Harvard University

Harvard University Biostatistics Working Paper Series

Year 2003 *Paper* 9

A Nonparametric Comparison of Conditional Distributions with Nonnegligible Cure Fractions

Yi Li* Jin Feng†

*Harvard University, yili@jimmy.harvard.edu

†University of Massachusetts, feng@math.umass.edu

A Nonparametric Comparison of Conditional Distributions with Nonnegligible Cure Fractions

BY YI LI
Department of Biostatistics
Harvard School of Public Health and Dana-Farber Cancer Institute
Boston, MA 02115, U.S.A.
yili@jimmy.harvard.edu

AND JIN FENG
Department of Mathematics and Statistics
University of Massachusetts, Amherst, MA 01002, U.S.A.
feng@math.umass.edu

Abstract

Survival data with nonnegligible cure fractions are commonly encountered in clinical cancer clinical research. Recently, several authors (e.g. Kuk and Chen, 1992; Maller and Zhou, 1993; Peng and Dear, 2000; Sy and Taylor, 2000) have proposed to use semiparametric cure models to analyze such data. Much of the existing work has been emphasized on cure detections and regression techniques. In contrast, this project focuses on the hypothesis testing in the presence of a cure fraction. Specifically, our interest lies in detecting whether there exists survival differences among non-cured patients between treatment arms. For this purpose, we investigate the use of a modified Cramér-von Mises statistic for two-sample survival comparisons within the framework of cure models. **Such a test has been studied by Tamura et al. (2000) using a bootstrap procedure. We will focus on developing asymptotic theory and convergent algorithms in this paper.** We show that the limiting distributions of the Cramér-von Mises statistic under the null hypothesis can be represented by stochastic integrals and a weighted noncentral chi-squares. Both representations lead to concrete numerical schemes for computing the limiting distributions. The algorithms can be easily implemented for data analysis **and significantly reduce computing time compared to the bootstrap approach**. For illustrative purposes, we apply the proposed test to a published clinical trial.

KEY WORDS: Cure Model; Cramér-von Mises Statistics; Asymptotic Theory.

RUNNING TITLE: Nonparametric Comparison of Cure Models

1 Introduction

In cancer trials the idea of "cure from cancer" is that the disease would be completely eliminated so that it never recurs, and the patient's lifespan is the same as that of someone who has never suffered cancer. Treatments are typically developed to increase patients' chances of being cured, but considerable interest has also been given to pursue treatments that prolong survival among non-cured patients as well. Because the study population is essentially an unobservable mixture of patients deemed curable and non-curable, evaluation of treatment effects in such a scenario is often complicated. A recently published study is presented below as an illustrative example.

> Between 1992 and 1999, a phase III clinical trial was conducted by Adelstein et al. (2003) of the Eastern Cooperative Oncology Group (ECOG) to evaluate treatment effects in patients with unresectable head and neck cancer. Patients were randomized among the following treatment arms: a standard single daily fractionated radiotherapy (control arm), and a split course of single daily fractionated radiotherapy and chemotherapy (experimental arm). One primary endpoint was whether these treatments had survival benefit in terms of preventing death from head and neck cancer. In this trial, a number of long term survivors have been observed. But, in addition to the comparison of the cure rates, the investigators were also interested in evaluating the treatment effect in terms of survival among patients who were not cured.

The concept of cure brings new clinical interests as well as statistical challenges. For instance, clinical objectives are not only focused on the comparison of unconditional distributions of the time to a medical event of interest (e.g. death), but also on that of the conditional distributions within non-cured patients (e.g. Berkson and Gage, 1952; Farewell, 1982; Greenhouse and Wolfe, 1984; Gray and Tsiatis, 1989; Laska and Meisner, 1992). Meanwhile, other characteristic problems in survival analysis need to be addressed. For example, random accrual and patients' dropout or loss to follow up are to be modeled through random right censoring processes. Several authors (e.g. Kuk and Chen, 1992; Peng and Dear, 2000; Sy and

Taylor, 2000) have proposed using semiparametric cure models to analyze such data, with emphasis placed on regression modeling. A complete review of the statistical methods using censored failure time to determine the presence of a cure fraction can be seen in Maller and Zhou (1992, 1993).

In this article, we will take a different perspective by focusing on the characterization of the conditional distribution among non-cured individuals. That is, we are interested in studying the distribution of the time-to-event variables given the event (e.g. disease-related death) will occur before a clinically meaningful terminal time τ. Explanations of these types of models from medical viewpoints, together with some available statistical methods (e.g. likelihood ratio tests and rank tests) can be found in, for example, Laska and Meisner (1992) and Gray and Tsiatis (1989), and the likelihood ratio test has been used to analyze a clinical trial (Laska, Siegel and Sunshine, 1991).

Indeed, the past two decades has seen a rapid development of statistical tools for detecting survival differences in clinical trials, but log-rank type tests, most powerful under proportional hazards alternatives, are routinely performed by practitioners in the absence of cure fractions. Schumacher (1984), however, demonstrated via simulation that Cramér-von Mises statistics have decent power under proportional hazards alternatives and are superior to log-rank tests in other cases. **A Cramér-von Mises type statistic was proposed by Tamura, Faries, and Feng (2000) for a two-sample survival comparison within the framework of cure models, and inference was drawn via bootstrap simulations.** On the other hand, in the absence of cure fractions the asymptotic theory of the Cramér-von Mises type statistic for one-sample comparisons under particular parametric models has been intensively investigated by Koziol and Green (1976) (with censoring) and Stute (1997) (without censoring). This article investigates the use of the Cramér-von Mises statistic for two-sample comparisons in the presence of censoring and cure fractions. Moreover, we develop the asymptotic theory in a more general framework: unlike in Koziol and Green (1976), our procedure does not require the specification of the censoring distributions. We also give concrete numerical algorithms for practical applications without resorting to resampling schemes. Though both deal with survival comparisons for cure models, our motivation differs from that of Gray and Tsiatis (1989)

in that their main interest was in testing differences in cure rates, while we are primarily interested in testing differences in survival among non-cured patients. **In contrast to Tamura et al (2000), we further the asymptotic theory: not only do we give two characterizations of the limit distribution in detail (one of which appeared sketchily in Tamura et al. (2000) as well), but also we construct practically usable algorithms based on them. As our simulation results (in Section 6) indicate, the proposed algorithms significantly reduce computing time compared to the bootstrap procedure. In addition, the involved techniques may have applications elsewhere, for example, numerically performing Loève principle component decomposition of a complex stochastic process.**

The rest of the article is structured as follows. We state a general cure model in Section 2, and formulate a two-sample comparison hypothesis in Section 3. We introduce in Section 4 a modified version of Cramér-von Mises statistic, and derive its large sample properties, and give in Section 5 two numerical simulation schemes for constructing the limiting distributions of the Cramér-von Mises statistics. We conduct simulation studies in Section 6 to examine the finite sample performance of the proposed two algorithms, and illustrate the use of the Cramér-von Mises statistics by analyzing a published clinical trial in head and neck cancer in Section 7. We conclude this article with general discussion in Section 8. In the Appendix, we give the technical proofs to the main theorems, and include a convergence result on stochastic integrals, which has largely facilitated our proofs.

Throughout this article, $F(t) = P(T \leq t)$ denotes the distribution for a nonnegative random variable T, $X_n \Rightarrow X$ means random variables X_n converge to X in distribution, $X_n \to X$ means the convergence is in probability, $X \stackrel{D}{=} Y$ means random variables X and Y are equally distributed, and for any real function $g(\cdot)$, $g(\infty-)$ means $\lim_{x\to\infty, x\neq\infty} g(x)$ if the limit exists.

