Phase II clinical trials with time-to-event endpoints: Optimal two-stage designs with one-sample log-rank test

Minjung Kwak[a] and **Sin-Ho Jung**[b,c,*]

[a]Department of Statistics, Yeungnam University, Kyeongbuk, 712-749, ROK

[b]Samsung Center Research Institute, Samsung Medical Center, Seoul, ROK

[c]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27710, USA

Summary

Phase II clinical trials are often conducted to determine whether a new treatment is sufficiently promising to warrant a major controlled clinical evaluation against a standard therapy. We consider single-arm phase II clinical trials with right censored survival time responses where the ordinary one-sample logrank test is commonly used for testing the treatment efficacy. For planning such clinical trials this paper presents two-stage designs that are optimal in the sense that the expected sample size is minimized if the new regimen has low efficacy subject to constraints of the type I and type II errors. Two-stage designs which minimize the maximal sample size are also determined. Optimal and minimax designs for a range of design parameters are tabulated along with examples.

Keywords

logrank test; minimax design; optimal design; single-arm trial; two-stage design; time to event

1 Introduction

The most common primary endpoint in phase II cancer clinical trials is tumor response which is a binary variable that indicates the size of an index tumor has changed substantially following treatment. Sometimes, this endpoint is not appropriate. Examples include studies involving blood cancers and surgical studies with adjuvant chemotherapies where the tumor is completely resected, so that tumor response is not a meaningful endpoint. Also, tumor response is not a good endpoint for phase II trials on experimental cytotoxic therapies as cytotoxic therapies are meant to prevent the growth of tumor rather than shrinking it. In these cases, a preferred clinical outcome may be the time to disease progression or death. Because of the loss to follow-up or termination of the study, event times are subject to right censoring. Following the standard terminology, we will use time-to-event, failure time, and survival time as synonymous in this paper.

*Correspondence to: Sin-Ho Jung, sinho.jung@duke.edu.

In this paper, we present optimal and minimax two-stage designs for single-arm phase II trials with a survival outcome as the primary endpoint. Several articles have already appeared in the literature which present single-arm multi-stage procedures specifically developed for phase II trials. Gehan [1] and Fleming [2] developed two-stage designs for estimating the response rate, and Simon's [3] optimal design is one of the most commonly used designs developed for phase II clinical trials. Fleming used normal approximation in multiple testing procedures, and Simon determined optimal designs by enumeration using exact binomial probabilities, but both of them considered the case when the primary response is binomial. Although tumor response is a frequently used endpoint in phase II clinical trials, we have increasingly observed time-to-event measure primary endpoints in phase II cancer trials [4]. Thus it is valuable to investigate optimal two-stage designs for time-to-event endpoints in phase II clinical trials.

Case and Morgan [5] proposed optimal two-stage designs for a two-sample log-rank type test allowing for an accrual suspension before stage 1 testing. Their method finds the optimal accrual and follow-up periods for each stage given accrual rate, $(\alpha, 1 - \beta)$, rejection and acceptance values at stage 1 and the number of events after each stage. At stage 2, they use the critical value $z_{1-\alpha}$ of single-stage designs. There are many differences between Case and Morgan [5] method and ours, but one major difference that the number of events is an output variable in our method, while it is an input variable in theirs. By specifying the number of events for each stage, Case and Morgan [5] method just finds an optimal allocation of the accrual period and follow-up period for each stage, while ours extends the optimality condition to the balance between stages 1 and 2. Our method can be extended to allow for an accrual suspension for additional follow-up after stage 1 as in Case and Morgan [5], but we do not consider this option since it is rarely carried out in real clinical trials.

Section 2 reviews a single-stage procedure for single-arm phase II clinical trials when the primary outcome is a survival endpoint. In Section 3 we propose two-stage designs and analysis methods for a continuous survival time outcome in single arm phase II clinical trials. We describe how to obtain optimal two-stage designs in such trials and present the required sample sizes and critical values at each of the two stages in Section 4. Simulations to illustrate empirical size and power for each design are also presented. We conclude this article with some discussion on the implications of these optimal two-stage designs in Section 5.

2 Single-Stage Procedure

The one-sample log-rank test has been investigated by many statisticians including Berry [6] and Finkelstein et al.[7].

2.1 One-Sample Log-Rank Test

Suppose that n patients are treated by the experimental therapy of a single-arm phase II clinical trial. For patient $i(= 1, \ldots, n)$, let T_i denote the survival time with a survival function $S(t) = P(T_i \geq t)$. Because of censoring, we observe $\{(X_i, \delta_i), i = 1, \ldots, n\}$ instead of the survival times, where X_i denotes the minimum of survival time T_i and censoring time C_i and

$\delta_i = I(T_i \quad C_i)$ denotes the event indicator taking 1 if an event is observed and 0 otherwise. We assume that the survival and censoring times are independent.

