Published in final edited form as:

Stat Med. 2012 December 20; 31(29): 3959–3971. doi:10.1002/sim.5465.

Sample Size Calculation for the Proportional Hazards Cure Model

Songfeng Wang^{a,*}, Jiajia Zhang^a, and Wenbin Lu^b

^aDepartment of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

^bDepartment of Statistics, North Carolina State University, Raleigh, NC 27695, USA

Abstract

In clinical trials with time-to-event endpoints, it is not uncommon to see a significant proportion of patients being cured (or long-term survivors), such as trials for the non-Hodgkins lymphoma disease. The popularly used sample size formula derived under the proportional hazards (PH) model may not be proper to design a survival trial with a cure fraction, since the PH model assumption may be violated. To account for a cure fraction, the PH cure model is widely used in practice, where a PH model is used for survival times of uncured patients and a logistic distribution is used for the probability of patients being cured. In this paper, we develop a sample size formula based on the PH cure model by investigating the asymptotic distributions of the standard weighted log-rank statistics under the null and local alternative hypotheses. The derived sample size formula under the PH cure model is more flexible since it can be used to test the differences in the short-term survival and/or cure fraction. Furthermore, the impacts of accrual methods and durations of accrual and follow-up periods on sample size calculation are also investigated as numerical examples. The results show that ignoring the cure rate in sample size calculation can lead to either underpowered or overpowered studies. The performance of the proposed formula is evaluated by simulation studies and an example is given to illustrate its application using data from a melanoma trial.

Keywords

Clinical trial; Proportional hazards cure model; Power; Sample size; Weighted log-rank test

1. Introduction

Sample size calculation is an important component in designing randomized controlled clinical trials with time-to-event endpoints. In order to get the accurate number of patients, we need to understand the disease of interest and define the aim of the study accordingly. For a survival trial, a main goal is to investigate the effects of treatment on survival times. Assuming constant hazard ratio between the treatment group and control group, the following sample size formula based on the proportional hazards (PH) model has been widely used in practice[1, 2]:

$$n = \frac{(Z_{\alpha/2} + Z_{\theta})^2}{p(1-p)\beta_0^2 P(\delta=1)}, \quad (1)$$

where $Z_{\alpha/2}$ and Z_{θ} are the upper $\alpha/2$ and θ percentiles of the standard normal distribution, respectively; p is the proportion of patients assigned to the treatment group; β_0 is the log-hazard ratio of the two treatments, and δ is the censoring indicator (1 for failure and 0 for censoring). It is worthwhile pointing out that standard survival models including the PH model is based on the assumption that all individuals under study are susceptible to the adverse event of interest, and they will experience the event eventually if there was no censoring.

However, in some recent clinical trials, a substantial proportion of patients may never experience the event, even after a sufficient follow-up period. For example, in late 1980s, more than 70% of children who had acute lymphoblastic leukemia (ALL) and were younger than 10 years old at diagnosis stayed disease-free for 4 years after treatment [3]. Due to advances in therapy and treatment, in survival trials for breast cancer, non-Hodgkin's lymphoma, leukemia, prostate cancer, melanoma, head cancer and neck cancer, there can be found significant proportions of long-term survivors, and these patients are usually referred to as "cured" [4]. For such trials, it is of interest to study the treatment effects on the probability of being cured and the distribution of survival times for susceptible patients.

The mixture cure model developed by Berkson and Gage [5] is known as the standard cure rate model, which models survival as a mixture of cure fraction and survival function of uncured patients, and provides a natural separation between the short-term and long-term survival. Another type of model is the bounded cumulative hazard (BCH) model (also known as the promotion time cure model) which is derived from a natural biological motivation and also accommodates cure rate and the PH structure [6, 7, 8, 9, 10]. An alternative model that has the proportional hazards structure when cure rates are present has also been suggested by Maller and Zhou [11], and extended by Zhao and Zhou to both continuous-time and discrete-time survival models [12, 13]. In general, when there are long-term survivors, the standard PH model can not be used since it fails to adjust for the cure fractions. Therefore, the standard sample size formula derived based on the PH model may not be valid. Some tests have been proposed when the interest is in the difference of cure rates, for example, a nonparametric linear rank test (equivalent to a weighted log-rank test) [14], a χ^2 test using the product-limit (PL) estimate [15], and a likelihood ratio test under parametric assumptions [16]. As for sample size calculation, Halpern and Brown [17, 18] develop a computer program to calculate the empirical power and sample size for the cure rate models based on Monte Carlo simulation. Recently, Ewell and Ibrahim [19] extend the work of Gray [14] by considering a general alternative that allows for treatment effects on both the short-term and long-term survival.

The PH cure model [20, 21, 22] has been widely used in practice to study the effects of treatment on the cure probability and/or survival distribution of uncured patients, where the PH model is used for survival times of uncured patients in the mixture cure model and a logistic distribution is used for the probability of patients being cured. However, there are limited discussions about the sample size and power calculation for the PH cure model, in particular, an analytical formula for the sample size calculation and its relationship with the standard sample size formula for the PH model. In this paper we study the large sample properties of the weighted log-rank test under both the null and local alternative hypotheses for the PH cure model, and then derive its associated sample size formula based on the asymptotic distributions. The proposed sample size formula naturally extends the standard one based on the PH model by allowing a cure fraction and provides more flexibility to test

the differences in the short-term survival and/or cure fraction that are of main interest in many survival trials.

In addition, the rate of patients' enrollment and the durations of accrual and follow-up periods are also important in planning a survival trial since they can have direct impact on the number of events to be observed, and thus influence the sample size calculation [23]. In this paper, we further investigate the impact of different accrual rate patterns, such as uniform accrual, increasing accrual and decreasing accrual, and the length of accrual and follow-up periods on sample size calculation based on the PH cure model.