2 A General Cure Model

Let T be a nonnegative random variable of interest, for example, the time from the start of treatment to disease-related death in our motivating example. We assume that its distribution on the extended real line $[0, +\infty]$, where $+\infty$ is identified to be a single point, is given by

$$F(dt) \equiv P\{T \in (t, t+dt]\} = pf(t)dt + q\delta_\infty(dt), \quad p, q > 0, p+q = 1,$$

where $f \geq 0, \int_0^\infty f(t)dt = 1$, q (or p) is the cure (or non-cure) proportion, and for any $a \in [0, +\infty], \delta_a(\cdot)$ is a counting measure with a point mass on a. This is equivalent to saying that T follows a mixed distribution with a dominating measure consisting of a Lebesgue measure and a singular point mass placed at $+\infty$. In reality it would be a reasonable assumption that the probability density f, corresponding to the non-cure proportion of T, have a compact support $[0, \tau] \subset [0, \infty)$, and

$$\tau = \inf\{t \geq 0 : \sup_{s \geq t} f(s) = 0\} < \infty. \tag{1}$$

For instance, if a disease-related death did not occur to a patient after a long time of observation, say roughly, 5 years in a head and neck cancer clinical trial, it would be "safe" to predict that patient would not die from this cancer. Similar formulations for this type of cure models can be found in Maller and Zhou (1992, 1993) and Peng and Dear (2000), among others.

We can either directly identify the τ, the clinically meaningful terminal time, with available medical knowledge or estimate it from empirical data. Specifically, denote by $\hat{F}$ the Kaplan-Meier estimate for the distribution function F and note that (1) is equal to

$$\tau = \inf\{t \geq 0 : F(t) = F(\infty-)\}, \tag{2}$$

then τ can be consistently estimated by

$$\hat{\tau} = \inf\{t > 0 : \hat{F}(t) = \hat{F}(\infty-)\}, \tag{3}$$

which is the turning point after which the estimated distribution curve becomes plateau, and is indeed the largest uncensored failure time (Maller and Zhou, 1992). Elementary probabilistic

arguments immediately imply that $\hat{\tau} \leq \tau$ almost surely. Hence, by (3) and the monotonicity of $\hat{F}$,

$$\hat{F}(\hat{\tau}) = \hat{F}(\tau) = \hat{F}(\infty-)$$

almost surely.

Both (1) and (3) indicate that τ can be identified and estimated from F, as a function of f and p. Additionally, since estimating and testing f or p is of major interest to us, we will focus only on these two parameters (and not τ) in this article; a detailed discussion on τ and its estimation can be found in Maller and Zhou (1992).

3 Two-Sample Comparison of Cure Models

Large-scale clinical trials, for example, phase III trials, are often designed to detect survival benefits among competitive regimen, and two-sample comparisons are frequently utilized. In what follows, we study the Cramér-von Mises type statistic for comparing two treatments in a clinical trial.

The notation below is similar to that in the general cure model, except that we use an additional subscript i to indicate treatment. Specifically, we denote the time-to-event variables by $T_{ij} \sim F_i$, $i = 1, 2; j = 1, \ldots, n_i$, where, for example, $i = 1$ corresponds to the control arm and $i = 2$ to the experimental arm, F_is are distribution functions, and j refers to the j-th patient in his respective treatment arm; we also assume the nonnegative censoring times $U_{ij} \sim G_i$ where G_is are distribution functions. We further assume that the $\{T_{ij}, U_{ij} : i = 1, 2, j = 1, \ldots, n_i\}$ are all independent. Because of censoring, we observe $V_{ij} = T_{ij} \wedge U_{ij}$ and $\delta_{ij} = I(T_{ij} \leq U_{ij})$. If our interest were in estimating and comparing the $F_i, (i = 1, 2)$, we would use standard methods for right censored data, such as the Kaplan-Meier estimators or the log-rank type statistics. However, the distributions of more direct interest are two conditional distributions $F_i^*(t) = P(T_{ij} \leq t | T_{ij} < \infty), i = 1, 2$. To explore how to compare these conditional distributions, first let

$$F_i(t) \equiv P(T_{ij} \leq t) = p_i \int_0^t f_i(s)ds + q_i I(t = \infty), \quad p_i + q_i = 1, 0 < p_i, q_i < 1.$$

Then,

$$F_i^*(t) = \int_0^t f_i(s)ds = p_i^{-1}F_i(t).$$

For a two-sample comparison, the statistical test is formulated as

$$H_0 : F_1^* = F_2^* \text{ vs } H_1 : F_1^* \neq F_2^*. \tag{4}$$

We assume both f_is have compact support $\tau_i \equiv \inf\{t \geq 0 : \sup_{s \geq t} f_i(s) = 0\} < \infty$, and denote $\tau = \max\{\tau_1, \tau_2\}$ such that $[0, \tau]$ covers the supports of both f_1 and f_2. Moreover, we assume that

$$P(U_{11} > \tau) > 0, \qquad P(U_{21} > \tau) > 0, \tag{5}$$

so that τ can be observed with a positive probability in both treatment arms.

As is the case without cure fractions, if it is reasonable to believe that $F_1^*(t) \leq F_2^*(t)$ (or vice versa), H_1 should be replaced by a one-sided hypothesis. Furthermore, if the proportional hazard assumption is approximately satisfied, a log-rank type statistic, powerful in detecting the stochastic dominance, should perform better than its Cramér-von Mises counterpart. However, without such a proportionality assumption, the Cramér-von Mises type statistic, which is particularly useful in detecting the deviation of two distributions, is more appropriate (see, e.g. Schumacher, 1984).

4 Cramér-von Mises Statistic and its Asymptotics Properties

Let n_1 and n_2 be the sample size of two arms, respectively, and $n = n_1 + n_2$. Following Laska and Meisner (1992), for each $i = 1, 2$, we derive a nonparametric maximum likelihood estimate of F_i^* based on the observed data $(V_{ij}, \delta_{ij}), j = 1, \ldots, n$, as

$$\hat{F}_i^*(t) = \hat{p_i}^{-1}\hat{F}_i(t),$$

where $\hat{F}_i(t)$ is the usual Kaplan-Meier estimator for the F_i and $\hat{p_i} = \hat{F}_i(\infty-)$ is the consistent estimate for p_i, the estimated non-cure fraction in the i-th arm (see, Maller and Zhou, 1992).

Denote the pooled conditional distribution by

$$\hat{F}^*_{pool} = \frac{n_1\hat{p}_1\hat{F}^*_1 + n_2\hat{p}_2\hat{F}^*_2}{n_1\hat{p}_1 + n_2\hat{p}_2}.$$

To test H_0 in (4), we define a modified version of Cramér-von Mises statistic to measure the discrepancy between the two empirical distributions $\hat{F}^*_1(\cdot)$ and $\hat{F}^*_2(\cdot)$ as follows

$$W_n = n\int_0^\infty \{\hat{F}^*_1(t-) - \hat{F}^*_2(t-)\}^2 d\hat{F}^*_{pool}(t). \tag{6}$$

The following theorem gives the asymptotic distribution of W_n, under $H_0 : F^*_1 = F^*_2 = F^*$.

