(X^n) \xrightarrow{pr} (1-p)$ 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 $\tau_{F_0} \leq \tau_X$.

A.3: Proof of Proposition 3

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

On the other hand, 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 (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 $\tau_U \leq \tau_X$.

A.4: Proof of Proposition 4

We first show by contradiction that if $\alpha_n \rightarrow 0$ in probability (or equivalently $N_n \rightarrow \infty$ in probability), then $\tau_{F_0} < \tau_U$. Otherwise, if we assume $\tau_U \leq \tau_{F_0}$, we show that N_n does not converge to ∞ in probability, which induces a contradiction.

Choose a constant $b > 0$ such that $\frac{1}{b} + e^{-b/M} < \frac{1}{2}$. Define $a_n = \inf\{a : h(2a) \geq \frac{b}{n}\}$. By the condition (d.1), a_n is well defined (or at least for large n), $a_n \downarrow 0$ and $h(2a_n) = \frac{b}{n}$ by the continuity of h . Define an event $A_n = \{X^{n*} > \tau^* - a_n\}$. The key idea of the proof is to show A_n happens with a large probability, while on A_n , N_n is bounded with a large probability, resulting in a contradiction.

Indeed

$$P(A_n) = P(X^{n*} > \tau^* - a_n) = P\{\cup_{i=1}^n (U_i \geq T_i > \tau^* - a_n)\}.$$

By independence across the subjects

$$P(A_n^c) = \prod P\{(U_i \geq T_i > \tau^* - a_n)^c\} = (1 - h(a_n))^n \leq (1 - h(2a_n)/M)^b \leq e^{-b/M}.$$

By assumption $\tau_U \leq \tau_{F_0}$, we have $X^{n*} \leq X^n \leq \tau^* \stackrel{def}{=} \tau_U \wedge \tau_{F_0} = \tau_U$ almost surely. Furthermore, on A_n we have $\tau^* - a_n \leq X^{n*} \leq X^n \leq \tau$, whence,

$$[2X^{n*} - X^n, X^{n*}] \subset [\tau^* - 2a_n, \tau^*].$$

Define indicator $Y_{in} = I(\tau^* - 2a_n \leq X_i \leq \tau^*, \delta_i = 1)$. As N_n is the number of uncensored observations on $[2X^{n*} - X^n, X^{n*}]$, it follows that if A_n happens, then $N_n \leq \sum_i Y_{in}$. Thus

$$E(Y_{in}) = P(\tau^* - 2a_n \leq T_1 \leq U_1 \leq \tau^*) = h(2a_n) = \frac{b}{n}.$$

and $\text{var}(Y_{in}) \leq E(Y_{in}^2) = E(Y_{in}) = \frac{b}{n}$.

Therefore, by Chebyshev's inequality

$$\begin{aligned} P(N_n > 2b) &\leq P(N_n > 2b, A_n) + P(A_n^c) \\ &\leq P(\sum Y_{in} > 2b, A_n) + P(A_n^c) \\ &\leq P(\sum Y_{in} > 2b) + e^{-b/M} \\ &\leq P(\sum (Y_{in} - b/n) > b) + e^{-b/M} \\ &\leq \frac{n\text{var}(Y_{1n})}{b^2} + e^{-b/M} \\ &\leq \frac{1}{b} + e^{-b/M} < 1/2. \end{aligned}$$

This contradicts with that $N_n \rightarrow \infty$ in probability. Hence the assumption $\tau_U \leq \tau_{F_0}$ must not hold, and we must have $\tau_{F_0} < \tau_U$.

Now we show the converse is true. That is, $\tau_{F_0} < \tau_U$ implies $N_n \rightarrow \infty$ in probability. Indeed, $X^{n*} \rightarrow \tau_{F_0}$ and $X^n \rightarrow \tau_U$ almost surely. Hence for any fixed $\epsilon > 0$

$$\tau_{F_0} - \epsilon \leq X^{n*} \leq \tau_{F_0} \leq \tau_U - \epsilon \leq X^n \leq \tau_U,$$

with probability 1 when n is sufficiently large.

