Published in final edited form as:

Pharm Stat. 2015 ; 14(3): 226–232. doi:10.1002/pst.1678.

Single-Arm Phase II Trial Design Under Parametric Cure Models

Jianrong Wu

Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Abstract

The current practice of designing single-arm phase II survival trials is limited under the exponential model. Trial design under the exponential model may not be appropriate when a portion of patients are cured. There is no literature available for designing single-arm phase II trials under the parametric cure model. In this article, a test statistic is proposed, and a sample size formula is derived for designing single-arm phase II trials under a class of parametric cure models. Extensive simulations showed that the proposed test and sample size formula perform very well under different scenarios.

Keywords

Clinical trial design; Cure model; Time-to-event; Sample size; Single-arm phase II trial

1 Introduction

The treatment of cancer has progressed dramatically in recent decades. It is not uncommon to see a significant proportion of patients with various types of cancers, e.g. breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer, being cured or surviving long term [1]. Traditional survival methods, such as the Cox model, may not be valid when there exists a proportion of patients who are cured. Thus, to adequately account for cured patients in survival data from cancer clinical trials, cure models are becoming increasingly useful. Various parametric and semiparametric cure models have been proposed by Farewell [2], Peng et al. [3] and Kuk and Chen [4], etc. A maximum-likelihood EM algorithm for parametric and semiparametric cure models has been proposed by Peng and Dear [5] and Sy and Taylor [6]. A SAS macro PSPMCM developed by Corbiere and Joly [7] is available to fit both parametric and semiparametric cure models.

The traditional methods for survival trial design may not be appropriate when there exists a cure rate. Sample size calculations for cure models have been developed, mainly for randomized phase III trials. For example, Halpern and Brown [8] developed a computer program to calculate the power and sample size for the exponential cure models based on Monte Carlo simulation. Ewell and Ibrahim [1] derived a power formula by considering a general alternative that allows for treatment effects on both the short and long-term survival. Recently, Wang et al. [9] derived a sample size formula under the proportional hazards cure model. However, there is no method available to design a single-arm phase II trial under the

cure models. In this paper, a one-sample test statistic is introduced for the purpose of designing single-arm phase II trials under the parametric cure models. The proposed test statistic can be used to detect improvement of the cure rate, short-term survival or both.

The rest of the paper is organized as follows. The parametric cure models are introduced in Section 2. In Section 3, a one-sample test statistic is proposed. A sample size formula is derived in Section 4. Simulation studies are conducted to study the performance of the proposed test in Section 5. An example is given in Section 6 to illustrate the trial design by using the proposed method. Concluding remarks are made in Section 7.

2 Cure Models

The failure time, T^* , is assumed to be $T^* = \upsilon T + (1 - \upsilon)\infty$, where υ is an indicator of whether a subject will eventually ($\upsilon = 1$) or never ($\upsilon = 0$) experience the failure, and T denotes the failure time if the subject is not cured, with a survival distribution $S(t)$, which is the conditional distribution for patients who will experience the failure, often called the latency distribution. Thus, the marginal survival distribution of T^* is a mixture model of a cure rate $\pi = P(\upsilon = 0)$ and a latency distribution $S(t)$ given by

$$S^*(t) = \pi + (1 - \pi)S(t). \quad (1)$$

Let $\lambda^*(t)$ and $\Lambda^*(t)$ be the hazard and cumulative hazard functions of T^* and $\lambda(t)$ be the hazard function of T , then we have the following relation for the hazard functions:

$$\lambda^*(t) = \frac{(1 - \pi)S(t)}{\pi + (1 - \pi)S(t)} \lambda(t), \quad (2)$$

and for the cumulative hazard function

$$\Lambda^*(t) = -\log\{\pi + (1 - \pi)S(t)\}. \quad (3)$$

A simple and popular parametric cure model is the exponential cure model

$$S^*(t) = \pi + (1 - \pi)e^{-\lambda t}, \quad (4)$$

which can be fully specified by the cure rate π and hazard rate λ for short-term survival. Suppose the historical survival data from previous trials for a certain type of disease follow the exponential cure model (4) with a cure rate of π_0 and a hazard rate of λ_0 . We want to design a study to determine whether the new treatment improves the cure rate π , the hazard rate λ or both. Specifically, we want to test the following null hypothesis:

$$H_0: \pi = \pi_0, \text{ and } \lambda = \lambda_0 \quad (5)$$

against various alternative hypotheses of interest: H_{1a} : improvements are in the short-term survival but not in the cure rate; H_{1b} : improvements are in the cure rate but not in the short-term survival; H_{1c} : improvements are in both the cure rate and short-term survival. Specifically, these alternative hypotheses can be formulated as

$$H_{1a}: \pi = \pi_0 \text{ and } \lambda = \lambda_1, \quad (6)$$

$$H_{1b}: \pi = \pi_1 \text{ and } \lambda = \lambda_0, \quad (7)$$

$$H_{1c}: \pi = \pi_1 \text{ and } \lambda = \lambda_1, \quad (8)$$

where the cure rate and hazard rate under the alternative are π_1 and λ_1 , respectively, with $\pi_1 > \pi_0$ and $\lambda_1 < \lambda_0$.