The rest of this paper is organized as follows. In Section 2, we describe the notations and derive the sample size formula for the PH cure model. In Section 3, we give sample size formulae for three different accrual rates including the uniform, increasing and decreasing accruals. In Section 4, numerical examples are presented to investigate the impacts of the accrual rates, the durations of accrual and follow-up times, and the hazard ratios and/or cure rates on sample size calculation under different scenarios. In Section 5, we conduct extensive simulations to evaluate the performance of the proposed method by comparing the asymptotic power from the proposed formula to the empirical power calculated from simulated datasets. A real trial example is given in Section 6 to illustrate the application in practice. Discussions are given in Section 7.

2. Method

For a survival trial involving n study patients, let T_i and C_i denote the survival and censoring times of patient i respectively, and $A_i = 1/0$ denote the treatment group indicator (0 for control group and 1 for treatment group). The observed data then consist of $\{U_i, \Delta_i, A_i, i = 1, \dots, n\}$, where $U_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. It is commonly assumed that the survival time T_i is independent of C_i given A_i . Let $S_j^*(\cdot)$ denote the overall survival function and $\lambda_j^*(\cdot)$ denote its corresponding hazard function for treatment group j , $j = 0, 1$. Similarly, let $S_j(\cdot)$ denote the survival function of uncured patients and $\lambda_j(\cdot)$ denote its hazard function. The cure rate in group j is defined by π_j , where $0 < \pi_j < 1$.

Then for a mixture cure model we have $S_j^*(t) = \pi_j + (1 - \pi_j)S_j(t)$ and

$$\lambda_j^*(t) = \frac{(1 - \pi_j)S_j(t)}{\pi_j + (1 - \pi_j)S_j(t)} \lambda_j(t)$$

for $j = 0, 1$. The PH cure model assumes

$$\lambda_1(t) = e^{\beta_0} \lambda_0(t) \quad \text{and} \quad \text{logit}(\pi_1) = \text{logit}(\pi_0) + \gamma_0, \quad (2)$$

where β_0 is the log hazard ratio of uncured patients and γ_0 is the log odds ratio of cure rates between the two groups. Note that $\pi_0 = \pi_1 = 0$ corresponds to the standard PH model. For a survival trial with a proportion of patients cured, we are interested in testing $H_0: \pi_1 = \pi_0$ and $S_1(t) = S_0(t)$, which is equivalent to $H_0: \beta_0 = \gamma_0 = 0$. Various alternative hypotheses are of our interest: $H_a: \beta_0 \neq 0, \gamma_0 = 0$, the differences are in both the short-term survival and the cure fraction; $H_a: \beta_0 \neq 0, \gamma_0 = 0$, the difference is in the short-term survival but not in the cure fraction; $H_a: \beta_0 = 0, \gamma_0 \neq 0$, the difference is in the cure fraction but not in the short-term survival.

Next we derive the asymptotical distributions of the weighted log-rank test statistic under both the null and local alternative hypotheses. We first introduce the standard counting process notations. Define $N_i(u) = I(U_i \leq u, \Delta_i = 1)$ and $Y_i(u) = I(U_i > u)$ for $i = 1, \dots, n$. The weighted log-rank test statistic is given by

$$T_n = \sum_{i=1}^n \int_0^\tau W_n(u) \{A_i - \bar{A}(u)\} dN_i(u),$$

where τ is a constant such as the maximum follow-up, $\bar{A}(u) = \sum_{i=1}^n A_i Y_i(u) / \sum_{i=1}^n Y_i(u)$ and $W_n(u)$ is a predictable weight function converging to a deterministic function $w(u)$. Various weight functions can be used, for example, the log-rank weight ($W_n(u) \equiv 1$) [24, 25], Gehan-Wilcoxon weight ($W_n(u) = n^{-1} \sum_{i=1}^n Y_i(u)$) [26] and Peto-Prentice weight ($W_n(u) \approx \hat{S}(u)e$) [27, 28].

Given the PH cure model (2), the hazard function of patient i is given by

$$\tilde{\lambda}_i(t) = \frac{\frac{1-\pi_0}{1-\pi_0+\pi_0 e^{\gamma_0 A_i}} \{S_0(t)\}^{\exp(\beta_0 A_i)}}{\frac{\pi_0 e^{\gamma_0 A_i}}{1-\pi_0+\pi_0 e^{\gamma_0 A_i}} + \frac{1-\pi_0}{1-\pi_0+\pi_0 e^{\gamma_0 A_i}} \{S_0(t)\}^{\exp(\beta_0 A_i)}} \lambda_0(t) e^{\beta_0 A_i}, \quad 0 < t < \infty.$$

Under the null hypothesis $H_0: \beta_0 = \gamma_0 = 0$, we have $\tilde{\lambda}_i(\cdot) = \lambda_0^*(\cdot)$ for all i . Then it is easy to show that $T_n = \sum_{i=1}^n \int_0^\tau W_n(u) \{A_i - \bar{A}(u)\} dM_i(u)$ under the null hypothesis, where $M_i(t) = N_i(t) - \int_0^t I(U_i \geq s) \tilde{\lambda}_i(s) ds$. Based on the counting process theory [29], it can be shown that $M_i(t)$ is a mean-zero martingale process. By the martingale central limit theorem [29], we have $n^{-1/2} T_n$ converges in distribution to normal with mean of 0 and variance of $\sigma^2 = \int_0^\tau w^2(u) E_{H_0} [\{A_i - \mu_A(u)\}^2 I(U_i \geq u)] \lambda_0^*(u) du$, where $\mu_A(u) = \lim_{n \rightarrow \infty} \bar{A}(u)$ and E_{H_0} is the expectation taken with respect to the distribution under the null hypothesis.