Theorem 1 *Assume that $n_1/n \to \gamma$. Then under the null hypothesis, the Cramér-von Mises statistic,*

$$W_n \Rightarrow X = \int_0^\infty G^2(t-)dF^*(t),$$

where the Gaussian process G is (distributionally) uniquely defined by

$$\begin{aligned} G(t) \;=\; & \frac{1}{\sqrt{\gamma}}\Big[\frac{1-p_1F^*(t)}{p_1}W_1\{c_1(t)\} - F^*(t)\frac{q_1}{p_1}W_1\{c_1(\infty-)\}\Big] \\ & -\frac{1}{\sqrt{1-\gamma}}\Big[\frac{1-p_2F^*(t)}{p_2}W_2\{c_2(t)\} - F^*(t)\frac{q_2}{p_2}W_2\{c_2(\infty-)\}\Big], \end{aligned}$$

$W_1(\cdot)$ and $W_2(\cdot)$ are independent Brownian motions and

$$c_i(t) = \int_0^t \frac{p_i dF^*(s)}{\pi_i(s)\{1-p_iF^*(s-)\}}.$$

Here $\pi_i(s) = P(V_{ij} \geq s) = \{1-G_i(s-)\}\{1-p_iF^*(s-)\}$ for $i = 1, 2$.

Remark: The validity of this theorem requires, $c_i(t), i = 1, 2$, the changes of time in the Gaussian processes, be finite over $[0, \infty)$. For each $i = 1, 2$, because $c_i(t)$ is non-decreasing we only need to show the finiteness of $c_i(\infty-)$. In fact, if $\tau < \infty$ as assumed and under (5), $c_i(\infty-) = c_i(\tau_i) < \infty$.

The proof relies on large sample results of the Kaplan-Meier product limit estimator and **can be found in Tamura et al. (2000).** Asymptotic results of the Kaplan-Meier estimates have been proved by Breslow and Crowley (1974), Gill (1984), and Fleming and Harrington (1991) in various degrees of generalization.

We note in this theorem the null distribution F^* is left unspecified, but it can be replaced by its empirical estimate $\hat{F}^*_{pool}$ when approximating the distribution of X under the null hypothesis; see Algorithm 1 and the associated large sample result (Theorem 4) in Section 5.

For gaining additional insight into the limiting distribution of the Cramér-von Mises statistics under H_0, we consider a Loève type expansion in terms of principal components. Specifically, we represent the X in Theorem 1 as a mixture of noncentral χ^2, which would facilitate numerical realizations. Similar results in the context of non-censored data or censored data without cured fraction are available, see for example Durbin et al. (1972, 1975) and Chapter 5 of Shorack and Wellner (1986). With some modifications, the arguments in these references would apply in the new context of cure models for establishing the existence of such decompositions. However, as is frequently the case in practice, the difficulty for making use of these results lies in the computation of the associated eigenvalues (e.g. the λ_ks below). By focusing on some special classes of parameterized distributions (of both survival and censoring times), the aforementioned references have obtained closed-form eigenvalues. But, in more general scenarios (e.g. censored data with a cure fraction as we have) we would not expect such explicit formulae to exist. Therefore, we approach this issue from a numerical perspective by deriving the kernel function for the principal component decomposition in more general situations, where the distribution functions for both survival and censoring times are left unspecified. Based on this decomposition, a numerical algorithm is developed in Section 5 for computing the associated eigenvalues.

By exploiting the independence of $W_1(\cdot)$ and $W_2(\cdot)$, we can compute the covariance function $K(s,t)$ of the Gaussian process $G(\cdot)$. Specifically,

$$\begin{aligned} K(s,t) &= E\{G(s)G(t)\} = \gamma^{-1}\{a_1(s,t)c_1(s\wedge t) + b_1(s,t)c_1(\tau) - d_1(s,t)c_1(t) - d_1(t,s)c_1(s)\} \\ &\quad +(1-\gamma)^{-1}\{a_2(s,t)c_2(s\wedge t) + b_2(s,t)c_2(\tau) - d_2(s,t)c_2(t) - d_2(t,s)c_2(s)\} \end{aligned} \tag{7}$$

where

$$a_i(s,t) = p_i^{-2}\{1-p_iF^*(t)\}\{1-p_iF^*(s)\}, b_i(s,t) = p_i^{-2}q_i^2F^*(t)F^*(s), d_i(s,t) = p_i^{-2}q_i\{1-p_iF^*(t)\}F^*(s)$$

for $i = 1, 2$. The result is summarized in the following theorem, whose proof can be found in the Appendix.

Theorem 2 *The distribution for the limiting random variable X in Theorem 1 can be represented as the following noncentral χ^2*

$$X \stackrel{D}{=} \sum_{k=1}^{\infty} \lambda_k Z_k^2 \tag{8}$$

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 $\Big(L^2([0,\infty]),(\cdot,\cdot)\Big)$ 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 the λ_k are decreasing in k to zero.

After identifying elements by their equivalence classes, $\Big(L^2([0,\infty]),(\cdot,\cdot)\Big)=\Big(L^2([0,\tau]),(\cdot,\cdot)\Big)$. In practice, unless some strong parametric assumptions are made, numerical computation of the λ_ks will typically involve in matrix approximation of the integral operator $\mathcal{T}$, which has F^* as an integrator. A direct discretization for $\mathcal{T}$ on interval $[0,\tau]$ with uniform mesh size may cause numerical instability. Through a change of variables, we can convert the eigenvalue problem for $\mathcal{T}$ in the space $\Big(L^2([0,\tau]),(\cdot,\cdot)\Big)$ to an equivalent eigenvalue problem for operator $\tilde{\mathcal{T}}$ in the Hilbert space on the unit interval $\Big(L^2([0,1]),\langle\cdot,\cdot\rangle\Big)$, wherein the inner product is $\langle\tilde{f},\tilde{g}\rangle=\int_0^1\tilde{f}(r)\tilde{g}(r)dr$ for $\tilde{f},\tilde{g}\in L^2([0,1])$. We give in the Appendix the explicit form of $\tilde{\mathcal{T}}$ and the justification that these two operators have the same set of eigenvalues on each individual Hilbert space. Hence, for numerical stability, our numerical scheme (outlined in Section 5) will be based on the transformed operator $\tilde{\mathcal{T}}$. In particular, assuming that F^* is strictly increasing and absolutely continuous and letting

$$(\tilde{\mathcal{T}}\tilde{g})(\tilde{t})=\int_0^1\tilde{K}(\tilde{s},\tilde{t})\tilde{g}(\tilde{s})d\tilde{s}$$

for any $\tilde{g}\in L^2([0,1])$, where $\tilde{K}$ is as defined in (14), one may show in the Appendix that the λ_k are also eigenvalues for the integral operator $\tilde{\mathcal{T}}$. The λ_k's can be computed more conveniently from operator $\tilde{\mathcal{T}}$ than from operator $\mathcal{T}$ directly.

The representation of X in (8) implies that the shape of distribution should be similar to that of the χ^2. We confirm this by a numerical approximation in our data example later.

Koziol and Green (1976) considered a very special case (in the absence of cure fractions) with $n_1 = n_2, p_1 = p_2 = 1, F(t) = F^*(t)$ being uniform on $[0,1]$ and $1 - G_1(t) = 1 - G_2(t) = (1-t)^\beta, \beta < 2$. In such a case $c_1(t) = c_2(t) = 1 - (1-t)^{-(1+\beta)}$ and the covariance function $K(s,t)$ has a simple form: $K(s,t) = 4(1-s)(1-t)\{1-(1-s\wedge t)^{-(1+\beta)}\}$. They computed the λ_i explicitly and gave some useful reference tables.