Among the parametric models, the Weibull, Log-normal, and Log-logistic distributions are widely used in the parametric cure model analysis [10]. These survival distribution functions

are as follows: 1) Weibull $S(t) = e^{-\lambda t^\kappa}$; 2) Log-normal $S(t) = 1 - \Phi\left(\frac{\log t - \lambda}{\kappa}\right)$; 3) Log-

logistic $S(t) = \frac{1}{1 + \lambda t^\kappa}$, with a shape parameter κ and a scale parameter λ . In this article, we will consider the study design for a single-arm phase II trial under the mixture model (1), with the latency distribution being one of these three distributions. In the design stage, we assume that the shape parameter κ can be estimated from historical data; thus, it is assumed to be known. Then the latency distribution $S(t)$ is fully specified by the scale parameter λ or equivalently by the survival probability at a landmark time point x . Specifically, let $S_0(x)$ and $S_1(x)$ be the survival probability at a landmark time point x under the null and alternative hypotheses, thus, the hypotheses of (5–8) are equivalent to the following null hypothesis:

$$H_0: \pi = \pi_0, \text{ and } S(x) = S_0(x), \quad (9)$$

against one of the following three alternative hypotheses:

$$H_{1a}: \pi = \pi_0 \text{ and } S(x) = S_1(x), \quad (10)$$

$$H_{1b}: \pi = \pi_1 \text{ and } S(x) = S_0(x), \quad (11)$$

$$H_{1c}: \pi = \pi_1 \text{ and } S(x) = S_1(x). \quad (12)$$

In general, let $S^*(t) = \pi + (1 - \pi)S(t)$ be the marginal survival distribution of a new study, where the latency distribution $S(t)$ is one of the three distributions. Suppose that the cure rate and latency distribution for the historical data are π_0 and $S_0(t)$, respectively, then the above hypotheses (5–8) or (9–12) are equivalent to the following null hypothesis:

$$H_0: S^*(t) = S_0^*(t) \text{ vs } H_1: S^*(t) = S_1^*(t) \quad (13)$$

where $S_i^*(t) = \pi_i + (1 - \pi_i)S_i(t)$, $i = 0, 1$, $\pi_1 > \pi_0$ and $S_1(t) > S_0(t)$, with at least one of the inequalities being strictly greater than.

3 Test Statistics

To test the hypothesis (13), we can use a one-sample test introduced briefly in this section [11]. Suppose during the accrual phase of the trial, n subjects are enrolled in the study. Let T_i^* and C_i denote, respectively, the failure time and censoring time of the i^{th} subject, where the failure time T_i^* follows the mixture cure model given in equation (1). We assume that the failure time T_i^* and censoring time C_i are independent and $\{T_i^*, C_i, i = 1, \dots, n\}$ are independent and identically distributed. Then the observed failure time and failure indicator are $X_i = T_i^* \wedge C_i$ and $\Delta_i = I(T_i^* \leq C_i)$, respectively, for the i^{th} subject. On the basis of the observed data $\{X_i, \Delta_i, i = 1, \dots, n\}$, we define $O = \sum_{i=1}^n \Delta_i$ as the observed number of failures, and $E = \sum_{i=1}^n \Lambda_0^*(X_i)$ as the expected number of failures (asymptotically), where $\Lambda_0^*(t) = -\log S_0^*(t)$ is the cumulative hazard of $S_0^*(t)$ under the null hypothesis (13). Then the one-sample test is defined by [11]

$$L = \frac{O - E}{\sqrt{(O + E)/2}}. \quad (14)$$

To study the asymptotic distribution of the one-sample test statistic, we formulate it using counting-process notations. Specifically, let $N_i(t) = I\{X_i \leq t\}$ and $Y_i(t) = I\{X_i \geq t\}$ be the failure and at-risk processes, respectively. Then

$$O = \sum_{i=1}^n \int_0^\infty dN_i(t), \text{ and } E = \sum_{i=1}^n \int_0^\infty Y_i(t) d\Lambda_0^*(t).$$

Thus, the counting-process formulation of the one-sample test is given by $L = W/\hat{\sigma}$, where

$$W = n^{-1/2} \sum_{i=1}^n \int_0^\infty \{dN_i(t) - Y_i(t) d\Lambda_0^*(t)\}$$

and

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \int_0^\infty \{dN_i(t) + Y_i(t) d\Lambda_0^*(t)\} / 2.$$

As shown in the Appendix 1, the exact mean of W under the null $E_{H_0}(W)$ is zero and $\hat{\sigma}^2$ is an consistent estimate of $\text{Var}_{H_0}(W)$. Therefore, by counting process central limit theorem [12], under the null hypothesis, L is asymptotically standard normal distributed. Hence, we reject the null hypothesis H_0 with one-sided type I error α if $L = W/\hat{\sigma} < -z_{1-\alpha}$, where $z_{1-\alpha}$ is the $100(1 - \alpha)$ percentile of the standard normal distribution.

4 Sample Size Calculation

To design the study, we must calculate the sample size to detect a specified survival difference at the alternative $H_1: S^*(t) = S_1^*(t) (> S_0^*(t))$, given the type I error α and power 1

$-\beta$. For the sample size calculation, the exact variance of W under the alternative has been derived by Wu [13]. Let the exact mean and variance of W at the alternative be

$E_{H_1}(W) = \sqrt{n}\omega$ and $\text{Var}_{H_1}(W) = \sigma^2$, respectively, where ω and σ^2 are given below. By counting process central limit theorem, $(W - \sqrt{n}\omega)/\sigma$ is asymptotically standard normal distributed under H_1 .