To derive the power and sample size calculation for the PH cure model, we need to consider a series of local alternatives. Specifically, we consider $H_{a,n}: \beta_0 = \beta_a / \sqrt{n}, \gamma_0 = \gamma_a / \sqrt{n}$. The weighted log-rank test statistic T_n can be rewritten as

$$T_n = \sum_{i=1}^n \int_0^\tau W_n(u) \{A_i - \bar{A}(u)\} dM_i(u) + \sum_{i=1}^n \int_0^\tau W_n(u) \{A_i - \bar{A}(u)\} I(U_i \leq u) \tilde{\lambda}_i(u) du \equiv T_{n,1} + T_{n,2}.$$

By the martingale central limit theorem, we can show that $n^{-1/2} T_{n,1}$ converges in distribution to $N(0, \sigma^2)$ under the local alternative $H_{a,n}$. Moreover, we have that $n^{-1/2} T_{n,2}$ converges in probability to a constant $\xi(w, \gamma_a, \beta_a)$, where

$$\xi(w, \gamma, \beta) = - \int_0^\tau w(u) E_{H_0} [\{A_i - \mu_A(u)\}^2 I(U_i \geq u)] \left[\gamma \frac{\pi_0}{S_0^*(u)} - \beta \left\{ 1 - \frac{\pi_0 \Lambda_0(u)}{S_0^*(u)} \right\} \right] \lambda_0^*(u) du,$$

and $\Lambda_0(u) = \int_0^u \lambda_0(s) ds$. The details of the derivation are given in Appendix. Therefore, under the local alternative hypothesis, we have that $n^{-1/2} T_n$ converges in distribution to $N\{\xi(w, \gamma_a, \beta_a), \sigma^2\}$.

Based on the above results, for a weighted log rank test with a significance level of α to achieve a power of $1 - \theta$, it must satisfy

$$1 - \theta = \Phi\left(\frac{|\xi(w, \gamma_a, \beta_a)|}{\sigma} - Z_{\alpha/2}\right),$$

where Φ is the cumulative distribution function of a standard normal random variable. Based on the relationship $\beta_0 = \beta_a / \sqrt{n}$ and $\gamma_0 = \gamma_a / \sqrt{n}$, the total sample size can then be determined by

$$n = \frac{(Z_\theta + Z_{\alpha/2})^2 \sigma^2}{\{\xi(w, \gamma_0, \beta_0)\}^2}. \quad (3)$$

Special case

Consider independent censoring, i.e. the censoring time is independent of the survival time and treatment group indicator. Let $S_C(t) = P(C > t)$ denote the survival function of the censoring time. Let $p = P(A_i = 1)$, the proportion of patients in the treatment group. For the standard log-rank weight ($w(u) = 1$), we have

$$\sigma^2 = p(1-p) \int_0^\tau S_C(t) (1 - \pi_0) S_0(t) \lambda_0(t) dt$$

$$\xi(w, \gamma_0, \beta_0) = -p(1-p) \beta_0 \int_0^\tau m(\gamma_0, \beta_0, \pi_0) S_C(t) (1 - \pi_0) S_0(t) \lambda_0(t) dt,$$

where $m(\gamma_0, \beta_0, \pi_0) = \pi_0 \{\gamma_0 / \beta_0 + \Lambda_0(t)\} / S_0^*(t) - 1$. The sample size formula is then

$$n = \frac{(Z_\theta + Z_{\alpha/2})^2 \int_0^\infty S_C(t) f_0(t) dt}{p(1-p) \beta_0^2 (1 - \pi_0) \left\{ \int_0^\infty m(\gamma_0, \beta_0, \pi_0) S_C(t) f_0(t) dt \right\}^2}, \quad (4)$$

where $f_0(t) = \lambda_0(t) S_0(t)$ is the density function of survival times for uncured patients in the control group.

When $\pi_0 = 0$, the above sample size formula is reduced to the standard PH model sample size formula as given in (1). Note that the integrations in the formula may need to be computed numerically.

3. Sample Size Calculation under Different Accrual Rate Functions

In survival trials, patients are usually enrolled over a period of time and not all patients have the same follow-up time. This is referred to as staggered entry or extended accrual [30]. The

most common entry method is uniform accrual, which assumes that patients enter the study at a constant rate over the accrual period. However, in practice the rate is hardly constant or piece-wise constant throughout the accrual period [31]. For example, in a Phase III trial comparing adjuvant treatments for colon cancer, the daily accrual is not constant and shows an increase pattern during the early period [32]. Barthel et al. (2006) [33] also give examples of trials where the actual accrual rates are higher in the later period of accrual. On the other hand, for trials dealing with small population or rare diseases, the limited number of cases can be exhausted quickly and further enrollment may be slow down [34]. Therefore, in this situation the accrual rate will decrease with the accrual going on. Maki [34] suggests different entry scenarios and investigates their relationship with sample size under the Weibull model. Here we adopt the uniform, increasing and decreasing accrual rate functions defined by Maki [34], and explore their effects on power and sample size under the PH cure model. More complicated accrual functions can also be easily incorporated in the proposed sample size calculation.

Let τ_a denote the accrual period and τ_f denote the follow-up time. Then the total study length is $\tau = \tau_a + \tau_f$. Let $g(t)$ denote the probability density function of accrual times. For simplicity, here we assume that the only censoring is due to administrative censoring at time τ , and there is no loss to follow-up or competing risks. The probability density functions $g(t)$ and their associated survival functions $S_C(t)$ of the censoring variable for the uniform, increasing and decreasing accruals are summarized in Table 1. Plugging the defined survival functions $S_C(t)$ into formula (4), we obtain the sample size formulae for the uniform, increasing and decreasing accruals.

An Example

In the special case considered above, assume a uniform accrual and an exponential distribution with the rate of λ for survival times of uncured patients in the control group, i.e. $\lambda_0(t) = \lambda$, $\Lambda_0(t) = \lambda t$ and $S_0(t) = e^{-\lambda t}$. Then the sample size is calculated as

$$n = \frac{(Z_{\theta} + Z_{\alpha/2})^2 \left(\int_0^{\tau_f} e^{-\lambda t} dt + \int_{\tau_f}^{\tau_a + \tau_f} \frac{\tau_a + \tau_f - t}{\tau_a} e^{-\lambda t} dt \right)}{p(1-p)\beta_0^2(1-\pi_0)\lambda \left\{ \int_0^{\tau_f} m(\gamma_0, \beta_0, \pi_0) e^{-\lambda t} dt + \int_{\tau_f}^{\tau_a + \tau_f} \frac{\tau_a + \tau_f - t}{\tau_a} m(\gamma_0, \beta_0, \pi_0) e^{-\lambda t} dt \right\}^2}, \quad (5)$$

where $m(\gamma_0, \beta_0, \pi_0) = \frac{(\gamma_0/\beta_0 + \lambda t)\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda t}} - 1$.