Therefore, with probability 1,

$$[\tau^* - \eta + \epsilon, \tau^*] \subset [2X^{n*} - X^n, X^{n*}],$$

where $\tau^* = \tau_{F_0} \wedge \tau_U = \tau_{F_0}$ and $\eta = \tau_U - \tau_{F_0}$. Let $Z_i = I(\tau^* - \eta + \epsilon \leq X_i \leq \tau^*, \delta_i = 1)$. Then $N_n \geq \sum_i Z_i$ almost surely.

But

$$E(Z_i) = \int_{\tau^* - \eta + \epsilon}^{\tau^*} \pi(t) d\tilde{\Lambda}(t) = h(\eta - \epsilon).$$

So if we take $\epsilon < \eta - \epsilon_0$, we have $E(Z_i) > h(\epsilon_0) > 0$ (by the condition (d.1)). Therefore the law of large numbers gives $\sum_i Z_i/n \rightarrow E(Z_i) > 0$ in probability. Hence $\sum_i Z_i \rightarrow \infty$ in probability, which indicates $N_n \rightarrow \infty$ in probability.

A.5: Proof of Proposition 5

Using the same argument as in Rivest and Wells (2001), up to an $o_p(1)$ term, we have that

$$\begin{aligned} Z_n(t) &= \sqrt{n} \left(-\frac{1}{n} \int_0^{t \wedge X^n} I(Y(s) > 0) \phi' \left\{ \frac{Y(s)}{n} \right\} dM(s) \right. \\ &\quad \left. + \int_0^{t \wedge X^n} I(Y(s) > 0) \left[\psi \left\{ \frac{Y(s)}{n} \right\} - \psi(\pi(s)) \right] d\tilde{\Lambda}(s) \right) \\ &= Z_{n,1}(t) + Z_{n,2}(t). \end{aligned} \tag{28}$$

Rivest and Wells (2001) showed, for any t_0 such that $\pi(t_0) > 0$, $Z_n(t)$ 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) neighborhood of τ_X in view of Theorems 13.2 and 16.8 of Billingsley (1999). That is, it suffices to show for any $\epsilon > 0$

$$\lim_{t \rightarrow \tau_X} \limsup_n P\left(\sup_{s \in (t, \tau_X]} |Z_n(s) - Z_n(t)| > \epsilon \right) = 0; \tag{29}$$

see, also, Gill (1980).

Fix a t . Then

$$\sup_{s \in (t, \tau_X]} |Z_n(s) - Z_n(t)| \leq \sup_{s \in (t, \tau_X]} |Z_{n,1}(s) - Z_{n,1}(t)| + \sup_{s \in (t, \tau_X]} |Z_{n,2}(s) - Z_{n,2}(t)|.$$

Since X^n is a stopping time, and by the optional sampling theorem,

$$Z_{n,1}(s) - Z_{n,1}(t) = -\frac{1}{\sqrt{n}} \int_{t \wedge X^n}^{s \wedge X^n} I(Y(s) > 0) \phi' \left\{ \frac{Y(s)}{n} \right\} dM(s)$$

is a local martingale and its predictable variation process is given by

$$\langle Z_{n,1}(s) - Z_{n,1}(t), Z_{n,1}(s) - Z_{n,1}(t) \rangle = \int_{t \wedge X^n}^{s \wedge X^n} I(Y(s) > 0) \left[\psi \left\{ \frac{Y(s)}{n} \right\} \right]^2 \frac{n}{Y(s)} d\tilde{\Lambda}(s),$$

hence,

$$(Z_{n,1}(s) - Z_{n,1}(t))^2 - \langle Z_{n,1}(s) - Z_{n,1}(t), Z_{n,1}(s) - Z_{n,1}(t) \rangle$$

is a martingale (again assume that t is fixed).