As shown in equations (A7) and (A8) in the Appendix 1, under the alternative $\sigma^2 \rightarrow \sigma^2 = (v_1 + v_0)/2$ and $W/\hat{\sigma} - \sqrt{n}\omega/\bar{\sigma} \rightarrow N(0, \sigma^2/\bar{\sigma}^2)$, thus, the power of the test $L = W/\sigma^2$ satisfies the following equations:

$$1 - \beta = P(L < -z_{1-\alpha}) \simeq \Phi\left(-\frac{\bar{\sigma}}{\sigma}z_{1-\alpha} - \frac{\sqrt{n}\omega}{\sigma}\right).$$

Therefore, the required sample size for the test statistic L is given by

$$n = \frac{(\bar{\sigma}z_{1-\alpha} + \sigma z_{1-\beta})^2}{\omega^2}, \quad (15)$$

where $\omega = v_1 - v_0$, $\sigma^2 = (v_1 + v_0)/2$, and $\sigma^2 = v_1 - v_1^2 + 2v_{00} - v_0^2 - 2v_{01} + 2v_0v_1$, with v_0 , v_1 , v_{00} , and v_{01} given as follows (see Appendix 1 equations (A3)–(A6)):

$$v_0 = \int_0^\infty G(t)S_1^*(t)d\Lambda_0^*(t),$$

$$v_1 = \int_0^\infty G(t)S_1^*(t)d\Lambda_1^*(t),$$

$$v_{00} = \int_0^\infty G(t)S_1^*(t)\Lambda_0^*(t)d\Lambda_0^*(t),$$

$$v_{01} = \int_0^\infty G(t)S_1^*(t)\Lambda_0^*(t)d\Lambda_1^*(t),$$

and $G(t)$ is the survival function of censoring time C .

To calculate the quantities v_0 , v_1 , v_{00} , and v_{01} , we assumed that subjects were recruited with a uniform distribution over the accrual period t_a and were followed for a period of t_f , and no loss to follow-up. Thus, the censoring distribution is a uniform distribution on $[t_f, t_a + t_f]$. Then for a parametric cure model in which the cure rate and scale parameter are determined by the null and alternative as described above, the quantities v_0 , v_1 , v_{00} , and v_{01} can be calculated by numeric integrations, e.g. using R function `integrate`. For example, for the exponential cure model,

$$v_0 = \int_0^{t_f} S_1^*(t) \lambda_0^*(t) dt + \frac{1}{t_a} \int_{t_f}^{\tau} (\tau - t) S_1^*(t) \lambda_0^*(t) dt,$$

$$v_1 = \int_0^{t_f} S_1^*(t) \lambda_1^*(t) dt + \frac{1}{t_a} \int_{t_f}^{\tau} (\tau - t) S_1^*(t) \lambda_1^*(t) dt,$$

$$v_{00} = \int_0^{t_f} S_1^*(t) \Lambda_0^*(t) \lambda_0^*(t) dt + \frac{1}{t_a} \int_{t_f}^{\tau} (\tau - t) S_1^*(t) \Lambda_0^*(t) \lambda_0^*(t) dt,$$

$$v_{01} = \int_0^{t_f} S_1^*(t) \Lambda_0^*(t) \lambda_1^*(t) dt + \frac{1}{t_a} \int_{t_f}^{\tau} (\tau - t) S_1^*(t) \Lambda_0^*(t) \lambda_1^*(t) dt,$$

where $\tau = t_a + t_f$; and

$$S_1^*(t) = \pi_1 + (1 - \pi_1) e^{-\lambda_1 t},$$

$$\Lambda_0^*(t) = -\log\{\pi_0 + (1 - \pi_0) e^{-\lambda_0 t}\},$$

$$\lambda_i^*(t) = \lambda_i \frac{(1 - \pi_i) e^{-\lambda_i t}}{\pi_i + (1 - \pi_i) e^{-\lambda_i t}}, i=0, 1.$$

Therefore, the study design can be conducted by calculating the sample size using formula (15).

5 Simulation Studies

To study the performance of the one-sample test under various parametric cure models, including the Weibull, Log-normal and Log-logistic distributions, we calculated sample sizes for each of these distributions under various design scenarios. Simulation studies were conducted to study the accuracy of the sample size estimation and performance of the one-sample test under small sample size. Under each of these distributions, the shape parameter was set to $\kappa=0.5, 1$, and 2 reflecting the different types of hazard functions. Instead of specifying the scale parameter λ equivalently, we specified the survival probability at a landmark time point $x = 2$. The survival probability under the null and alternative was set to $S_0(2) = 0.2 - 0.6$ and under the alternative was set to $S_1(2) = 0.2 - 0.75$; the cure rate under the null was set to $\pi_0 = 0 - 0.3$ and under the alternative was set to $\pi_1 = 0 - 0.45$. We assumed that subjects were recruited with a uniform distribution over the accrual period $t_a =$

3 and were followed for $t_f = 1$. The censoring distribution is a uniform distribution on the interval $[t_f, t_a + t_f]$. Thus under any one of these survival distributions and the uniform censoring distribution, quantities v_0 , v_1 , v_{00} , and v_{01} can be calculated by numeric integrations.

Three scenarios of the alternative hypotheses were considered in the simulations. Scenario 1: improvements are in the short-term survival but not in the cure rate. Scenario 2: improvements are in the cure rate but not in the short-term survival. Scenario 3: improvements are in both the cure rate and short-term survival. Therefore, given the nominal significance level 0.05 and power of 80%, the required sample sizes for each parameter configuration, survival distribution and hypothesis scenario were calculated. The empirical type I error and power for the corresponding sample size were simulated based on 100,000 runs (Tables 1–3).

The simulation results can be summarized as follows. First, from sample size calculations, we observed that the sample sizes were in the typical range of a phase II trial for various designs and three hypotheses scenarios. Second, the empirical powers were close to the nominal level of 80%. Thus, the sample sizes were correctly estimated. The empirical type I errors were, in general, also close to the nominal level of 5%. Overall, the results showed that the one-sample test preserved type I error well and provided adequate power for designing single-arm phase II survival trials under the parametric cure models.