The numerical integral can be realized by the “QUAD” subroutine in SAS IML or the “integrate” function in R.

More flexible parametric distributions such as Weibull, log-normal or gamma distributions can also be considered. Parametric distributions are easy to use but they may be sensitive to model misspecification, and thus lead to erroneous sample size calculation. When there are historical data available, nonparametric estimators of survival function such as the life-table or Kaplan-Meier (KM) estimators can be used, but they need to be adjusted for the cure fraction [11].

4. Numerical Example

The following settings are used in the numerical example: (1) an exponential distribution is assumed for survival times of uncured patients in the control group with $\lambda_0(t) \equiv 0.5$; (2) an equal allocation is used, i.e. $p = 0.5$; (3) a two-sided log-rank test with a significance level of 0.05 is considered.

4.1. Impact of accrual and follow-up durations

We investigate the impacts of accrual and follow-up durations under three different accrual rate functions for sample size calculation under the PH cure model. We set the log hazard ratio $\beta_0 = \log(2/3)$, and the cure rates $\pi_0 = 0.1$ and $\pi_1 = 0.2$ in the control and treatment groups, respectively (which gives $\gamma_0 = 0.811$). In Figure 1, sample sizes for the log-rank test to achieve a power of 0.9 under the PH cure model are plotted against the lengths of accrual and follow-up durations, which vary from 0.2 to 6 years.

From Figure 1, we find that the impact of the accrual and follow-up periods on sample size is quite complicated under the PH cure model, which may be nonlinear and nonmonotone. First, for all three accrual rate functions, the relationship between the follow-up period and sample size is nonlinear. For example, when the accrual period is 3 years (see Figure 2(a)), sample size for all three accrual rate functions first decreases as the follow-up period increases, and then it starts to increase after the follow-up period passes a certain time point. Second, the relationship between the accrual period and sample size depends on the follow-up time. When the follow-up time is less than a certain value (2.5 years for uniform accrual, 3.4 years for increasing accrual and 1.6 years for decreasing accrual in this example), sample size monotonously decreases as the accrual period increases. For example, in Figure 2(b), we plot sample size against the accrual period given one-year follow-up time for all three accrual rate functions. They all show monotone decreasing patterns. On the other hand, when the follow-up time is longer than these values, sample size first decreases and then increases as the accrual period increases. To see this, in Figure 2(c), we plot sample size against the accrual period for all three accrual rate functions with follow-up time of 2.8 years for uniform accrual, 3.5 years for increasing accrual, and 1.8 years for decreasing accrual, respectively.

4.2. Impact of hazard ratio and cure rates

We investigate the impact of the log hazard ratio and cure rates on sample size for different accrual rates under the PH cure model. Here we consider the case with an accrual period of 3 years and a follow-up time of 4 years.

Scenario 1: $\beta_0 \neq 0$, $\gamma_0 \neq 0$ —We first study the effect of β_0 . Here we set $\pi_0 = 0.1$ and $\pi_1 = 0.2$, which gives $\gamma_0 = 0.811$. The mean survival time for uncured patients in the control group is assumed to be 2 years, and in the treatment group it is assumed to vary from 2.5 to 4 years with an increment of 0.5 year. Therefore, the log hazard ratio β_0 takes the value of $\log(2/2.5)$, $\log(2/3)$, $\log(2/3.5)$ and $\log(2/4)$, accordingly. Sample sizes under all three accrual rates are plotted in Figure 3. First, for all three accrual rates, sample size always increases as the log hazard ratio increases. This indicates that more patients are needed to detect a smaller difference in the hazard functions between two groups, which is in line with the standard PH model. Second, it is noted that the increasing accrual requires the most patients, while the decreasing accrual needs the least patients. For example, when $\beta_0 = \log(2/2.5)$, to maintain a power of 0.9, the increasing accrual requires 437 patients, the uniform accrual needs 429 patients, while the decreasing accrual needs only 421 patients.

Next we investigate the impact of cure rates π_0 and π_1 . The mean survival time for the uncured patients in the control group is assumed to be 2 years and the log hazard ratio is assumed to be $\beta_0 = \log(2/3)$. The cure rate for the control group is assumed to be $\pi_0 = 0.1$ while for the treatment group it varies from 0.15 to 0.3 with an increment of 0.05. So γ_0 increases accordingly from 0.463 to 1.350. Sample sizes under different accrual rates are plotted in Figure 4. Based on Figure 4, we have the following observations. First, when power increases, sample size also increases regardless of cure rates. Under all the three accrual rates, sample size always increases with the decrease of γ_0 , which means a study

needs more patients to detect a smaller difference between cure rates, and this is what we expect to see. Second, the impact of accrual rates depends on γ_0 (or π_1). For example, when $\gamma_0 = 0.811$, to maintain a power of 0.9, all the three accrual rates require the same sample size of 221. However, when $\gamma_0 = 0.463$, the uniform accrual requires 312 patients, the increasing accrual needs 307 patients while the decreasing accrual requires 317 patients. Third, compared to the uniform accrual, in general, the decreasing accrual needs a larger sample size but the increasing accrual needs a smaller sample size, although the differences may be subtle. Lachin and Foulkes [35] point out that for a fixed sample size, compared to the uniform accrual, the convex accrual function (increasing accrual) has greater power while the concave accrual function (decreasing accrual) has less power, which is consistent with our findings here.

For comparison purpose, in Figure 4 we also plot the sample sizes based on the standard PH model using formula (1), ignoring the cure rates. We can see that the calculated sample sizes from the PH model are deviated from those based on the PH cure model, and the deviations depend on the value of γ_0 . When γ_0 is large ($\gamma_0 = 0.811, 1.099$, or 1.350), the sample sizes from the PH model are larger than those from the PH cure model, while when γ_0 gets smaller ($\gamma_0 = 0.463$), the sample sizes from the PH model become smaller than those from the PH cure model. One interpretation is that when the difference in the cure rates between two groups is large (i.e. γ_0 is large), it is easier for the log-rank test to detect the difference between two groups. Therefore, the sample size derived from the PH cure model requires less patients than the one based on the standard PH model by taking into account the cure fractions in calculation. When the difference in the cure rates between two groups is small, due to the presence of cured patients, the probability of observing an event of interest becomes smaller than expected, and thus the sample size from the standard PH model tends to underestimate the number of totally required patients.