Therefore, by Lenglar's inequality we have

$$\begin{aligned}
& P(\sup_{[t, \tau_X]} |Z_{n,1}(s) - Z_{n,1}(t)| > \epsilon) \\
& < \frac{\eta}{\epsilon^2} + P\left(\int_{t \wedge X^n}^{\tau_X \wedge X^n} I(Y(s) > 0) \left[\psi \left\{ \frac{Y(s)}{n} \right\}\right]^2 \frac{n}{Y(s)} d\tilde{\Lambda}(s) > \eta\right) \\
& \leq \frac{\eta}{\epsilon^2} + P\left(\int_t^{\tau_X} \left[\psi \left\{ \frac{Y(s)}{n} \right\}\right]^2 \frac{n}{Y(s)} d\tilde{\Lambda}(s) > \eta\right). \tag{30}
\end{aligned}$$

Because of the uniform convergence of $\frac{Y(s)}{n}$ to $\pi(s)$ on $[0, \tau_X]$,

$$\int_t^{\tau_X} \left[\psi \left\{ \frac{Y(s)}{n} \right\}\right]^2 \frac{n}{Y(s)} d\tilde{\Lambda}(s) \xrightarrow{pr.} \int_t^{\tau_X} \frac{(\psi(\pi(s)))^2}{\pi(s)} d\tilde{\Lambda}(s),$$

for any $t < \tau_X$. Hence by the regularity condition (c.4), the second term in (30) converges to 0 for any $\eta > 0$ as $t \rightarrow \tau_X$. Hence, we have

$$\lim_{t \rightarrow \tau_X} \limsup_n P(\sup_{[t, \tau_X]} |Z_{n,1}(s) - Z_{n,1}(t)| > \epsilon) = 0. \tag{31}$$

Now we turn to show that

$$\lim_{t \rightarrow \tau_X} \limsup_n P(\sup_{[t, \tau_X]} |Z_{n,2}(s) - Z_{n,2}(t)| > \epsilon) = 0. \tag{32}$$

As

$$Z_{n,2}(s) - Z_{n,2}(t) = \int_{t \wedge X_n}^{s \wedge X_n} I(Y(s) > 0) \psi'(\pi(s)) \sqrt{n} \left\{ \frac{Y(s)}{n} - \pi(s) \right\} d\tilde{\Lambda}(s) + o_p(1),$$

it follows that (32) holds as $\sqrt{n}(\frac{Y(s)}{n} - \pi(s))$ converges weakly to a tight Gaussian process over $[0, \infty)$. Combining (31) and (32) gives (29). Hence the proposition follows.

For completeness we compute below the covariance function for the limiting process $Z(t)$, which is needed in computing the asymptotic distribution of the test statistic derived later. The derivation of this covariance function, which is not given in Rivest and Wells (2001), is involved as $Z(t)$ is not an independent increment process.

For any $t < \tau_X$, as $X^n \rightarrow \tau_X$ almost surely and following Rivest and Wells (2001), we can show that (28) is asymptotically equal to (up to an $o_p(1)$ term)

$$W_n(t) = \frac{1}{\sqrt{n}} \int_0^t -\phi'(\pi(u)) dM(u) + \int_0^t X_n(s) \psi'(\pi(s)) d\tilde{\Lambda}(s) = W_{n,1}(t) + W_{n,2}(t),$$

where $X_n(s) = \sqrt{n} \left\{ \frac{Y_n(s)}{n} - \pi(s) \right\}$. Hence we only need to compute the limiting covariance function for $W_n(t)$.

Consider $0 \leq t_1 \leq t_2 \leq \tau_X$. Then

$$\begin{aligned} \text{cov}\{W_n(t_1), W_n(t_2)\} &= E\{W_{n,1}(t_1)W_{n,1}(t_2)\} + E\{W_{n,2}(t_1)W_{n,2}(t_2)\} \\ &\quad + E\{W_{n,1}(t_1)W_{n,2}(t_2)\} + E\{W_{n,1}(t_2)W_{n,2}(t_1)\}. \end{aligned}$$

Since $W_{n,1}(\cdot)$ is a square integrable martingale,

$$E\{W_{n,1}(t_1)W_{n,1}(t_2)\} = \frac{1}{n} E \left\{ \int_0^{t_1} [\phi'\{\pi(s)\}]^2 Y(s) d\tilde{\Lambda}(s) \right\} = \int_0^{t_1} [\phi'\{\pi(s)\}]^2 \pi(s) d\tilde{\Lambda}(s).$$