6 An Example

In this section, we illustrate the application of the study design, under a parametric cure model by using the data from the Eastern Cooperative Oncology Group (ECOG) trial e1684 as an example. The ECOG trial e1684 was a two-arm phase III clinical trial to compare the relapse-free survival (RFS) of patients with melanoma who were treated with high-dose interferon alpha-2b and placebo as postoperative adjuvant therapy. Researchers have extensively studied this dataset using the cure models [14]. There were 92 deaths among 146 patients in the treatment group. The SAS macro PSPMCM was applied to this data to fit the treatment arm data under the Weibull cure model (Figure), with an estimated shape parameter $\kappa = 1.018$, scale parameter $\lambda = 0.836$, and cure rate of 35%. Suppose a new therapy is available, and we want to design a single-arm phase II study to compare the RFS of the new treatment to that of the treatment arm of the ECOG trial, which will serve as a literature control. Three scenarios are considered here: 1) improve the short-term survival by reducing the hazard rate 1.75 times but not the cure rate; 2) increase the cure rate by 20% but not the short-term survival; 3) increase the cure rate by 10% and improve the short-term survival by reducing the hazard rate 1.25 times. With a type I error of 0.05, power of 80% at the alternative, 3 years accrual period and 1 year follow-up, the required sample sizes calculated using formula (15) under the Weibull cure model are 93, 94 and 95 patients for scenarios 1–3, respectively. The corresponding simulated empirical type I error and power are 0.053 and 79.7%, 0.053 and 80%, and 0.054 and 79.6% for scenarios 1–3, respectively. Thus, the calculated sample sizes preserve the type I error and provide adequate power for the trial designs.

7 Conclusion

In this article, a sample size formula is derived for designing single-arm phase II trials under the parametric cure models. Three parametric cure models are considered, including the Weibull, Log-normal, Log-logistic distributions that have been implemented in a SAS macro PSPMCM. Thus, we can estimate the cure rate and model parameters by using the SAS macro for the trial design. The extensive simulations showed that the proposed test preserves type I error well and provides adequate power for the study design. Finally, the proposed method does not limit to the three cure models mentioned above; it can be applied to other parametric cure model too, such as the Gamma and Gompertz models, etc. Therefore, the proposed method has a wide range of applications for designing single-arm phase II trials under various parametric cure models. For convenience, two R functions SIZE and TEST are given (Appendix 2) for the sample size and p-value calculations using the proposed methods under the Weibull cure model. One can easily change the distribution in the R functions for trial designs under the Log-normal and Log-logistic distributions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The author acknowledges two anonymous reviewers for their valuable comments that improved an earlier version of the paper. This work was supported in part by the National Cancer Institute (NCI) support grant CA21765 and ALSAC.

Appendix 1: Derivation of the exact variance of W under the alternative

First, we calculate the mean and variance of W under the null hypothesis H_0 by noting that $E_{H_0}(O) = nE_{H_0}()$ and $E_{H_0}(E) = nE_{H_0}(\Lambda_0^*(X))$.

Let $f_0^*(t)$, $S_0^*(t)$, and $\Lambda_0^*(t)$ be the density, survival, and cumulative hazard functions of failure time T under the null and $g(t)$ and $G(t)$ be the density and survival functions of censoring time C . Then, by exchange of integrations, we have

$$E_{H_0}(\Delta) = \int_0^\infty \left(\int_0^y f_0^*(t) g(y) dt \right) dy = \int_0^\infty \left(f_0^*(t) \int_t^\infty g(y) dy \right) dt = \int_0^\infty G(t) S_0^*(t) d\Lambda_0^*(t).$$

Let $S_x^*(t)$ be the survival distribution of $X = T \wedge C$ under the null, then $S_x^*(t) = S_0^*(t)G(t)$ and by integration by parts, we have

$$E_{H_0}(\Lambda_0^*(X)) = - \int_0^\infty \Lambda_0^*(t) dS_x^*(t) = \int_0^\infty S_x^*(t) d\Lambda_0^*(t) = \int_0^\infty G(t) S_0^*(t) d\Lambda_0^*(t) = E_{H_0}(\Delta). \quad (A1)$$

Therefore, the mean of W under the null is $E_{H_0}(W) = \sqrt{n}\{E_{H_0}(\Delta) - E_{H_0}(\Lambda_0^*(X))\} = 0$. By similar calculation, we have

$$E_{H_0}(\Delta\Lambda_0^*(X)) = \int_0^\infty \left(\int_0^y f_0(t)g(y)\Lambda_0^*(t)dt \right) dy = \int_0^\infty \left(f_0(t)\Lambda_0^*(t) \int_t^\infty g(y)dy \right) dt = \int_0^\infty G(t)S_0^*(t)\Lambda_0^*(t)d\Lambda_0^*(t),$$

and

$$E_{H_0}(\Lambda_0^{*2}(X)) = - \int_0^\infty \Lambda_0^{*2}(t)dS_X(t) = 2 \int_0^\infty S_X(t)\Lambda_0^*(t)d\Lambda_0^*(t) = 2 \int_0^\infty G(t)S_0^*(t)\Lambda_0^*(t)d\Lambda_0^*(t).$$

So far, we have shown that $E_{H_0}(\Delta) = E_{H_0}(\Lambda_0^*(X))$ and $E_{H_0}(\Lambda_0^{*2}(X)) = 2E_{H_0}(\Delta\Lambda_0^*(X))$.