Scenario 2: $\beta_0 \neq 0, \gamma_0 = 0$ —There is no difference between the cure rates, i.e. $\gamma_0 = 0$, and we test the difference in the hazard functions adjusting for different cure fractions. Here the log hazard ratio is set as $\beta_0 = \log(2/3)$, and the cure rates in both groups vary from 0.1 to 0.4 with an increase of 0.1. Sample sizes under different accrual rates are plotted in Figure 5. We make the following observations. First, the required sample sizes increase with the increase of cure rates under all three accrual rates. This is expected since with higher cure rates, less patients will experience the event of interest. As a result, the PH cure model requires more patients than the standard PH model, and the difference depends on the cure rates. For example, under uniform accrual, to achieve a power of 0.9, the sample size from the PH model is 275. However, for the PH cure model, when $\pi_0 = \pi_1 = 0.1$ the sample size is 557; when $\pi_0 = \pi_1 = 0.2$ the sample size is 949; when $\pi_0 = \pi_1 = 0.3$ the sample size is 1533; when $\pi_0 = \pi_1 = 0.4$ the sample size is 2435. Second, to achieve a specific power, the decreasing accrual requires the most patients, followed by the uniform accrual and then the increasing accrual. For example, when $\pi_0 = \pi_1 = 0.2$, to achieve a power of 0.9, the decreasing accrual requires 1031 patients, the uniform accrual requires 949 patients, while the increasing accrual requires only 875 patients.

Scenario 3: $\beta_0 = 0, \gamma_0 \neq 0$ —There is no difference between the two hazard functions, i.e. $\beta_0 = 0$, and we test the difference in the cure rates (long-term effects). Sample sizes under different accrual rates are plotted in Figure 6, where $\pi_0 = 0.1$ and π_1 varies from 0.2 to 0.5 with an increment of 0.1. Since the sample size formula (1) for the standard PH model gives only the significance level α when $\beta_0 = 0$, only the sample sizes from the PH cure model are plotted. It can be seen from Figure 6 that for all three accrual rates, sample size increases with the increase of γ_0 . To maintain a specific power, the increasing accrual requires the most patients, followed by the uniform accrual, and then the decreasing accrual. For example, when $\pi_1 = 0.2$, to maintain a power of 0.9, the increasing accrual requires 1746

patients, the uniform accrual requires 1604 patients, but the decreasing accrual requires only 1480 patients. It is also noted that the sample sizes required in this scenario are much larger than those in scenario 2 where $\beta_0 = 0$, $\gamma_0 = 0$, when the differences between the cure rates are similar to the difference between the hazards. For example, when $\pi_0 = \pi_1 = 0.1$, $\gamma_0 = 0$, $\beta_0 = \log(2/3) = -0.405$ in scenario 2, the required sample size to achieve a power of 0.9 for the uniform accrual is 557. However, in the current scenario, when $\pi_0 = 0.1$, $\pi_1 = 0.143$, $\gamma_0 = 0.405$, $\beta_0 = 0$, the sample size is 6428, which is ten times larger. This implies that it is much harder to test the difference in the cure rates than that in the hazard functions. Accordingly, the sample size for the long-term effect (cure) is greater than that for a similar short-term effect (survival).

5. Simulation Study

In this section, we conduct simulation studies to investigate the performance of the proposed sample size formula based on the PH cure model, by comparing the theoretical asymptotic power derived from the new sample size formula to the empirical power of the standard log-rank test. For comparisons, we also include the asymptotic power derived from the sample size formula based on the standard PH model. Moreover, we compute the sample size based on the empirical power for both the PH cure model and the PH model, and compare them to the true values. We assume a two-arm clinical trial in which patients are uniformly accrued and are assigned to each treatment arm with equal probability. Standard exponential survival times are assumed for uncured patients on both treatment arms. We vary hazard rates of uncured patients and cure rates to allow the sample size formula to be examined under three scenarios: (1) there are differences in both hazard rates and cure rates, i.e., $\beta_0 \neq 0$, $\gamma_0 \neq 0$; (2) there are only differences between hazard rates, i.e., $\beta_0 \neq 0$, $\gamma_0 = 0$; (3) there are only differences between cure rates, i.e., $\beta_0 = 0$, $\gamma_0 \neq 0$. The hazard rates for uncured patients vary from 0.3 to 0.5 while the cure rates vary from 0.1 to 0.6. It is always assumed that treatment arm 1 has lower hazard rate and higher cure rate than treatment arm 0 if there are differences. In addition, we consider an accrual period of 1 year and a follow-up period of either 3 or 4 years. The empirical powers are calculated as the rejection rates of two-sided log-rank tests with the size of 0.05 via simulations with 10^6 replicates.

The simulation results are summarized in Table 2. For the PH cure model, we observe that the asymptotic powers derived from our sample size formula are quite close to those empirical powers under all the three scenarios, and the differences between them become smaller when the sample sizes increase. A similar conclusion can be made regarding the estimated sample size to achieve the empirical power and the actual sample size for the PH cure model. As for the asymptotic powers derived from the sample size formula based on the standard PH model and the corresponding sample sizes calculated to achieve the empirical power, we make the following observations. First, when there are differences in both cure rates and hazard rates (scenario 1), the associated asymptotic powers are always lower than the empirical powers, and the differences get larger when the cure rates increase. That is, when there is a difference in cure rates between two groups, by ignoring the cure fraction, the standard sample size formula may underestimate the power since it only considers the difference in the short-term survival. Due to this reason, the design based on the standard PH model may require more patients than necessary in this situation, as shown in Table 2. Second, when there are only differences between hazard rates (scenario 2), the associated asymptotic powers are reasonably close to the empirical powers when the cure rates are low ($\pi_0 = \pi_1 = 0.1$), but they are much greater than the empirical powers when cure rates are high ($\pi_0 = \pi_1 = 0.5$). The possible reason is that when the relatively large cure rates exist and are the same in the two groups, the standard sample size formula based on the PH model tends to overestimate the power since in the presence of cure fraction, it can only observe less number of failures than expected. So the actual power of the log-rank test is

lower than the desired level if the cure fraction is ignored. Thus, the design based on the PH model requires less patients than necessary, as we can see that the sample sizes based on the empirical powers from the PH model are much smaller than those from the PH cure model. Finally, when there are only differences between cure rates (scenario e3), the associated asymptotic powers are all equal to the significance level of 0.05, since we assume no difference between hazard rates in this situation. As a result, the sample sizes under the PH model can not be calculated.