We will develop an approximation algorithm (Algorithm 2 in Section 5) to calculate these eigenvalues for more general and practical settings, where the null distribution F^* and the censoring distributions G_i are unspecified, and, hence, functions such as $\pi_i(s)$ and $c_i(s)$ in the limiting distribution are unknown. In particular, we consistently estimate these unknown quantities by

$$\hat{\pi}_i(s) = 1 - \frac{1}{n_i}\sum_{j=1}^{n_i} I(V_{ij} < s), \tag{9}$$

and

$$\hat{c}_i(t) = \int_0^t \frac{d\hat{F}_i(s)}{\hat{\pi}_i(s-)\{1-\hat{F}_i(s-)\}}, \tag{10}$$

for $i = 1, 2$ and use them to replace π_i, c_i in the kernels K and $\tilde{K}$. We justify the use of (9) and (10) with the following theorem, whose proof is given in the Appendix.

Theorem 3 *$\hat{c}_i$ consistently estimates c_i. That is, $\hat{c}_i \to c_i$ in probability on $D[0,\tau]$ and this convergence is in conjunction with that of $\hat{p}_i$ to p_i and $\hat{F}_i(t)$ to $F_i(t)$.*

5 Numerical Approximation/ Simulation Schemes

In this section we develop two "data driven" numerical schemes to construct random variables that approximate the limiting distribution of the Cramér-von Mises statistics. We start with estimating $\tau \equiv \max\{\tau_1, \tau_2\}$ by $\hat{\tau} = \max\{\hat{\tau}_1, \hat{\tau}_2\}$, where $\hat{\tau}_i = \inf\{t \geq 0 : \hat{F}_i(t) = \hat{F}_i(\infty-)\}$. We first present a numerical algorithm based on the stochastic integral representation theorem (Theorem 1) (see also Remark 2 of Tamura et al. (2000)), followed by a numerical scheme based on the principal decomposition theorem (Theorem 2). The associated limit theorems are also provided.

Algorithm 1: An approximation scheme based on Theorem 1

1. Compute statistics $\hat{\pi}_i$ in (9), $\hat{c}_i$ in (10), $\hat{p}_i$, for $i = 1, 2$, and $\hat{F}^*_{pool}$.

2. For each positive integer m, generate i.i.d standard normal random variables $\xi_1, \ldots, \xi_m \sim N(0,1)$ and $\eta_1, \ldots, \eta_m \sim N(0,1)$; define

$$W_1^{m,n}(t) = \sum_{k=1}^{[tm/\hat{\tau}]} \{\hat{c}_1(k\hat{\tau}/m) - \hat{c}_1((k-1)\hat{\tau}/m)\}^{1/2}\xi_k$$

and

$$W_2^{m,n}(t) = \sum_{k=1}^{[tm/\hat{\tau}]} \{\hat{c}_2(k\hat{\tau}/m) - \hat{c}_2((k-1)\hat{\tau}/m)\}^{1/2}\eta_k;$$

3. Let

$$\begin{aligned} G_{m,n}(t) &= \frac{1}{\sqrt{\gamma}}\Big\{\frac{1-\hat{p}_1\hat{F}^*_{pool}(t)}{\hat{p}_1}W_1^{m,n}(t) - \hat{F}^*_{pool}(t)\frac{\hat{q}_1}{\hat{p}_1}W_1^{m,n}(\hat{\tau})\Big\} \\ &\quad -\frac{1}{\sqrt{1-\gamma}}\Big\{\frac{1-\hat{p}_2\hat{F}^*_{pool}(t)}{\hat{p}_2}W_2^{m,n}(t) - \hat{F}^*_{pool}(t)\frac{\hat{q}_2}{\hat{p}_2}W_2^{m,n}(\hat{\tau})\Big\} \end{aligned}$$

4. Finally, compute

$$X_{m,n} = \int_0^{\hat{\tau}} G^2_{m,n}(t-)d\hat{F}^*_{pool}(t).$$

Note that m is the size of the approximation scheme and n is the size of the real clinical trial dataset. Since every integral appearing in the above scheme has a piece-wise constant integrand, these integrals are effectively finite summations. **In practice, we shall repeatedly apply this algorithm to obtain a series of independent realizations of $X_{m,n}$ in order to approximate the asymptotic distribution.** Indeed, the following theorem, proved in the Appendix, shows the convergence in distribution of $X_{m,n}$ to the desired limit.

Theorem 4 *Under the conditions stated in Theorem 1, the $X_{m,n}$ defined above satisfies $X_{m,n} \Rightarrow X$ as $m, n \to \infty$, where X is the limiting random variable in Theorem 1.*

Algorithm 2: An approximation scheme based on Theorem 2

In this approximation scheme, we again use m to denote the size of the approximation scheme and n the sample size of the dataset. The larger these parameters, the better for the approximant $X_{m,n}$ to approach the limiting X.

1. Compute statistics $\hat{\pi}_i$ in (9), $\hat{c}_i$ in (10), $\hat{p}_i$ for $i = 1, 2$, and $\hat{F}^*_{pool}$.

2. For each $0 \leq \tilde{s}, \tilde{t} \leq 1$, define the empirical version of $\tilde{K}(\tilde{s}, \tilde{t})$ (given in (14)) by

$$\hat{\tilde{K}}(\tilde{s}, \tilde{t}) = \hat{K}\Big\{(\hat{F}^*_{pool})^{-1}(\tilde{s}), (\hat{F}^*_{pool})^{-1}(\tilde{t})\Big\},$$

   where $\hat{K}(s, t)$, the empirical version of $K(s, t)$, is obtained by replacing all the unknown quantities in (7) by their estimates. As the estimator $\hat{F}^*_{pool}$ is piecewise constant, we interpret its inverse as $(\hat{F}^*_{pool})^{-1}(\tilde{t}) \equiv \inf\{t \geq 0 : \hat{F}^*_{pool}(t) \geq \tilde{t}\}$ for any $\tilde{t} \in [0, 1]$.

3. For a large positive integer m, construct a working matrix $A^{(m,n)} = (a_{u,v})_{m \times m}$ such that

$$a_{u,v} \equiv \frac{1}{m}\hat{\tilde{K}}(\frac{u}{m}, \frac{v}{m}) \text{ for } u, v = 1, \ldots, m.$$

4. Compute the eigenvalues of matrix $A^{(m,n)}$ and rank them from the largest to the smallest to obtain $\lambda_1^{(m,n)} \geq \ldots \geq \lambda_m^{(m,n)}$.

5. Select the first l eigenvalues and let

$$X_{m,n} = \sum_{k=1}^{l \wedge m} \lambda_k^{(m,n)} Z_k^2$$

   where $Z_1, \cdots, Z_l$ are i.i.d. standard normal random variables. Here, $l = \inf\{k : \lambda_k^{(m,n)}/\lambda_1^{(m,n)} \leq \epsilon\}$ where ϵ is a prespecified small constant controlling the accuracy of approximation.

Since $A^{(m,n)}$ is a random matrix, the $\lambda_k^{(m,n)}$s are random quantities. The following theorem gives the asymptotic results concerning the $\lambda_k^{(m,n)}$s (see the appendix for the detailed proof).

Theorem 5 *Under the conditions stated in Theorem 1,* $\lim_{m,n \to +\infty} \lambda_k^{(m,n)} = \lambda_k$ *in probability, for each* $k = 1, 2, \ldots$, *where the* λ_k *are the eigenvalues in (8).*

6 Simulation Studies

Simulations were performed to examine the finite sample performance of the proposed test. Our main objectives are three folds, namely, assessment of the level of the test under varying

sample sizes and degree of censoring, evaluation of the power of the test under the same varying scenarios, and comparison of our two proposed asymptotical distribution-based algorithms with the bootstrap procedure adopted by Tamura et al. (2000).