Therefore,

$$\text{Var}_{H_0}(W) = E_{H_0}(\Delta - \Lambda_0^*(X))^2 = E_{H_0}(\Delta) - 2E_{H_0}(\Delta\Lambda_0^*(X)) + E_{H_0}(\Lambda_0^{*2}(X)) = E_{H_0}(\Delta) = \int_0^\infty G(t)S_0^*(t)d\Lambda_0^*(t). \quad (\text{A2})$$

Under the null hypothesis H_0 , $n^{-1}O \rightarrow E_{H_0}(\cdot)$ and

$n^{-1} \sum_{i=1}^n \int_0^\infty Y_i(t)d\Lambda_0^*(t) \rightarrow \int_0^\infty G(t)S_0^*(t)d\Lambda_0^*(t) = E_{H_0}\{\Lambda_0^*(X)\}$. Thus from equation (A1), $\hat{\sigma}^2 \rightarrow \sigma_0^2 = \int_0^\infty G(t)S_0^*(t)d\Lambda_0^*(t)$, which is the exact variance of W under the null hypothesis (A2). Thus,

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \int_0^\infty \{dN_i(t) + Y_i(t)d\Lambda_0^*(t)\} / 2$$

is a consistent estimate of $\text{Var}_{H_0}(W)$ under the null, and

$$L = W/\hat{\sigma} \rightarrow N(0, 1).$$

Now we derive the exact variance of W under the alternative. Let $f_1(t)$, $S_1^*(t)$, and $\Lambda_1(t)$ be the density, survival, and cumulative hazard functions, respectively, of failure time T under the alternative. Then by similar calculation, we have

$$E_{H_1}(\Delta) = \int_0^\infty \left(\int_0^y f_1^*(t)g(y)dt \right) dy = \int_0^\infty \left(f_1^*(t) \int_t^\infty g(y)dy \right) dt = \int_0^\infty G(t)S_1^*(t)d\Lambda_1^*(t) = v_1. \quad (\text{A3})$$

Let $S_X(t)$ be the survival distribution of $X = T \wedge C$ under the alternative, then

$S_X^*(t) = G(t)S_1^*(t)$, and by integration by parts, we have

$$E_{H_1}(\Lambda_0^*(X)) = - \int_0^\infty \Lambda_0^*(t)dS_X^*(t) = \int_0^\infty S_X^*(t)d\Lambda_0^*(t) = \int_0^\infty G(t)S_1^*(t)d\Lambda_0^*(t) = v_0. \quad (\text{A4})$$

Thus, $E_{H_1}(W) = \sqrt{n}\{E_{H_1}(\Delta) - E_{H_1}(\Lambda_0^*(X))\} = \sqrt{n}(v_1 - v_0)$. Similarly, we have

$$E_{H_1}(\Delta\Lambda_0^*(X)) = \int_0^\infty \left(\int_0^y f_1^*(t)g(y)\Lambda_0^*(t)dt \right) dy = \int_0^\infty \left(f_1^*(t)\Lambda_0^*(t) \int_t^\infty g(y)dy \right) dx = \int_0^\infty G(t)S_1^*(t)\Lambda_0^*(t)d\Lambda_1^*(t) = v_{01}, \quad (\text{A5})$$

and

$$E_{H_1}(\Lambda_0^{*2}(X)) = -\int_0^\infty \Lambda_0^{*2}(t) dS_X^*(t) = 2\int_0^\infty S_X^*(t) \Lambda_0^*(t) d\Lambda_0^*(t) = 2\int_0^\infty G(t) S_1^*(t) \Lambda_0^*(t) d\Lambda_0^*(t) = 2v_{00}. \quad (A6)$$

Therefore, the exact variance of W under the alternative is given by

$$\begin{aligned} \text{Var}_{H_1}(W) &= \text{Var}_{H_1}(\Delta) + \text{Var}_{H_1}(\Lambda_0^*(X)) - 2\text{Cov}_{H_1}(\Delta, \Lambda_0^*(X)) \\ &= E_{H_1}(\Delta) - E_{H_1}^2(\Delta) + E_{H_1}(\Lambda_0^{*2}(X)) - E_{H_1}^2(\Lambda_0^*(X)) \\ &\quad - 2E_{H_1}(\Delta \Lambda_0^*(X)) + 2E_{H_1}(\Delta)E_{H_1}(\Lambda_0^*(X)) \\ &= v_1 - v_1^2 + 2v_{00} - v_0^2 - 2v_{01} + 2v_0v_1 = \sigma^2. \end{aligned}$$

Under the alternative H_1 ,

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \int_0^\infty \{dN_i(t) + Y_i(t) d\Lambda_0^*(t)\} / 2; \rightarrow \bar{\sigma}^2 = (v_0 + v_1) / 2; \quad (A7)$$

thus, $W - \sqrt{n}\omega \rightarrow N(0, \sigma^2)$, and by Slutsky's theorem, it follows that

$$\frac{W}{\hat{\sigma}} - \frac{\sqrt{n}\omega}{\bar{\sigma}} \rightarrow N(0, \frac{\sigma^2}{\bar{\sigma}^2}) \text{ as } n \rightarrow \infty. \quad (A8)$$