6. Trial Example

In this section, we illustrate the application of our sample size formula using data from the Eastern Cooperative Oncology Group (ECOG) trial e1684 as an example. Note that our intention here is not to re-design the trial but to show the sample size requirement under the PH cure model when the hazard ratio and cure rates are estimated from the e1684 data. The ECOG trial e1684 was a two-arm phase III clinical trial comparing high dose interferon alpha-2b with an observation arm. The primary endpoint was relapse-free survival (RFS), with RFS defined as the time from randomization until progression of the tumor or death. The trial was designed to detect a 50% improvement in median RFS from 1.5 to 2.25 years, with an accrual period of 4 years and a follow-up period of 3 years. A sample size of 286 patients was considered, which is expected to yield a power of 83% based on the PH model.

This dataset has been extensively studied using the PH cure models [36, 10, 37], because the PH cure model can capture the long and stable plateau of the Kaplan-Meier curves well. We fit the data using the PH cure model with treatment as the covariate for the purpose of sample size calculation. Two estimation methods are considered in this section. The semiparametric estimation method [21], which uses an EM algorithm with a nonparametric estimator of the baseline survival function, is referred to as Cox PH cure model. The parametric estimation approach [11], which assumes a Weibull distribution for the survival distribution of uncured patients, is referred to as Weibull PH cure model. For Cox PH cure model, the log-hazard ratio is estimated as -0.14 and cure rates for the observation arm and the interferon arm are estimated as 0.225 and 0.331, respectively. For Weibull PH cure model, the estimated log-hazard ratio is -0.126, and the estimated cure rates are 0.242 for the observation arm and 0.351 for the interferon arm. As shown in Figure 7, the estimated survival curves based on Cox PH cure model and Weibull mixture cure model are reasonably close to the Kaplan-Meier estimates for each arm. The goodness-of-fit for a fitted model is further evaluated by the correlation coefficient (r) between the Kaplan-Meier curve and the fitted survival curve [11, 37]. For Cox PH cure model, $r = 0.997$ for both treatment arms; while for Weibull PH cure model, $r = 0.998$ for the interferon arm and 0.986 for the observation arm, indicating the adequate fit of both considered cure models. The results are consistent with [37].

Assuming equal number of patients on each arm, a uniform accrual period of 4 years and a follow-up period of 3 years, and taking the estimates from either Cox PH cure model or Weibull PH cure model as the effect sizes under the alternative hypothesis, we calculate the sample size using the proposed method. In order to achieve a power of 83%, based on the estimates from Cox PH cure model, a sample size of 483 is required; while the required sample size is 551 based on the estimates from Weibull PH cure model. This implies that the sample size of 286 calculated based on the standard PH model assumption may lead to a underpowered trial if there is a cure fraction. In fact, we calculated the theoretical power of the log-rank test with sample size 286, which is 0.61 under Cox PH cure model and 0.56 under Weibull PH cure model, and both are far below the desired power of 0.83.

7. Discussion

In this article, we propose a sample size calculation method under the PH cure model for survival trials, which naturally extends the commonly used sample size formula under the standard PH model by allowing a cure fraction and provides flexible tools to design studies that aim to test the treatment effects in the short-term and/or the long-term survival. The extensive simulations show that the proposed sample size formula performs reasonably well under different scenarios, while the sample size formula based on the standard PH model may lead to an either underpowered or overpowered study when there exists a cure rate. There are several alternatives of cure models in the literature, as mentioned in the introduction section, and it is also of interest to investigate the sample size calculation under alternative cure models in our future research.

Acknowledgments

We sincerely thank the Associate Editor and the two reviewers for their valuable comments and suggestions for improving the manuscript. The project is supported by National Institutes of Health (NIH) research grants R03CA150077 and R01CA140632.

References

1. Schoenfeld DA. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981; 68(1):316–319. [10.1093/biomet/68.1.316](https://doi.org/10.1093/biomet/68.1.316)
2. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983; 39:499–503. [PubMed: 6354290]
3. Bleyer WA. Acute lymphoblastic leukemia in children. advances and prospectus. *Cancer*. 1990; 65:689–695. [PubMed: 2405996]
4. Gu Y, Sinha D, Banerjee S. Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis*. 2011; 17(1):123–134. [PubMed: 20521166]
5. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*. 1952; 47(259):501–515.
6. Yakovlev, AY.; Tsodikov, AD.; Asselain, B. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific Pub Co Inc; Singapore: 1996.
7. Tsodikov A. A proportional hazards model taking account of long-term survivors. *Biometrics*. 1998; 54(4):1508–1516. [PubMed: 9883549]
8. Chen MH, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*. 1999; 94:909–919.
9. Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data. *Journal of the American Statistical Association*. 2003; 98(464):1063–1078. [10.1198/01622145030000001007](https://doi.org/10.1198/01622145030000001007) [PubMed: 21151838]
10. Ibrahim, JG.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. Springer-Verlag Inc; New York: 2001.
11. Maller, RA.; Zhou, X. *Survival Analysis with Long-term Survivors*. John Wiley & Sons; Chichester, UK: 1996.
12. Zhao X, Zhou X. Proportional hazards models for survival data with long-term survivors. *Statistics & Probability Letters*. 2006; 76(15):1685–1693. [10.1016/j.spl.2006.04.018](https://doi.org/10.1016/j.spl.2006.04.018)
13. Zhao X, Zhou X. Discrete-time survival models with long-term survivors. *Statistics in Medicine*. 2008; 27(8):1261–1281. [10.1002/sim.3018](https://doi.org/10.1002/sim.3018) [PubMed: 17676708]
14. Gray RJ, Tsiatis AA. A linear rank test for use when the main interest is in differences in cure rates. *Biometrics*. 1989; 45:899–904. [PubMed: 2790128]
15. Spisto R, Sather HN, Baker SA. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*. 1992; 48:87–99. [PubMed: 1581495]
16. Laska EM, Meisner MJ. Nonparametric estimation and testing in a cure model. *Biometrics*. 1992; 48:1223–1234. [PubMed: 1290799]