We considered a similar simulation setup utilized by Tamura et al. (2000). Specifically, we set the noncure proportion $p_1 = 0.6$ in treatment group 1 (for example, the control arm) and generated the survival times for the noncured patients in this arm by the truncated Weibull distribution

$$1 - F_1^*(t) = [\exp\{-(t/\lambda)^\alpha\} - \exp\{-(43/\lambda)^\alpha\}/[1 - \exp\{-(43/\lambda)^\alpha], \tag{11}$$

where we set $\lambda = 20$ and $\alpha = 2$. This survival distribution would yield a median time to death of 16.6 time units with 90% of events occurring within 30 time units among noncured patients. On the other hand, we simulated the survival times for the noncured patients in treatment group 2 (e.g. the experiment arm) by

$$1 - F_2^*(t) = \{1 - F_1^*(t)\}^\beta,$$

where $\beta = 1$ corresponds to the null hypothesis and $\beta \neq 1$ corresponds to an alternative hypothesis. In our simulations we set $\beta = 1$ when estimating the level of the test and varied β to be 1.5, 2 and 2.5 when evaluating the power of the test under various alternative hypotheses. Two different proportions of noncure patients in Arm B, p_2, were considered in the simulations. We first let $p_2 = 0.6$, which was equal to its counterpart in Arm A. In the second case we set $p_2 = 0.9$, much higher than its counterpart in Arm B. The censoring times in both arms were independently simulated by a uniform distribution on $[0, c]$, with $c = 60$ inducing approximately 50% censoring prior to time 43 and $c = 80$ inducing approximately 35% censoring prior to time 43. At the opposite extreme, we also considered the case of no censoring prior to time 43. Same levels of censoring were adopted by Tamura et al. (2000). Finally, we considered an equal sample size of 100 per treatment group, which is typical in a two-sample clinical trial.

Table 1 lists the estimates of level and power of test under various conditions, all of which were based on 1000 data realizations. More specifically, for each simulated data set, the empirical p-values were computed based on the asymptotic distributions approximated by

Algorithms 1 and 2, respectively, and were later dichotomized by whether they fell below 0.05. When applying Algorithms 1 and 2, we chose the mesh size $m = 40$ and generated 1000 realized values of "$X_{m,n}$" to approximate the desired asymptotic distribution. We set $\epsilon = 0.001$ when using Algorithm 2. All programming was conducted in the environment of R and each entry in the table took approximately 2 hours on a mainframe computer.

Based on these simulation results we observed that the size of the test ranged 0.048-0.051 for no censoring, 0.060-0.063 for moderate censoring, and 0.075-0.089 for high censoring, indicating that the test is slightly liberal if censoring is moderate or less and is more considerably 'anticonservative' if censoring is high. Same observations have been documented for the bootstrap procedure by Tamura et al. (2000). We also noticed, as expected, that the power of the test increased rapidly as β increased from 1.5 to 2.5 and the power decreased as the amount of censoring increased. In particular, the powers obtained by the large sample approximation were slightly better than the powers obtained by the bootstrap procedure reported in Table II of Tamura et al. (2000). Last we noted that algorithms 1 and 2 yielded nearly identical sizes and powers.

7 A Data Example

We applied the Cramér-von Mises statistic to analyze the cancer clinical trial described previously in Section 1. Originally there were three treatment arms in this study. For simplicity, we only considered the comparison of disease-specific survival between two treatment arms in this clinical trial. In particular, the disease-specific survival was measured for a total of 184 patients, 95 in the control arm and 89 in the experimental arm. Disease-specific survival is defined as the time from the start of treatment to the death caused by the disease. Patients dying from any other causes (for example, traffic accident or suicide) would be regarded as being censored at the death dates (see Adelstein et al., 2003).

The top panel of Figure 1 gives the comparison in disease-specific survival by treatment among all these 184 patients, including those who were cured and non-cured, while the bottom panel shows the conditional time to disease-specific death among noncured patients.

We estimated from the data that the cured patients were those who had survived from head and neck cancer for more than 4.2 years, i.e. $\hat{\tau} = 4.2$. Using the Kaplan-Meier estimate, we also estimated that the cure rate in the control arm was $\hat{q}_1 = 1 - \hat{F}_1(\infty-) = 0.25$ with a 95% confidence interval $(0.13, 0.37)$, while the cure rate in the experimental arm was $\hat{q}_2 = 1 - \hat{F}_2(\infty-) = 0.41$ with a 95% interval $(0.29, 0.53)$. The Wald test for equal cure rates between the two arms, i.e. $H_0 : q_1 = q_2$, is marginally significant (p-value=0.056), favoring the experimental arm. However, our interest was whether the distribution of disease-specific survival differs between the two treatment arms among non-cured patients. We applied the two-sample Cramér-von Mises statistic to test the null hypothesis that the two treatment arms have the same survival function among the non-cured patients. We calculated the limiting distribution of the Cramér-von Mises statistic under the null hypothesis by exploiting the numerical algorithms (Algorithms 1 and 2), where we set $m = 40$ and $n = 184$. The resulting density curves of the limit distribution obtained by both algorithms are displayed in Figure 2. The two curves overlap for the most part and resemble a χ^2 distribution. The estimated quantiles of the distributions by these two algorithms are listed in Table 2 and the leading eigenvalues of the linear operator in Algorithm 2 are tabulated in Table 3. Based on (6), the resulting test statistic is 1.16 with a p-value of 0.356 according to the limiting distribution obtained by Algorithm 1 and 0.362 by Algorithm 2, indicating there is no strong evidence for a significant difference in disease-specific survival between the two treatment arms among the non-cured patients, though the cure rate in the treatment arm is marginally superior to that in the control arm.

8 Discussion

In this article, we have developed large sample results and given concrete numerical schemes for implementing a modified Cramér-Von Mises statistic for two-sample comparisons of conditional survival curves in the presence of cure fractions. We second Tamura et al.'s (2000) opinion that the cure model is a proper model which separates the survival information into the proportion of the cured population, and the time to event conditional on being non-cured patients. Both parts of information are important when evaluating treatment effects in clinical

trials. Our motivation stems from comparisons of conditional survival curves, given that the tests of cure rates have been well documented in statistical literature (e.g. Gray and Tsiatis, 1989).

Our estimators and numerical schemes can be easily implemented and are fully nonparametric, which is in contrast to the statistics proposed by Koziol and Green (1976) for one-sample comparisons in the absence of cure fractions and under very specific parametric models of censorship. On the other hand, as opposed to the bootstrap-based inferential methods proposed by Tamura et al. (2000), our work directly employs large sample results and does not require any complicated resampling schemes in data analysis, thereby significantly reducing the computational burden.

Future research is needed for studying the behavior of the proposed Cramér-von Mises statistics under the alternative hypothesis, which might facilitate the design of a comparative clinical trial in a cure rate model. This will also enable one to compare the efficiency between the Cramér-von Mises statistics and other commonly used statistics in survival problems (e.g. log rank tests), under a variety of alternatives.

Related Splus or R programs for computing the proposed Cramér-von Mises test statistic and its distribution under the null hypothesis are available upon request.

Acknowledgement

The authors thank the Editor-in-Chief and two anonymous referees for their helpful comments. The authors also owe thanks to Dr. Roy N. Tamura, who has carefully read this revised manuscript and provided many insightful suggestions.

Appendix: Technical Details

A Convergence Theorem on Stochastic Integrals

The following is extracted from Theorem 2.2 in Kurtz and Protter (1991), of which we have made repeated usage in our proofs.