Appendix 2: R codes for the sample size calculation and the test statistic under the Weibull model

```
SIZE=function(kappa, pi0, pil, lambda0, lambda1, ta, tf, alpha, beta)
{### sample size calculation under the Weibull cure model
### kappa: shape parameter of the Weibull distribution
### pi0 and pil: cure rates under the null and alternative
### lambda0 and lambda1: hazard parameters under the null and alternative
### ta and tf: accrual period and follow-up time
### alpha and beta: type I and type II error. 1-beta is the study power.
S0=function(t) ### the Weibull cure model under the null
{ans=pi0+(1-pi0)*exp(-lambda0*t^kappa); return(ans)}
S1=function(t) ### the Weibull cure model under the alternative
{ans=pil+(1-pil)*exp(-lambda1*t^kappa); return(ans)}
G=function(t) ### uniform censoring distribution on interval [tf, tau]
{ans=1-punif(t, tf, tau); return(ans)}
Lambda0=function(t)# cumulative hazard function of the cure model under the
null
{ans=-log(S0(t)); return(ans)}
h0=function(t) # hazard function of the Weibull cure model under null
{a=lambda0*kappa*t^(kappa-1)*(1-pi0)*exp(-lambda0*t^kappa)}
```

```

b=pi0+(1-pi0)*exp(-lambda0*t^kappa); ans=a/b; return(ans) }
h1=function(t)      # hazard function of the Weibull cure model under
alternative
{a=lambda1*kappa*t^(kappa-1)*(1-pil)*exp(-lambda1*t^kappa)
b=pil+(1-pil)*exp(-lambda1*t^kappa); ans=a/b; return(ans)}
f0=function(t){ans=G(t)*S1(t)*h0(t);return(ans)}
f1=function(t){ans=G(t)*S1(t)*h1(t);return(ans)}
f2=function(t){ans=G(t)*S1(t)*h0(t)*Lambda0(t);return(ans)}
f3=function(t){ans=G(t)*S1(t)*h1(t)*Lambda0(t);return(ans)}
tau=ta+tf          # tau is the study duration
z0=qnorm(1-alpha); z1=qnorm(1-beta)
v0=integrate(f0, 0, tau)$value; v1=integrate(f1, 0, tau)$value
v00=integrate(f2, 0, tau)$value; v01=integrate(f3, 0, tau)$value
n=(sqrt((v0+v1)/2)*z0+sqrt(v1+2*v0*v1-v0^2-v1^2+2*v00-2*v01)*z1)^2/(v0-v1)^2
ans=ceiling(n); return(ans)
}
### Sample size calculation for the example based on ECOG data ####
SIZE(kappa=1.018,pi0=0.35,pil=0.45,lambda0=0.836,lambda1=0.669,ta=3,tf=1,alph
a=0.05,beta=0.2)
95
TEST=function(kappa, lambda0, pi0, data)
{#p-value calculation for the one-sample test under the Weibull distribution
#kappa and lambda0: shape parameter and hazard parameter under the null.
#pi0: cure rate under the null.
#data: data frame for survival data set with variable "time" for the
observed
#failure time and variable "cens" for failure indicator, 0-censoring, 1-
failure.
S0=function(t)      # the Weibull cure model under the null
{ans=pi0+(1-pi0)*exp(-lambda0*t^kappa);return(ans)}
Lambda0=function(t) # cumulative hazard function under the null
{ans=-log(S0(t)); return(ans)}
delta=data$cens; X=data$time; O=sum(delta); M=Lambda0(X); E=sum(M)
L0<-(O-E)/((O+E)/2)^(1/2); pvalue=round(pnorm(L0),4); return(pvalue)
}

```

References

1. Ewell M, Ibrahim JG. The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis*. 1997; 3:5–12. [PubMed: 9384622]
2. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 1982; 38:1041–1046. [PubMed: 7168793]
3. Peng Y, Dear KBG, Denham JW. A generalized F mixture model for cure rate estimation. *Statistics in Medicine*. 1998; 17:813–830. [PubMed: 9595613]
4. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992; 79:531–541.

5. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000; 56:237–243. [PubMed: 10783801]
6. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
7. Corbiere F, Joly P. A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine*. 2007; 85:173–180. [PubMed: 17157948]
8. Halpern J, Brown BW. Designing clinical trials with arbitrary specification of survival functions and for the log rank or generalized Wilcoxon test. *Controlled Clinical Trials*. 1987; 8:177–189. [PubMed: 3665521]
9. Wang S, Zhang J, Lu W. Sample size calculation for the proportional hazards cure model. *Statistics in Medicine*. 2012; 31:3959–3971. [PubMed: 22786805]
10. Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*. 2004; 23:1733–1747. [PubMed: 15160405]
11. Wu J. A new one-sample log-rank test. *Journal of Biometrics and Biostatistics*. 2014; 5(4):1–5.
12. Fleming, TR.; Harrington, DP. *Counting processes and survival analysis*. New York: John Wiley and Sons; 1991.
13. Wu J. Sample size calculation for the one-sample log-rank test. *Pharmaceutical Statistics*. 2015; 14:26–33. [PubMed: 25339496]
14. Ibrahim, JG.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer-Verlag Inc.; 2001.

Figure 1.
The Kaplan - Meier survival curve (step function); the fitted Weibull cure model (solid curve).

Scenario 1 in which sample size, simulated empirical type I error (α), and power ($1 - \beta$) were based on 100,000 simulation runs for various distributions. The nominal type I error was 0.05 and power was 80%; the censoring distribution was uniform on $[t_i, t_u + t_f]$, where $t_u = 3$ and $t_f = 1$.

Table 1

Scenario 1			$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
Dist.	(π_0, π_1)	S_0 vs S_1	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$
WB	(.0, .0)	.2 vs .35	44	.052	.803	44	.051	.810	44	.050	.812
	(.1, .1)	.2 vs .35	64	.052	.802	67	.052	.803	80	.050	.805
	(.2, .2)	.2 vs .35	88	.054	.801	96	.054	.800	124	.052	.802
	(.0, .0)	.3 vs .45	55	.052	.803	53	.052	.802	51	.051	.803
	(.1, .1)	.3 vs .45	72	.053	.797	74	.053	.800	84	.052	.799
	(.2, .2)	.3 vs .45	95	.053	.802	99	.053	.795	124	.052	.800
	(.0, .0)	.5 vs .65	58	.054	.794	55	.054	.800	49	.054	.805
	(.1, .1)	.5 vs .65	70	.055	.795	68	.054	.800	67	.053	.802
	(.2, .2)	.5 vs .65	85	.056	.792	84	.054	.798	88	.053	.797
	(.0, .0)	.6 vs .75	52	.054	.798	48	.057	.799	41	.055	.801
	(.1, .1)	.6 vs .75	61	.056	.796	57	.056	.793	52	.054	.801
	(.2, .2)	.6 vs .75	72	.057	.796	68	.055	.793	65	.054	.797
LN	(.0, .0)	.2 vs .35	39	.051	.815	37	.051	.810	36	.052	.806
	(.1, .1)	.2 vs .35	63	.053	.808	54	.052	.803	52	.053	.804
	(.2, .2)	.2 vs .35	94	.052	.800	78	.052	.800	72	.052	.797
	(.0, .0)	.3 vs .45	52	.053	.808	49	.052	.805	48	.052	.802
	(.1, .1)	.3 vs .45	79	.052	.806	67	.054	.800	64	.053	.798
	(.2, .2)	.3 vs .45	113	.053	.799	91	.054	.798	85	.053	.800
	(.0, .0)	.5 vs .65	60	.054	.803	57	.053	.801	56	.054	.795
	(.1, .1)	.5 vs .65	82	.053	.799	71	.054	.799	68	.055	.793
	(.2, .2)	.5 vs .65	109	.054	.801	89	.055	.797	84	.054	.794
	(.0, .0)	.6 vs .75	54	.055	.805	52	.055	.801	51	.055	.795
	(.1, .1)	.6 vs .75	70	.055	.800	62	.057	.794	61	.054	.797
	(.2, .2)	.6 vs .75	90	.055	.801	75	.056	.793	72	.056	.791
LG			(.0, .0)	.2 vs .35	35	.052	.800	.806	37	.051	.802