Also

$$\begin{aligned} &E\{W_{n,2}(t_1)W_{n,2}(t_2)\} \\ &= E \int_0^{t_2} \int_0^{t_1} (Y_1(u) - \pi(u))(Y_1(s) - \pi(s)) \psi'(\pi(u)) d\tilde{\Lambda}(u) \psi'(\pi(s)) d\tilde{\Lambda}(s) \\ &= E \int_0^{t_1} \int_0^{t_1} (Y_1(u) - \pi(u))(Y_1(s) - \pi(s)) \psi'(\pi(u)) d\tilde{\Lambda}(u) \psi'(\pi(s)) d\tilde{\Lambda}(s) \\ &\quad + E \int_{t_1}^{t_2} \int_0^{t_1} (Y_1(u) - \pi(u))(Y_1(s) - \pi(s)) \psi'(\pi(u)) d\tilde{\Lambda}(u) \psi'(\pi(s)) d\tilde{\Lambda}(s) \\ &= 2 \int_0^{t_1} \int_0^s \pi(s)(1 - \pi(u)) \psi'(\pi(u)) d\tilde{\Lambda}(u) \psi'(\pi(s)) d\tilde{\Lambda}(s) \\ &\quad + \int_{t_1}^{t_2} \pi(s) \psi'(\pi(s)) d\tilde{\Lambda}(s) \int_0^{t_1} (1 - \pi(u)) \psi'(\pi(u)) d\tilde{\Lambda}(u), \end{aligned}$$

where the calculation of $E \int_0^{t_1} \int_0^{t_1}$ comes from Rivest and Wells (2001) (after correcting a typographic error in their formula).

Introduce $A(s) = -\phi'(\pi(s))$ and $B(s)ds = \psi'(\pi(s))d\tilde{\Lambda}(s)$. Applying the result $E(M_1(u)Y_1(s)) = -\pi(s)\tilde{\Lambda}(u \wedge s)$ and integration by parts, we have

$$\begin{aligned} &\text{cov}\{W_{n,2}(t_1), W_{n,1}(t_2)\} \\ &= E \left\{ \int_0^{t_1} A(u) dM_1(u) \int_0^{t_2} (Y_1(s) - \pi(s)) B(s) ds \right\} \\ &= E \left\{ \int_0^{t_2} A(t_1) M_1(t_1) Y_1(s) B(s) ds \right\} \end{aligned} \tag{33}$$

$$+ E \left\{ \int_0^{t_2} \int_0^{t_1} -M_1(u) Y_1(s) dA(u) B(s) ds \right\} \tag{34}$$

Using $\int_0^{t_2} = \int_0^{t_1} + \int_{t_1}^{t_2}$, (33) is

$$-\tilde{\Lambda}(t_1) \int_0^{t_1} \pi(s) \tilde{\Lambda}(s) B(s) ds - \tilde{\Lambda}(t_1) A(t_1) \int_{t_1}^{t_2} \pi(s) B(s) ds \quad (35)$$

while using $\int_0^{t_2} \int_0^{t_1} = \int_0^{t_1} \int_0^{t_1} + \int_{t_1}^{t_2} \int_0^{t_1}$, (34) is

$$\int_0^{t_1} \int_0^{t_1} \pi(s) \tilde{\Lambda}(u \wedge s) dA(u) B(s) ds + \int_{t_1}^{t_2} \int_0^{t_1} \pi(s) \tilde{\Lambda}(u) dA(u) B(s) ds. \quad (36)$$

Adding the first term of (35) and the first term of (36) gives $-\int_0^{t_1} \int_0^s A(u) d\tilde{\Lambda}(u) \pi(s) B(s) ds$ following Rivest and Wells (2001, though a minus sign is missing in their formulation). Integration by parts with respect to $dA(u)$ in the second term of (36) gives the summation of the second term in (35) and the second term in (36) is

$$-\int_{t_1}^{t_2} \pi(s) B(s) ds \int_0^{t_1} A(u) d\tilde{\Lambda}(u)$$

So,

$$\text{cov}\{W_{n,1}(t_1), W_{n,2}(t_2)\} = -\int_0^{t_1} \int_0^s A(u) d\tilde{\Lambda}(u) \pi(s) B(s) ds - \int_{t_1}^{t_2} \pi(s) B(s) ds \int_0^{t_1} A(u) d\tilde{\Lambda}(u). \quad (37)$$