Let $\Lambda_0(t)$ denote the cumulative hazard function of a historical control that is chosen for a new single-arm phase II trial. If the historical control data come from a previous study, $\Lambda_0(t)$ may be the Nelson-Aalen estimate [8],[9] from the data. Let $\Lambda(t)$ denote the cumulative hazard function of an experimental therapy that will be observed from a new phase II trial. We want to test $H_0 : \Lambda(t) = \Lambda_0(t)$ against $H_1 : \Lambda(t) < \Lambda_0(t)$ for all $t > 0$.

Let $N_i(t) = \delta_i I(X_i \quad t)$ and $Y_i(t) = I(X_i \quad t)$ be the event and at-risk processes, respectively. Under H_0 for large n,

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

is approximately normal with mean 0 and its variance can be consistently estimated by

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

Hence, we reject H_0 with one-sided type I error rate α if $Z = W/\hat{\sigma} < -z_{1-\alpha}$ [7]. Here $Z_{1-\alpha}$ denotes the $100(1 - \alpha)$ percentile of the standard normal distribution.

Note that the standardized test statistic, $W/\hat{\sigma}$ is expressed as

$$Z = \frac{O - E}{\sqrt{E}},$$

where $O = \sum_{i=1}^{n} \int_0^\infty dN_i(t) = \sum_{i=1}^{n} \delta_i$ denotes the observed number of events and $E = \sum_{i=1}^{n} \int_0^\infty Y_i(t)d\Lambda_0(t)$. Note that, under H_0, $n^{-1} \sum_{i=1}^{n} Y_i(t) \rightarrow S_0(t)G(t)$ uniformly, so that $E = \sum_{i=1}^{n} \int_0^\infty Y_i(t)d\Lambda_0(t) = \sum_{i=1}^{n} \Lambda_0(X_i)$ is asymptotically identical to $-n \int_0^\infty G(t)dS_0(t)$ which is the expected number of events under H_0, where $S_0(t) = \exp\{-\Lambda_0(t)\}$ and $G(t) = P(C_i \quad t)$.

2.2 Sample Size Calculation

We calculate the required sample size n for a specified power under a specific alternative hypothesis $H_1 : \Lambda(t) = \Lambda_1(t)(< \Lambda_0(t))$. Under H_1, $n^{-1} \sum_{i=1}^{n} Y_i(t)$ uniformly converges to $G(t)S_1(t)$, so that $\hat{\sigma}^2$ converges to

$$\sigma_0^2 = \int_0^\infty G(t)S_1(t)d\Lambda_0(t), \quad (1)$$

where $S_1(t) = \exp\{-\Lambda_1(t)\}$.

On the other hand, we have

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\infty \{dN_i(t) - Y_i(t)d\Lambda_1(t)\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\infty Y_i(t)d\{\Lambda_1(t) - \Lambda_0(t)\},$$

so that, under H_1, W is approximately normal with mean

$$\sqrt{n}\omega \equiv \sqrt{n} \int_0^\infty G(t)S_1(t)d\{\Lambda_1(t) - \Lambda_0(t)\}.$$

Assuming that $\Lambda_1(t)$ and $\Lambda_0(t)$ are close, we calculate its variance by

$$\sigma_1^2 = \int_0^\infty G(t)\bar{S}(t)d\bar{\Lambda}(t), \quad (2)$$

where $\bar{\Lambda}(t) = \{\Lambda_0(t) + \Lambda_1(t)\}/2$ and $\bar{S}(t) = \exp\{-\bar{\Lambda}(t)\}$.

Hence, we have the power function

$$1 - \beta = P(W/\hat{\sigma} < -z_{1-\alpha}|H_1) = P(\frac{W - \sqrt{n}\omega}{\sigma_1} \times \frac{\sigma_1}{\sigma_0} < -\frac{\sqrt{n}\omega}{\sigma_0} - z_{1-\alpha}|H_1).$$

By solving this equation and replacing $\omega = \sigma_1^2 - \sigma_0^2$, we obtain the required sample size

$$n = \frac{(\sigma_0 z_{1-\alpha} + \sigma_1 z_{1-\beta})^2}{\omega^2}. \quad (3)$$

Although the one-sample log-rank test is nonparametric, its sample size calculation requires specification of survival and censoring distributions. We now derive sample size formulas under some practical design settings.

Under Uniform Censoring and Exponential Survival Models—For a practical sample size calculation, we assume that the survival distribution of the experimental therapy is exponential with hazard rate λ_0 under H_0 and λ_1 under H_1. The survival functions are given as $S_0(t) = \exp(-\lambda_0 t)$ under H_0 and $S_1(t) = \exp(-\lambda_1 t)$ under H_1. Under exponential models for survival distributions, we have a proportional hazards model, $= \lambda_0/\lambda_1$.

Assuming that patients are accrued at a constant rate during period a and followed for an additional period of b, the censoring distribution is given as $U(b, a + b)$ with a survivor function $G(t) = P(C \quad t) = 1$ if $t \quad b$; $= (a + b)/a - t/a$ if $b \quad t \quad a + b$; $= 0$ otherwise.