Theorem 6 *For each n, let (X_n, Y_n) be an $\{\mathcal{F}_t^n\}$-adapted process with paths in $D_{R^2}[0,\tau]$. Let $Y_n = M_n + A_n$, where $M_n(t)$ is an $\{\mathcal{F}_t^n\}$-martingale and $A_n(t)$ be a process with finite variation. Suppose for each t,*

$$\sup_n E\Big\{[M_n](t) + T_t(A_n)\Big\} < \infty, \tag{12}$$

where $T_t(A_n)$ denotes the total variation of A_n up to time t and $[M](t)$ denotes the quadratic (or called optional) variation process for a local martingale $M(\cdot)$ (see, e.g. Andersen et al., 1993, p.69).

Suppose $(X_n, Y_n) \Rightarrow (X, Y)$ in the Skorohod topology on $D_{R^2}[0,T]$, then Y is a semimartingale with respect to a filtration to which (X, Y) are adapted and

$$\left\{X_n(t), Y_n(t), \int_0^t X_n(s-)dY_n(s)\right\} \Rightarrow \left\{X(t), Y(t), \int_0^t X(s-)dY(s)\right\} \tag{13}$$

in the Skorohod topology. If $(X_n, Y_n) \to (X, Y)$ in probability, then (13) also converges in probability.

Proof of Theorem 2

From (7), $0 \le K(s,t) = K(t,s) \le \sup_{0\le s,t\le\tau} K(s,t) < \infty$. Hence, $\mathcal{T}$ is self-adjoint by the symmetry of K; $\mathcal{T}$ is positive since it maps nonnegative functions to nonnegative functions; $\mathcal{T}$ is compact because K is square integrable. Specifically, $\int_0^\tau \int_0^\tau K^2(s,t)F(ds)F(dt) < \infty$ (in fact, K is even bounded). Therefore, $\mathcal{T}$ maps a set of uniformly bounded sequence of functions to a compact sequence of functions in the Hilbert space (Dunford and Schwartz, 1958). Hence, spectrum theory for linear compact operators implies that

$$K(s,t) = \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t)$$

almost everywhere by the product measure $F^*(ds) \times F^*(dt)$, where $\{\lambda_k \geq 0\} \in \ell^2$ and $\{f_k(\cdot)\}$ is the orthonormal basis in $L^2([0,\tau])$. Consequently, define a Gaussian process $H(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} f_k(t) Z_k$ so that

$$E\Big\{H(t)H(s)\Big\} = \sum_{k=1}^{\infty} \lambda_k f_k(t) f_k(s = K(s,t) = E\Big\{G(t)G(s)\Big\}.$$

Therefore, the two Gaussian processes $H(\cdot)$ and $G(\cdot)$ agree in distribution, and

$$\begin{aligned} \int_0^\tau G^2(t-) dF^*(t) &\stackrel{D}{=} \int_0^\tau H^2(t-) dF^*(t) \\ &= \sum_i \sum_j \sqrt{\lambda_i \lambda_j} Z_i Z_j \int_0^\tau f_i(t) f_j(t) dF^*(t) \\ &= \sum_k \lambda_k Z_k^2. \end{aligned}$$

Since $\mathcal{T}$ is a self-adjoint compact positive linear integral operator, there are at most countably many $\lambda_k \geq 0$ and the only possible point of accumulation for λ_k is 0 (see, e.g. Dunford and Schwartz, 1958). Without loss of generality, we may assume $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. □

Proof of Theorem 3

The Glivenko-Cantelli theorem implies that the empirical distribution $\hat{\pi}_i(s) \to P(V_{ij} \geq s)$ uniformly on $[0,\tau]$ and this convergence is joint with the convergence in probability of $\hat{F}_i$ to F_i. Thus, a direct application of Theorem 6 yields the result. □

Proof of Theorem 4

Joint with the convergence in probability (in the Skorohod topology) of $\hat{p}_i \to p_i$ and $\hat{F}^*_{pool} \to F^*$, and because c_i is continuous, then $W_i^{m,n}(t) \Rightarrow W_i\{c_i(t)\}$. This result follows from the martingale central limit theorem. Therefore, jointly we have $G_{m,n}(t) \Rightarrow G(t)$.

Note that $\hat{F}^*_{pool}(t)$ satisfies assumption (12). Applying Theorem 6, $X_{m,n} \Rightarrow X$. □

Justification of $\mathcal{T}$ and $\tilde{\mathcal{T}}$ Having Same Eigenvalues in Section 3

For simplicity, we assume that $f(t) > 0$ whenever $t \in [0, \tau]$. Therefore, F^* is strictly increasing and absolutely continuous on $[0, \tau]$. Define a one-to-one correspondence

$$F^*(s) = \tilde{s}, F^*(t) = \tilde{t}, \quad \Big(L^2([0,1]), \langle \cdot, \cdot \rangle\Big) \ni \tilde{f}(\tilde{s}) = f(s) \in \Big(L^2([0,\tau], (\cdot,\cdot))\Big),$$

and define

$$\tilde{K}(\tilde{s}, \tilde{t}) = K(s,t) = K\Big((F^*)^{-1}(\tilde{s}), (F^*)^{-1}(\tilde{t})\Big). \tag{14}$$

Let $(\tilde{\mathcal{T}}\tilde{f})(\tilde{t}) = \int_0^1 \tilde{K}(\tilde{s}, \tilde{t})\tilde{f}(\tilde{s})d\tilde{s}$. Then $\tilde{\mathcal{T}}$ is a compact positive linear integral operator on $\Big(L^2([0,1]), \langle \cdot, \cdot \rangle\Big)$ and the λ_ks are also the eigenvalues for $\tilde{\mathcal{T}}$. That is, $\lambda_k \tilde{f}_k = \tilde{\mathcal{T}}\tilde{f}_k$. □

Proof of Theorem 5

First note that the eigenvalues of the matrix $A^{(m,n)}$ in Algorithm 2 are equal to those of the (discretized) linear operator $\tilde{\mathcal{T}}_m$ in $L^2([0,1])$,

$$\tilde{\mathcal{T}}_m \tilde{f}(\tilde{t}) = \int_0^1 \hat{\tilde{K}}_m(\tilde{s}, \tilde{t})\tilde{f}(\tilde{s})d\tilde{s}, \quad \tilde{f} \in L^2([0,1]),$$

where

$$\hat{\tilde{K}}_m(\tilde{s}, \tilde{t}) = \hat{\tilde{K}}(\frac{u}{m}, \frac{v}{m}) \quad \text{if} \quad \frac{u-1}{m} \le \tilde{s} < \frac{u}{m}, \frac{v-1}{m} \le \tilde{t} < \frac{v}{m}$$

for some $u, v \in \{1, \ldots, m\}$.