Similarly we obtain

$$\text{cov}\{W_{n,1}(t_2), W_{n,2}(t_1)\} = -\int_0^{t_1} \int_0^s A(u) d\tilde{\Lambda}(u) \pi(s) B(s) ds. \quad (38)$$

Plugging back $A(u) = -\phi'(\pi(s))$ and $B(s) ds = \psi'(\pi(s)) d\tilde{\Lambda}(s)$ in (37) and (38) and using the weak convergence of a tight process W_n to $Z(t)$, we have thus obtained the covariance function $\mathcal{C}(t_1, t_2)$ as stated in the proposition.

A.6: Proof of Proposition 6

Note that

$$\begin{aligned} \sqrt{n}\{\phi(1 - \hat{p}) - \phi(1 - p)\} &= \sqrt{n}\{\phi(\hat{S}_T(X^n)) - \phi(S_T(X^n))\} + \sqrt{n}\{\phi(S_T(X^n)) - \phi(1 - p)\} \\ &= Z_n(\tau_X) + \sqrt{n}\{\phi(S_T(X^n)) - \phi(1 - p)\} \end{aligned}$$

where $Z_n(\tau_X)$, as defined in (7), converges weakly to Z^∞ by Proposition (5).

We only need to show $\sqrt{n}\{\phi(S_T(X^n)) - \phi(1-p)\} \xrightarrow{pr} 0$. For a fixed $\epsilon > 0$, consider an increasing sequence a_n such that

$$F_0(a_n -) \leq 1 - \frac{\epsilon}{\sqrt{np}} \leq F_0(a_n).$$

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

$$P(\sqrt{n}|p - F(X^n)| > \epsilon) = P(\sqrt{n}|p - pF_0(X^n)| > \epsilon) = P(X^n \leq a_n).$$

Assume (9), which implies that the tail of the observed survival times is heavier than that of the latency distribution. Indeed when $\tau_{F_0} < \tau_X (= \tau_U)$ this assumption holds immediately as $\lim_{t \rightarrow \tau_{F_0}} \frac{1-F_0(t)}{\pi(t)} = 0$. Under assumption (9) we have $\pi(t) \geq 1 - F_0(t)$ when t is sufficiently close to τ_{F_0} . Hence when n is sufficiently large

$$P(X^n \leq a_n) = (1 - \pi(a_n))^n \leq \left(1 - \frac{\epsilon}{\sqrt{np}}\right)^n \rightarrow 0.$$

Thus, $\sqrt{n}|p - F(X^n)|$ converges to 0 in probability and so does $\sqrt{n}|\phi(1-p) - \phi(S_T(X^n))|$ to 0 by the boundedness condition on $\phi'(\cdot)$ (the regularity condition (c.1)). Therefore (10) holds, which also implies (11) by the Slutsky theorem.

A.7: Proof of Proposition 7

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 $\hat{p} - p \xrightarrow{pr} 0$.

A.8: Proof of Proposition 8

First observe that

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

Now since $\sqrt{n}\{\phi(\hat{S}_T(t \wedge X^n) - \phi(S_T(t \wedge X^n)))\} \xrightarrow{w} Z(t)$ on $D[0, \tau_X]$ (Proposition 5) in conjunction with $\sqrt{n}(\hat{p} - p) \xrightarrow{d} \frac{Z^\infty}{-\phi'(1-p)}$ and $X^n \rightarrow \tau_X$ almost surely, it follows that $\sqrt{n}\{\hat{F}_0(t) - F_0(t)\} \xrightarrow{w} G(t)$ on $D[0, \tau_X)$, where

$$G(t) = -\frac{Z(t)}{p\phi'(S_T(t))} + \frac{\{1 - S_T(t)\}Z^\infty}{p^2\phi'(1-p)}. \quad (39)$$