17. Halpern J, Wm B, Jun B. 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]
18. Halpern J, Wm B, Jun B. Cure rate models: Power of the logrank and generalized Wilcoxon tests. *Statistics in Medicine*. 1987; 6:483–489. [PubMed: 3629049]
19. 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]
20. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992; 79:531–541.
21. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000; 56(1):237–243. [PubMed: 10783801]
22. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56(1): 227–236.10.1111/j.0006-341X.2000.00227.x [PubMed: 10783800]
23. Case LD, Morgan TM. Duration of accrual and follow-up for two-stage clinical trials. *Lifetime Data Analysis*. 2001; 7:21–37.10.1023/A:1009621009283 [PubMed: 11280845]
24. Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. *Biometrika*. 1977; 64(1):156–160.10.1093/biomet/64.1.156
25. Breslow N. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*. 1970; 57(3):579–594.10.1093/biomet/57.3.579
26. Gehan EA. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*. 1965; 52:203–223. [PubMed: 14341275]
27. Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society, Series A (General)*. 1972; 135:185–206.
28. Prentice RL. Linear rank tests with right censored data. *Biometrika*. 1978; 65(1):167–179.10.1093/biomet/65.1.167
29. Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. Wiley; New York: 1991.
30. Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*. 1988; 44(1):229–241. [PubMed: 3358991]
31. Zhang X, Long Q. Stochastic modeling and prediction for accrual in clinical trials. *Statistics in Medicine*. 2010; 29:649–658. [PubMed: 20082363]
32. Van Cutsem E, Labianca R, Bodoky G, Barone C, Aranda E, Nordlinger B, Topham C, Tabernero J, Andr T, Sobrero AF, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *Journal of Clinical Oncology*. 2009; 27(19):3117–3125.10.1200/JCO.2008.21.6663 [PubMed: 19451425]
33. Barthel FMS, Babiker A, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine*. 2006; 25(15):2521–2542.10.1002/sim.2517 [PubMed: 16479563]
34. Maki E. Power and sample size considerations in clinical trials with competing risk endpoints. *Pharmaceutical Statistics*. 2006; 5(3):159–171. [PubMed: 17080750]
35. Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification (Corr: V42 p1009). *Biometrics*. 1986; 42:507–519. [PubMed: 3567285]
36. Ibrahim JG, Chen MH, Sinha D. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*. 2001; 57(2):383–388. [PubMed: 11414560]
37. Corbière F, Joly P. A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine*. 2007; 85:173–180.10.1016/j.cmpb.2006.10.008 [PubMed: 17157948]

9. Appendix

By Taylor expansion, we have

$$\begin{aligned}\tilde{\lambda}_i(t) &= -\frac{\pi_0(1-\pi_0)S_0(t)}{\{\pi_0+(1-\pi_0)S_0(t)\}^2}\lambda_0(t)A_i\gamma_0 + \left[\frac{(1-\pi_0)S_0(t)}{\pi_0+(1-\pi_0)S_0(t)} - \frac{\pi_0(1-\pi_0)S_0(t)\Lambda_0(t)}{\{\pi_0+(1-\pi_0)S_0(t)\}^2} \right] \lambda_0(t)A_i\beta_0 + \frac{(1-\pi_0)S_0(t)}{\pi_0+(1-\pi_0)S_0(t)}\lambda_0(t) + O(1/n) \\ &= -\frac{\pi_0}{\pi_0+(1-\pi_0)S_0(t)}\lambda_0^*(t)A_i\gamma_0 + \left[1 - \frac{\pi_0\Lambda_0(t)}{\pi_0+(1-\pi_0)S_0(t)} \right] \lambda_0^*(t)A_i\beta_0 + \lambda_0^*(t) + O(1/n).\end{aligned}$$

Since $W_n(u) \rightarrow w(u)$, $n^{1/2} \gamma_0 \rightarrow \gamma_a$ and $n^{1/2} \beta_0 \rightarrow \beta_a$, it follows that

$$\begin{aligned}T_{n,2} &= n^{-1/2} \sum_{i=1}^n \int_0^\tau W_n(u) \{A_i - \bar{A}(u)\} I(U_i \geq u) \left\{ -\frac{\pi_0}{\pi_0+(1-\pi_0)S_0(u)} \lambda_0^*(u) A_i \gamma_0 + \left[1 - \frac{\pi_0 \Lambda_0(u)}{\pi_0+(1-\pi_0)S_0(u)} \right] \lambda_0^*(u) A_i \beta_0 + \lambda_0^*(u) + O(1/n) \right\} du \\ &\rightarrow P - \int_0^\tau w(u) E_{H_0} [\{A_i - \mu_A(u)\}^2 I(U_i \geq u)] \left\{ \gamma_a \frac{\pi_0}{\pi_0+(1-\pi_0)S_0(u)} - \beta_a \left[1 - \frac{\pi_0 \Lambda_0(u)}{\pi_0+(1-\pi_0)S_0(u)} \right] \right\} \lambda_0^*(u) du \\ &= - \int_0^\tau w(u) E_{H_0} [\{A_i - \mu_A(u)\}^2 I(U_i \geq u)] \left[\gamma_a \frac{\pi_0}{S_0^*(u)} - \beta_a \left\{ 1 - \frac{\pi_0 \Lambda_0(u)}{S_0^*(u)} \right\} \right] \lambda_0^*(u) du \\ &= \tau(w, \gamma_a, \beta_a).\end{aligned}$$

Figure 1.
Sample size for different accrual and follow-up durations under three accrual rate functions

Figure 2.