By the Cauchy-Schwartz inequality, we bound the norm difference between linear operators $\tilde{\mathcal{T}}_m$ and $\tilde{\mathcal{T}}$ as follows:

$$\begin{aligned}
&\|\tilde{\mathcal{T}}_m - \tilde{\mathcal{T}}\| \equiv \sup_{\|\tilde{f}\|_{L^2}=1} \|(\tilde{\mathcal{T}}_m - \tilde{\mathcal{T}})\tilde{f}\|_{L^2} \\
\le\ & \Big(\int_0^1 \int_0^1 (\hat{\tilde{K}}_m - \tilde{K})^2(\tilde{s}, \tilde{t})d\tilde{s}d\tilde{t}\Big)^{1/2} \\
\le\ & \Big(\int_0^1 \int_0^1 (\hat{\tilde{K}}_m - \hat{\tilde{K}})^2(\tilde{s}, \tilde{t})d\tilde{s}d\tilde{t}\Big)^{1/2} + \Big(\int_0^1 \int_0^1 (\hat{\tilde{K}} - \tilde{K})^2(\tilde{s}, \tilde{t})d\tilde{s}d\tilde{t}\Big)^{1/2}
\end{aligned} \tag{15}$$

For any fixed n, $\hat{\tilde{K}}(\tilde{s}, \tilde{t})$ is a bounded piece-wise constant function with finite discontinuities, so the first term in (15) converges to 0 as the mesh size $m \to \infty$. For the second term in (15), notice that

$$\Big(\int_0^1 \int_0^1 (\hat{\tilde{K}} - \tilde{K})^2(\tilde{s}, \tilde{t})d\tilde{s}d\tilde{t}\Big)^{1/2} \le \sup_{0 \le \tilde{s} \le 1, 0 \le \tilde{t} \le 1} |\hat{\tilde{K}}(\tilde{s}, \tilde{t}) - \tilde{K}(\tilde{s}, \tilde{t})|.$$

Then by the consistency of $\hat{F}^*_{pool}$ to F^*, $\hat{c}_i$ to c_i, $\hat{K}$ to K, and by the continuity of K, the continuous mapping theorem gives

$$\lim_{n\to+\infty} \sup_{0\le\tilde{s}\le 1, 0\le\tilde{t}\le 1} |\widehat{\tilde{K}}(\tilde{s},\tilde{t}) - \tilde{K}(\tilde{s},\tilde{t})| = 0 \text{ in probability.}$$

Thus $\lim_{m,n\to+\infty} \|\tilde{\mathcal{T}}_m - \tilde{\mathcal{T}}\| = 0$ in probability. The desired result then follows from Theorem 4.10 in Chapter 5 of Kato (1980). □

References

Adelstein, D., Li, Y., Adams, G., Wagner, H., Kish, J., Ensley, J., Schuller and D., Forastiere, A., "A phase III comparison of standard radiation therapy (RT) versus RT plus concurrent cisplatin (DDP) versus split-course RT plus concurrent DDP and 5FU in patients with unresectable squamous cell," *Journal of Clinical Oncology* vol 21, pp. 92-98, 2003.

Andersen, P.K. , Borgan, O. , Gill, R. D., Keiding, N., Statistical Models Based on Counting Processes, Springer-Verlag: New York, 1993

Berkson, J. and Gage, R.P., "Survival Curves for Cancer Patients Following Treatment," *Journal of the American Statistical Association* vol. 47 pp. 501-515, 1952.

Breslow, N. and Crowley, J., "A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship," *The Annals of Statistics* vol. 2, pp. 437-453, 1974

Dunford, N. and Schwartz, J., *Linear Operators, Part 1, General Theory,* Interscience: New York, 1958

Durbin, J. and Knott, M., "Components of Cramer-von Mises Statistics, I," *Journal of Royal Statistical Society, Series B* vol. 34, 2, pp. 290-307, 1972.

Durbin, J., Knott, M. and Taylor, C.C., "Components of Cramer-von Mises Statistics, II," *Journal of Royal Statistical Society, Series B* vol. 37, 2, 216-237, 1975.

Ethier, S., Kurtz and T.G., *Markov Processes, Characterization and Convergence,* John Wiley and Sons: New York, 1976

Farewell, V.T., "The use of mixture models for the analysis of survival data with long-term survivors," *Biometrics* vol. 38 pp. 1041-1046, 1982.

Fleming, T. and Harrington, D., *Counting Processes and Survival Analysis,* John Wiley and Sons: New York, 1991.

Gill, R.D., "Understanding Cox's Regression Model: A Martingale Approach, " *Journal of the American Statistical Association* vol. 79, pp. 441-447, 1984.

Gray, R.J. and Tsiatis, A.A., "A Linear Rank Test for Use When the Main Interest is in Differences in Cure Rates," *Biometrics* vol 45, pp. 899-904, 1989.

Greenhouse, J.B. and Wolfe, R.A., "A Competing Risks Derivation of a Mixture Model for the Analysis of Survival Data, *Communications in Statistics, Part A – Theory and Methods* vol. 13 pp. 3133-3154, 1984.

Kato, T., *Perturbation Theory for Linear Operators,* Corrected Printing of the Second Edition. Springer: New York, 1980.

Koziol, J. and Green, S., "A Cramér-von Mises Statistic For Randomly Censored Data," *Biometrika* vol. 63 pp. 466-474, 1976.

Kuk, A. Y. C. and Chen, C., "A Mixture Model Combining Logistic Regression with Proportional Hazards Regression," *Biometrika* vol. 79, pp. 531-541, 1992.

Kurtz, T. and Protter, P., "Weak Limit Theorems For Stochastic Integrals and Stochastic Differential Equations, " *Annals of Probability* vol. 19, 1035-1070, 1991.

Laska, E., Siegel, C. and Sunshine, A., "Onset and Duration: Measurement and Analysis," *Clinical Pharmacology and Therapeutics* vol. 49, pp. 1-5, 1991.

Laska, E. and Meisner, M., "Nonparametric Estimation and Testing in a Cure Model," *Biometrics* vol. 48, pp. 1223-1234, 1992.

Maller, R.A. and Zhou, X., "Estimating the proportion of immunes in a censored sample," *Biometrika* vol. 79, pp. 731-739, 1992.

Maller, R.A., Zhou, X., "The probability that the largest observation is censored," *Journal of Applied Probability* vol. 30, 602-615, 1993.

Orazem, J., "A Nonparametric Analysis of Survival Data Within A Mixture of Susceptibles and Nonsusceptibles," *Ph.D thesis,* Columbia University, 1991.

Peng, Y., Dear, K.B.G., "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics* vol. 56, pp. 237-243, 2000

Schumacher, M., "Two-sample Tests of Cramér-von Mises- and Kolmogorov-Smirnov-type for Randomly Censored Data," *International Statistical Review* vol. 52, pp.263-282, 1984.

Shorack, G.R. and Wellner, J.A., *Empirical Processes With Applications to Statistics*, Wiley: New York, 1986.

Stute, W., "Nonparametric model checks for regression," *Annals of Statistics* vol. 25, pp. 613-641, 1997.

Sy, J.P. and Taylor, J.M., " Estimation in a Cox proportional Hazards Cure Model," *Biometrics* vol. 56, pp. 227-236, 2000.

Tamura, R., Faries, D. and Feng, J., "Comparing time to Onset of Response in Antidepressant Clinical Trials Using the Cure Model and the Cramér-von Mises test," *Statistics in Medicine* vol. 19 pp. 2169-2184, 2000.

von Mises, R., "On the Asymptotic Distribution of Differentiable Statistical Functions," *The Annals of Mathematical Statistics* vol. 18, pp. 309-348, 1947.