A.9: Proof of Proposition 9

Denote by $X^n = X^{n_1} \wedge X^{n_2}$ and define residual processes $\epsilon_{n_i}(t) = \sqrt{n_i}\{\hat{F}_{0,i}(t \wedge X^n) - F_0(t \wedge X^n)\}$. The sample paths of stochastic processes ϵ_{n_i} reside in the Skorohod space $D_R[0, \tau]$. Then it follows by Proposition 8) $\epsilon_{n_i}(t) \xrightarrow{w} G_i(t)$. Hence by the continuous mapping theorem, when $r < \infty$

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

When $r = \infty$,

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

Reference

- Bartholomew, D. J. (1957), "A problem in life testing," *Journal of the American Statistical Association*, 52, 350-355.
- Berkson, J. and Gage, R.P. (1952), "Survival Curves for Cancer Patients Following Treatment," *Journal of the American Statistical Association*, 47, 501-515.

- Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987), *Regular variation*, Cambridge, UK: Cambridge University Press.
- Billingsley, P. (1999), *Convergence of probability measures*, New York: John Wiley & Sons.
- Boag, J. W. (1949), "Maximum likelihood estimates of the proportion of patients cured by cancer therapy," *JRSSB*, 11, 15-53.
- Genest, C. and MacKay, R. J. (1986), "Archimedean copulas and bivariate families with continuous marginals," *Canadian Journal of Statistics*, 14, 145-159.
- Gill, R. D. (1980), "Censoring and stochastic integrals," CWI, Math. Centrum.
- Fleming, T. R. and Harrington, D. P. (1991) *Counting processes and survival analysis*, New York: John Wiley & Sons.
- Klebanov, L. B. and Yakovlev, A. Y. (2005), "A new approach to testing for sufficient follow up in cure-rate analysis," unpublished manuscript.
- 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 Nonnegligible Cure Fractions," *Lifetime Data Analysis*, 11, 367-387.
- Lin, D.Y., Robins, J.M. and Wei, L. J. (1996), "Comparing two failure time distributions in the presence of dependent censoring," *Biometrika*, 83, 381-393.
- 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," *Journal of the American Statistical Association*, 89, 1499-1506.
- Maller, R. A. and Zhou, S. (1996), *Survival analysis with long-term survivors*, New York: John Wiley & Sons.

- Rivest, L.P. and Wells, M.T. (2001), "A martingale approach to the copula-graphic estimator for the survival function under dependent censoring," *J. of Multivariate Analysis*, 79, 138-155.
- Peng, Y. and Dear, K.B. (2000), "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics* 56, 237-243.
- Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Public-Use Data (1973-2001), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2004, based on the November 2003 submission.
- 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," *Scandinavian Actuarial Journal*, 235-239.
- Wang, W.J. (2003), "Estimating the association parameter for copula models under dependent censoring," *JRSSB*, 65: 257-273.
- Zheng, M. and Klein, J. P. (1995), "Estimates of marginal survival for dependent competing risks based on an assumed copula," *Biometrika*, 82, 127-138.