(a) Relationship between the follow-up duration and sample size for 3 years of accrual. (b) Relationship between the accrual duration and sample size for 1 year of follow-up. (c) Relationship between the accrual duration and sample size: 2.8 years of follow-up for uniform accrual, 3.5 years of follow-up for increasing accrual, 1.8 years of follow-up for decreasing accrual.

Figure 3.
Sample sizes for different β_0 under three accrual rate functions

Figure 4.
Sample sizes for different γ_0 under three accrual rate functions

Figure 5.
Sample sizes for $\gamma_0 = 0$ under three accrual rate functions

Figure 6.
Sample sizes for $\beta_0 = 0$ under three accrual rate functions

Figure 7. Marginal survival curves for ECOG e1684 data. Upper lines are for the interferon arm and lower lines are for the observation arm

Table 1

Density functions $g(t)$ of accrual times and the corresponding survival functions $S_C(t)$ of censoring times.

accrual	$g(t)$	$S_C(t)$
uniform	$g(t) = \begin{cases} \frac{1}{\tau_a} & \text{if } 0 < t \leq \tau_a \\ 0 & \text{otherwise} \end{cases}$	$S_C(t) = \begin{cases} 1 & \text{if } t \leq \tau_f \\ \frac{\tau_a + \tau_f - t}{\tau_a} & \text{if } \tau_f < t \leq \tau_a + \tau_f \\ 0 & \text{if } t > \tau_a + \tau_f \end{cases}$
increasing	$g(t) = \begin{cases} \frac{2t}{\tau_a^2} & \text{if } 0 < t \leq \tau_a \\ 0 & \text{otherwise} \end{cases}$	$S_C(t) = \begin{cases} 1 & \text{if } t \leq \tau_f \\ \frac{(\tau_a + \tau_f - t)^2}{\tau_a^2} & \text{if } \tau_f < t \leq \tau_a + \tau_f \\ 0 & \text{if } t > \tau_a + \tau_f \end{cases}$
decreasing	$g(t) = \begin{cases} \frac{2(\tau_a - t)}{\tau_a^2} & \text{if } 0 < t \leq \tau_a \\ 0 & \text{otherwise} \end{cases}$	$S_C(t) = \begin{cases} 1 & \text{if } t \leq \tau_f \\ 1 - \frac{(\tau_f - t)^2}{\tau_a^2} & \text{if } \tau_f < t \leq \tau_a + \tau_f \\ 0 & \text{if } t > \tau_a + \tau_f \end{cases}$

Table 2

Empirical power, asymptotic power from the PH cure model and the standard PH model, and the sample sizes based on the empirical power ($H_0: \beta_0 = \gamma_0 = 0$)

	Sample Size	Empirical Power	Asymptotic Power		Sample Size for Empirical Power		
			Cure Model	PH Model	Cure Model	PH model	
scenario 1	$\pi_0 = 0.4$	100	0.276	0.286	0.247	96	
	β_0 0	$\pi_1 = 0.5$	200	0.490	0.505	0.438	193
		$\lambda_0 = 0.4$	400	0.780	0.796	0.723	384
	$\lambda_1 = 0.3$	500	0.863	0.877	0.814	480	
		$\tau_a = 1$	700	0.951	0.958	0.922	672
	$\tau_f = 4$	1000	0.991	0.993	0.981	963	
		$\pi_0 = 0.1$	100	0.425	0.398	0.247	109
	$\pi_1 = 0.2$	200	0.707	0.673	0.438	217	
		$\lambda_0 = 0.4$	300	0.868	0.838	0.599	327
	$\lambda_1 = 0.3$	400	0.942	0.926	0.723	431	
scenario 2	$\tau_a = 1$	500	0.977	0.968	0.814	541	
		600	0.992	0.986	0.878	660	
	$\pi_0 = 0.1$	100	0.182	0.181	0.178	101	
		200	0.311	0.317	0.311	196	
	$\gamma_0 = 0$	$\gamma_0 = 0.4$	500	0.636	0.651	0.640	484
		$\lambda_1 = 0.3$	1000	0.905	0.913	0.906	971
	$\tau_a = 1$	1500	0.980	0.982	0.996	1462	
		$\tau_f = 3$	2000	0.996	0.997	1	1368
	$\gamma_0 = 0.5$	100	0.089	0.080	0.225	124	
		$\pi_1 = 0.5$	200	0.127	0.119	0.398	220
$\lambda_0 = 0.4$	500	0.248	0.234	0.768	537		
	$\lambda_1 = 0.3$	1000	0.442	0.416	0.968	1079	
$\tau_a = 1$	2000	0.727	0.695	1	2154		
	$\tau_f = 3$	4000	0.951	0.937	1	4282	

	Sample Size	Empirical Power	Asymptotic Power		Sample Size for Empirical Power	
			Cure Model	PH Model	Cure Model	PH model
scenario 3	$\pi_0 = 0.4$	100	0.134	0.05	108	NA
	$\beta_0 = 0$	200	0.225	0.05	213	NA
	$\pi_1 = 0.5$	500	0.477	0.05	538	NA
	γ_0	1000	0.768	0.05	1065	NA
	$\lambda_1 = 0.5$	1500	0.909	0.05	1603	NA
	$\tau_a = 1$	2000	0.968	0.05	2142	NA
	$\tau_f = 4$	100	0.144	0.05	102	NA
	$\pi_0 = 0.5$	200	0.245	0.05	204	NA
	$\pi_1 = 0.6$	500	0.519	0.05	504	NA
	$\lambda_0 = 0.5$	1000	0.810	0.05	1007	NA
	$\lambda_1 = 0.5$	1500	0.935	0.05	1518	NA
	$\tau_a = 1$	2000	0.980	0.05	2020	NA
	$\tau_f = 4$					