Table 1: Rejection rates for the Cramer-von Mises statistic with 1000 realizations for each parameter setting. The sample size is 100 per treatment group.

p_1	p_2	censoring percentage prior to time 43	β	rejection rate algorithm 1	algorithm 2	bootstrap(*)
Null hypothesis cases						
0.6	0.6	0	1.0	0.051	0.051	0.048
0.6	0.6	35	1.0	0.063	0.063	0.056
0.6	0.6	50	1.0	0.086	0.089	0.082
0.6	0.9	0	1.0	0.048	0.047	0.051
0.6	0.9	35	1.0	0.063	0.060	0.065
0.6	0.9	50	1.0	0.075	0.076	0.064
Alternative hypothesis cases						
0.6	0.6	0	1.5	0.477	0.478	0.425
0.6	0.6	35	1.5	0.410	0.414	0.365
0.6	0.6	50	1.5	0.404	0.411	0.300
0.6	0.9	0	1.5	0.541	0.546	0.474
0.6	0.9	35	1.5	0.472	0.470	0.416
0.6	0.9	50	1.5	0.421	0.424	0.291
0.6	0.6	0	2.0	0.894	0.894	0.895
0.6	0.6	35	2.0	0.821	0.827	0.817
0.6	0.6	50	2.0	0.807	0.802	0.682
0.6	0.9	0	2.0	0.941	0.940	0.930
0.6	0.9	35	2.0	0.898	0.894	0.878
0.6	0.9	50	2.0	0.838	0.832	0.754
0.6	0.6	0	2.5	0.988	0.989	0.988
0.6	0.6	35	2.5	0.977	0.974	0.981
0.6	0.6	50	2.5	0.963	0.963	0.914
0.6	0.9	0	2.5	0.996	0.996	0.997
0.6	0.9	35	2.5	0.981	0.985	0.985
0.6	0.9	50	2.5	0.977	0.975	0.945

* adopted from Table II in Tamura et al. (2000).

Table 2: Percentiles of the Limit Distribution of the Cramér-von Mises Statistics Obtained by Two Algorithms.

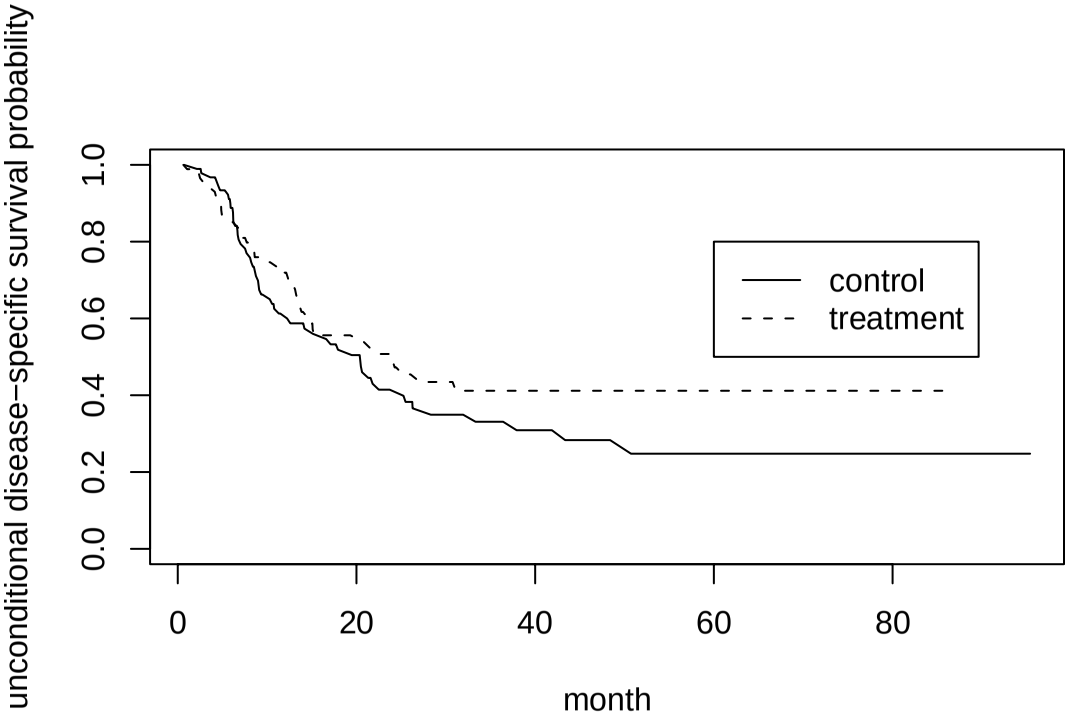
	5%	10%	15%	20%	25%	30%	35%	40%	45%
Alg. 1	0.2635	0.3374	0.3993	0.4531	0.5089	0.5680	0.6328	0.7010	0.7758
Alg. 2	0.2600	0.3308	0.3899	0.4510	0.5110	0.5698	0.6339	0.7045	0.7846

50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
0.8573	0.9543	1.0647	1.1864	1.3283	1.4919	1.7083	2.0144	2.4629	3.2606
0.8638	0.9635	1.0717	1.1923	1.3287	1.5003	1.7326	2.0434	2.4519	3.2715

Table 3: Leading Eigenvalues of the Linear Operator Used in Algorithm 2.

7.181848e-01	1.953676e-01	7.928266e-02	4.648866e-02	2.991972e-02
2.083118e-02	1.550592e-02	1.218492e-02	9.903118e-03	8.482065e-03
6.716481e-03	5.820127e-03	4.730430e-03	3.897105e-03	3.749634e-03
3.389445e-03	2.913652e-03	2.697358e-03	2.510337e-03	2.327357e-03
2.083282e-03	2.010754e-03	1.681561e-03	1.554187e-03	1.511168e-03

Figure 1: Comparison of Disease-specific Survival by Treatment Arm



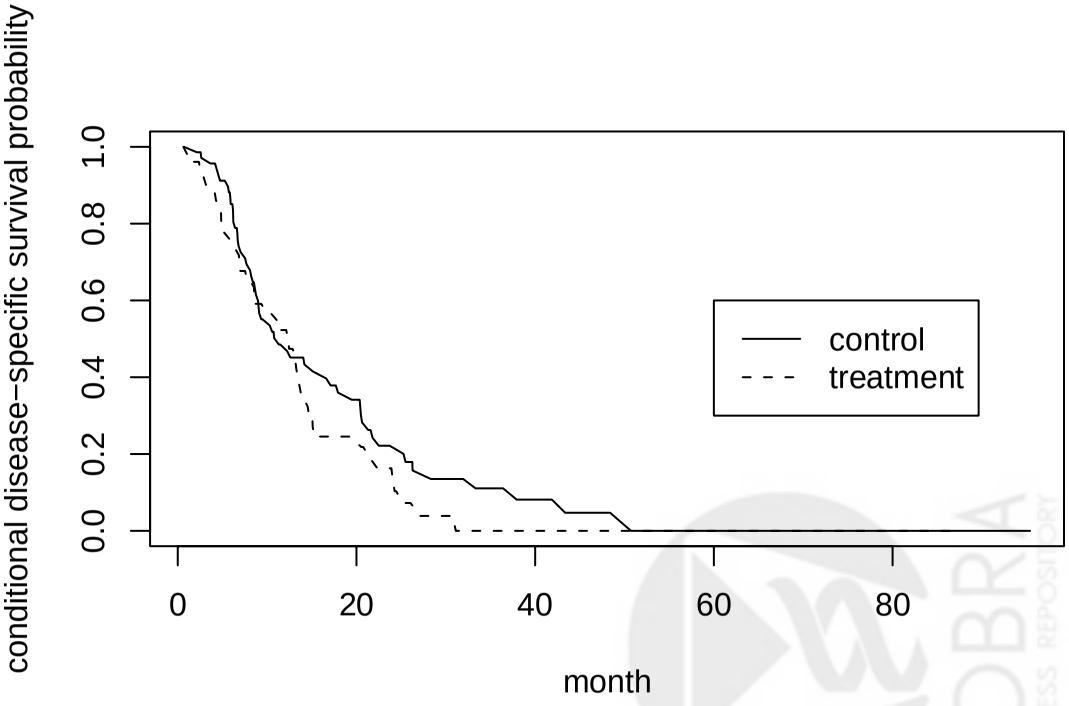


Figure 2: Density Curves of Cramér-von Mises Statistic