Dist.	Scenario 1			$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
	(π_0, π_1)			S_0 vs S_1			α			$1 - \beta$		
	π_0	π_1	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$	n
	(.1, .1)		50	.052	.799	52	.052	.805	56	.051	.802	
	(.2, .2)		70	.052	.797	73	.052	.799	81	.053	.800	
	(.0, .0)		49	.053	.800	50	.054	.803	52	.051	.801	
	(.1, .1)		65	.053	.801	67	.054	.801	73	.053	.799	
	(.2, .2)		85	.053	.798	89	.054	.799	101	.053	.802	
	(.0, .0)		58	.054	.797	58	.054	.794	61	.053	.803	
	(.1, .1)		70	.055	.796	71	.055	.794	77	.053	.802	
	(.2, .2)		85	.055	.793	87	.056	.791	94	.054	.793	
	(.0, .0)		53	.055	.796	52	.056	.797	52	.054	.796	
	(.1, .1)		62	.057	.794	62	.056	.798	64	.055	.798	
	(.2, .2)		73	.057	.795	73	.056	.791	78	.055	.792	

Abbreviations: Dist., distribution; LG, log-logistic; LN, log-normal; WB, Weibull.

Table 2

Scenario 2 in which sample size, simulated empirical type I error (α), and power ($1 - \beta$) were based on 100,000 simulation runs for various distributions. The nominal type I error was 0.05 and power was 80%; the censoring distribution was uniform on $[t_d, t_d + t_f]$, where $t_d = 3$ and $t_f = 1$.

Scenario 2		$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
		S_0 vs S_1	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α
Dist.	(π_0, π_1)									
WB	(.0, .2)	.2 vs .2	47	.051	.804	40	.051	.807	28	.051
	(.1, .3)	.2 vs .2	53	.053	.800	49	.052	.804	44	.051
	(.2, .4)	.2 vs .2	56	.052	.796	53	.054	.799	50	.053
	(.0, .2)	.3 vs .3	68	.052	.802	58	.052	.805	38	.051
	(.1, .3)	.3 vs .3	72	.052	.798	64	.052	.804	52	.052
	(.2, .4)	.3 vs .3	72	.053	.798	66	.052	.796	58	.052
	(.0, .25)	.5 vs .5	85	.055	.797	73	.055	.802	48	.053
	(.1, .35)	.5 vs .5	82	.055	.795	72	.055	.799	52	.053
	(.2, .45)	.5 vs .5	77	.055	.797	68	.055	.795	53	.055
	(.0, .25)	.6 vs .6	122	.053	.795	104	.054	.795	68	.053
	(.1, .35)	.6 vs .6	114	.055	.792	99	.053	.794	70	.054
	(.2, .45)	.6 vs .6	105	.055	.793	92	.056	.793	69	.055
LN	(.0, .2)	.2 vs .2	34	.051	.812	43	.053	.806	48	.052
	(.1, .3)	.2 vs .2	47	.053	.807	51	.052	.805	54	.051
	(.2, .4)	.2 vs .2	52	.053	.802	55	.053	.805	57	.053
	(.0, .2)	.3 vs .3	45	.051	.809	60	.051	.803	69	.052
	(.1, .3)	.3 vs .3	56	.054	.804	66	.053	.801	72	.053
	(.2, .4)	.3 vs .3	61	.056	.802	68	.052	.798	73	.053
	(.0, .25)	.5 vs .5	50	.053	.806	72	.054	.801	84	.054
	(.1, .35)	.5 vs .5	54	.055	.801	71	.053	.798	81	.053
	(.2, .45)	.5 vs .5	55	.056	.805	68	.055	.794	76	.055
	(.0, .25)	.6 vs .6	65	.055	.801	99	.053	.800	119	.053
	(.1, .35)	.6 vs .6	68	.055	.803	96	.055	.795	112	.054
	(.2, .45)	.6 vs .6	67	.056	.798	89	.056	.792	103	.055
LG	(.0, .2)	.2 vs .2	51	.051	.807	48	.051	.807	43	.052
										.810

Scenario 2											
Dist.	(π_0, π_1)	S_0 vs S_1	$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
			n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$
	(.1, .3)	.2 vs .2	56	.051	.804	54	.053	.801	51	.052	.805
	(.2, .4)	.2 vs .2	58	.054	.797	57	.052	.800	55	.053	.802
	(.0, .2)	.3 vs .3	72	.052	.801	68	.052	.805	58	.054	.803
	(.1, .3)	.3 vs .3	75	.054	.798	71	.053	.797	65	.052	.804
	(.2, .4)	.3 vs .3	75	.053	.797	72	.054	.798	67	.053	.800
	(.0, .25)	.5 vs .5	89	.055	.800	81	.055	.796	66	.055	.802
	(.1, .35)	.5 vs .5	85	.054	.795	79	.054	.795	67	.054	.801
	(.2, .45)	.5 vs .5	80	.056	.798	74	.056	.793	65	.053	.800
	(.0, .25)	.6 vs .6	126	.055	.796	114	.054	.797	90	.054	.799
	(.1, .35)	.6 vs .6	118	.054	.796	108	.054	.796	88	.053	.800
	(.2, .45)	.6 vs .6	108	.056	.794	99	.055	.792	83	.054	.793

Abbreviations: Dist., distribution; LG, log-logistic; LN, log-normal; WB, Weibull.