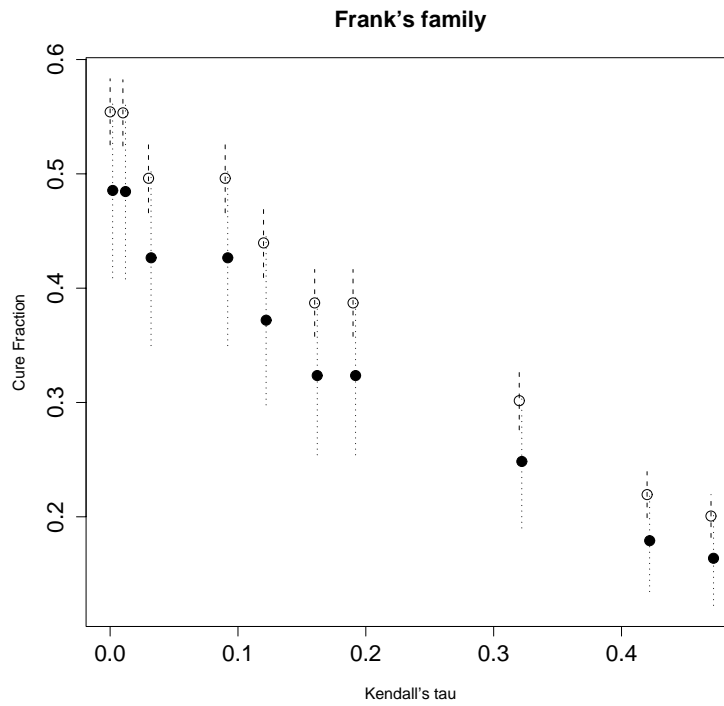


Figure 1: Estimated prostrate cancer cure fractions for whites (○) and non-whites (●) under Frank's family of copulas. The lines represent margins of two standard errors.



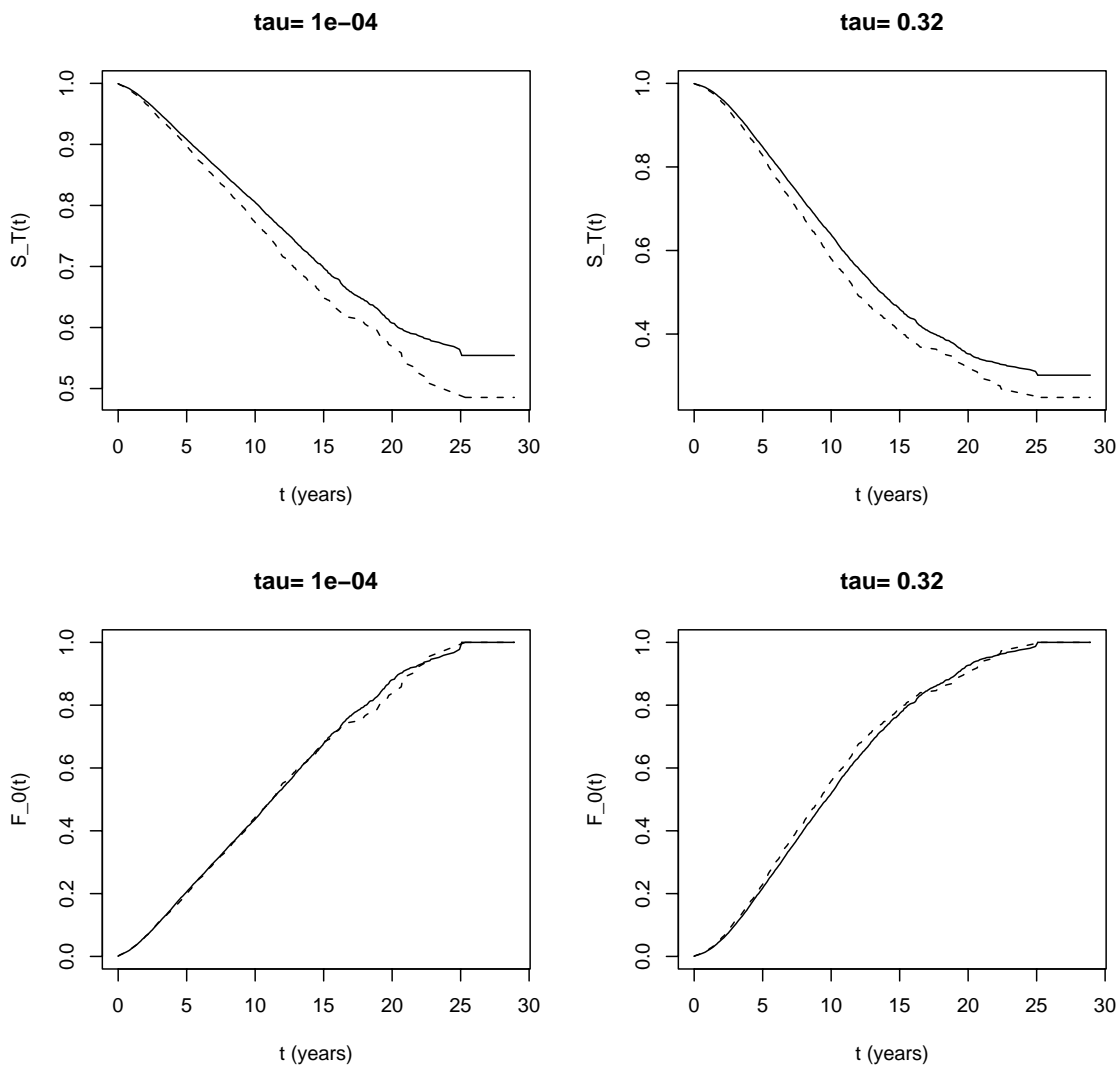
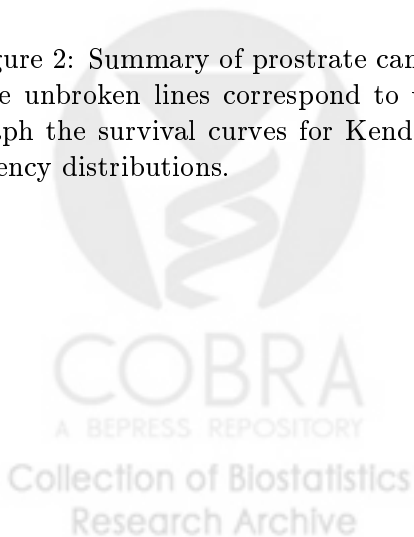