Scenario 3 in which sample size, simulated empirical type I error (α), and power ($1 - \beta$) were based on 100,000 simulation runs for various distributions. The nominal type I error was 0.05 and power was 80%; the censoring distribution was uniform on $[t_f, t_d + t_f]$, where $t_d = 3$ and $t_f = 1$.

Table 3

Scenario 3			$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
Dist.	(π_0, π_1)	S_0 vs S_1	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$
WB	(.0, .1)	.2 vs .3	37	.051	.802	40	.051	.811	30	.049	.817
	(.1, .2)	.2 vs .3	48	.052	.799	49	.052	.803	50	.052	.807
	(.2, .3)	.2 vs .3	60	.052	.803	53	.053	.799	65	.052	.797
	(.0, .1)	.3 vs .4	50	.053	.807	58	.053	.806	38	.051	.811
	(.1, .2)	.3 vs .4	60	.054	.803	64	.053	.804	57	.052	.805
	(.2, .3)	.3 vs .4	70	.053	.797	66	.052	.798	72	.053	.800
	(.0, .15)	.5 vs .6	51	.055	.797	73	.054	.797	37	.052	.806
	(.1, .25)	.5 vs .6	56	.055	.797	72	.054	.800	45	.054	.805
	(.2, .35)	.5 vs .6	61	.056	.794	68	.055	.797	52	.054	.801
	(.0, .15)	.6 vs .7	55	.056	.793	104	.055	.794	40	.054	.805
LN	(.1, .25)	.6 vs .7	60	.058	.790	99	.056	.792	46	.056	.799
	(.2, .35)	.6 vs .7	65	.056	.793	92	.056	.792	52	.056	.797
	(.0, .1)	.2 vs .3	30	.050	.811	32	.052	.806	33	.051	.806
	(.1, .2)	.2 vs .3	44	.052	.799	43	.054	.802	43	.052	.802
	(.2, .3)	.2 vs .3	58	.052	.802	54	.054	.799	54	.054	.802
	(.0, .1)	.3 vs .4	40	.054	.807	44	.052	.808	46	.052	.806
	(.1, .2)	.3 vs .4	55	.053	.801	55	.053	.802	56	.052	.804
	(.2, .3)	.3 vs .4	70	.053	.801	66	.054	.797	66	.053	.799
	(.0, .15)	.5 vs .6	41	.053	.811	47	.054	.801	49	.055	.792
	(.1, .25)	.5 vs .6	49	.055	.803	53	.054	.800	55	.056	.796
LG	(.2, .35)	.5 vs .6	56	.055	.795	58	.056	.794	60	.055	.796
	(.0, .15)	.6 vs .7	45	.056	.805	51	.056	.794	54	.055	.790
	(.1, .25)	.6 vs .7	52	.057	.799	57	.057	.796	59	.055	.793
	(.2, .35)	.6 vs .7	59	.055	.800	62	.058	.795	64	.055	.790
	(.0, .1)	.2 vs .3	33	.052	.807	33	.051	.809	32	.052	.809

Scenario 3										$\kappa=0.5$			$\kappa=1$			$\kappa=2$		
Dist.	(π_0, π_1)	S_0	S_1	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$	n	α	$1 - \beta$
	(.1, .2)	.2 vs .3	.3	43	.054	.806	43	.052	.804	43	.052	.801	43	.052	.801	43	.052	.801
	(.2, .3)	.2 vs .3	.3	53	.053	.798	53	.052	.798	55	.054	.799	55	.054	.799	55	.054	.799
	(.0, .1)	.3 vs .4	.4	47	.053	.804	46	.053	.802	45	.051	.806	45	.051	.806	45	.051	.806
	(.1, .2)	.3 vs .4	.4	57	.054	.802	56	.052	.796	57	.052	.801	57	.052	.801	57	.052	.801
	(.2, .3)	.3 vs .4	.4	67	.054	.800	67	.053	.797	69	.054	.801	69	.054	.801	69	.054	.801
	(.0, .15)	.5 vs .6	.6	45	.056	.793	50	.056	.800	47	.055	.804	47	.055	.804	47	.055	.804
	(.1, .25)	.5 vs .6	.6	57	.055	.796	55	.056	.796	53	.054	.798	53	.054	.798	53	.054	.798
	(.2, .35)	.5 vs .6	.6	61	.055	.791	60	.056	.797	59	.056	.798	59	.056	.798	59	.056	.798
	(.0, .15)	.6 vs .7	.7	56	.056	.791	54	.056	.792	50	.056	.797	50	.056	.797	50	.056	.797
	(.1, .25)	.6 vs .7	.7	61	.056	.791	59	.057	.793	56	.056	.796	56	.056	.796	56	.056	.796
	(.2, .35)	.6 vs .7	.7	66	.057	.793	64	.056	.793	62	.056	.797	62	.056	.797	62	.056	.797

Abbreviations: Dist., distribution; LG, log-logistic; LN, log-normal; WB, Weibull.