Figure 2: Summary of prostrate cancer results for Archimedean copulas based on Frank's family. The unbroken lines correspond to whites and the dashed lines to non-whites. The top panels graph the survival curves for Kendall's tau equal to 0 and 0.32. The bottom panels graph the latency distributions.



τ	Whites		Non-whites	
	Estimate	SE	Estimate	SE
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

Table 1: Cure fractions based on Frank's family of Archimedean copulas.

τ	CvM Test	P-Value	K-S Test	P-Value	Wald z	P-Value	95% CI for Δ
0	27.79	0.26	15.97	0.06	-3.04	0.00	(0.195 ,1)
0.01	27.67	0.25	15.94	0.06	-3.04	0.00	(0.195 ,1)
0.03	23.18	0.26	14.28	0.07	-2.93	0.00	(0.198 ,1)
0.09	23.18	0.26	14.28	0.07	-2.93	0.00	(0.198 ,1)
0.12	24.51	0.19	12.53	0.08	-2.94	0.00	(0.198 ,1)
0.16	30.74	0.09	11.44	0.07	-3.05	0.00	(0.196 ,1)
0.19	30.74	0.10	11.44	0.09	-3.05	0.00	(0.194 ,1)
0.32	48.01	0.01	13.04	0.01	-3.47	0.00	(0.177 ,1)
0.42	66.24	0.00	13.47	0.00	-4.31	0.00	(0.148 ,1)
0.47	69.82	0.00	13.42	0.00	-4.59	0.00	(0.137 ,1)

Table 2: Results based on Frank's family of Archimedean copulas.

a	r	n	\hat{p}	Empirical SE	Asymptotic SE	95% C.I. 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

Table 3: Summary of simulation results investigating the asymptotic behavior of estimator $\hat{p} = 1 - \hat{S}_T(X^n)$. The true cure fraction was assumed to be $p = 0.3$.

a	r	n	$\hat{S}_T(t_1 = 1)$	$\hat{S}_T(t_2 = 2)$	Empirical Cov	Asymptotic Cov
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	0.0005433187
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	0.0005302537	0.0004365148
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	0.0003851077	0.0003236679
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	0.0004329692	0.0004030368
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	0.0003670738	0.0003549008
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	0.0003249204	0.0002944627
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	0.0002964434	0.0003404564
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	0.0003221655
5.1	0.2	50	0.4862388	0.2987405	0.003476443	0.002784509
5.1	0.2	100	0.4871271	0.3011671	0.001435091	0.001484796
5.1	0.2	500	0.4897375	0.3021698	0.0003097326	0.0003086277

Table 4: Summary of simulation results verifying the covariance expression (8) for $cov(\hat{S}_T(1), \hat{S}_T(2))$. The true values $S_T(1) = 0.4882$ and $S_T(2) = 0.3$.