Under a uniform censoring and an exponential survival models, it is easy to show that

$$\sigma_0^2 = \Delta \left\{ 1 - \frac{e^{-\lambda_1 b}}{a\lambda_1}(1 - e^{-\lambda_1 a}) \right\}, \quad \sigma_1^2 = 1 - \frac{e^{-\bar{\lambda}b}}{a\bar{\lambda}}(1 - e^{-\bar{\lambda}a}),$$

and

$$\omega = (1 - \Delta) \left\{ 1 - \frac{e^{-\lambda_1 b}}{a\lambda_1}(1 - e^{-\lambda_1 a}) \right\},$$

where $\Delta = \lambda_0/\lambda_1$.

When Accrual Rate is Given—Note that $\omega = \omega(a)$, $\sigma_0^2 = \sigma_0^2(a)$, and $\sigma_1^2 = \sigma_1^2(a)$ are functions of an accrual period a. So, sample size formula (3) requires specification of an accrual period as an input variable. However, at the design stage of a clinical trial, we usually know the expected accrual for the trial, rather than the accrual period. Let r denote the expected accrual rate. For large n, we have $n \approx a \times r$, so that, given r, a should be an output variable of a sample size calculation. In (3), by replacing n with $a \times r$, we obtain an equation on a,

$$a \times r = \frac{\{\sigma_0(a)z_{1-\alpha} + \sigma_1(a)z_{1-\beta}\}^2}{\omega(a)^2}. \quad (4)$$

We solve this equation using a numerical method, such as the bisection method. Let a^* denote the solution to one of these equations. Then, the required sample size is given as $n = a^* \times r$.

Example 1 Suppose that the progression-free survival (PFS) for a standard therapy (historical control) has the exponential distribution with a median of $\theta_0 = 1$ year. We will be interested in the experimental therapy if its median PFS is $\theta_1 = 1.5$ years or longer ($\lambda_0 = 0.693$ and $\lambda_1 = 0.462$). We want to design a study to detect this improvement in PFS by the experimental therapy with $1 - \beta = 90\%$ power using the one-sample log-rank test with one-sided $\alpha = 10\%$ ($z_{1-\alpha} = z_{1-\beta} = 1.282$). Further, suppose that this trial is expected to accrue about 2 to 3 patients per month (or $r = 30$ per year) based on the recent accrual rate of patients at the study institution. By solving equation (4) with respect to a, we obtain an accrual period of $a^* = 1.96$ years (or about 24 months), and a required sample size of $n = a^* \times r = 30 \times 1.96 = 59$ patients. For $a = 1.96$ and the given (b, λ_0, λ_1) values, we have $\omega = -0.293$, $\sigma_0^2 = 0.878$ and $\sigma_1^2 = 0.664$.

From 10,000 simulations under the design setting with $n = 59$, we observed an empirical type I error of 9.3% and power of 89%.

3 Two-Stage Procedure

We consider an interim analysis only for futility testing as in most traditional single-arm phase II trials, but the cases with efficacy only or with both futility and efficacy can be easily derived.

3.1 Two-Stage One-Sample Log-Rank Test

Suppose that n_1 is the number of patients treated by the experimental therapy during the first stage and n_2 is the additional number of patients treated by the experimental therapy during the second stage. Let $n = n_1 + n_2$ denote the maximal sample size. We conduct an interim analysis at time τ which may be determined in terms of number of events or by calendar time. In order to avoid treating too many patients when the experimental therapy is shown to be inefficacious, we may assume that τ is smaller than the planned accrual period a.

For subject $i = 1, \ldots, n$, let T_i denote the survival time with survival distribution $S_h(t)$ and cumulative hazard function $\Lambda_h(t)$ under H_h, $h = 0, 1$, and $e_i \in [0, a]$ denote the entering time in the trial. Let C_i denote the censoring time of patient i at the end of stage 2 with survivor function $G(t) = P(C_i \quad t)$ that is defined by the accrual trend and additional follow-up period. The censoring time at the interim analysis is denoted as $C_{1i} = \max\{\min(\tau - e_i, C_i), 0\}$. If patient i enters the study during stage 1, i.e. $e_i < \tau$, then the censoring variable at the interim analysis has a survivor function $G_1(t) = P\{\min(\tau - e_i, C_i) \quad t\}$.

The observed survival data are expressed as (X_{1i}, δ_{1i}) at the interim analysis and (X_i, δ_i) at the final analysis, where $X_{1i} = \min(T_i, C_{1i})$, $\delta_{1i} = I(T_i \quad C_{1i})$, $X_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \quad C_i)$. We define at-risk processes $Y_{1i}(t) = I(X_{1i} \quad t)$ and $Y_i(t) = I(X_i \quad t)$, and event processes $N_{1i}(t) = \delta_{1i}I(X_{1i} \quad t)$ and $N_i(t) = \delta_i I(X_i \quad t)$.

The test statistics at the interim and final analyses are expressed as

$$W_1 = n_1^{-1/2} \sum_{i=1}^{n_1} \int_0^\infty \{dN_{1i}(t) - Y_{1i}(t)d\Lambda_0(t)\}$$

and

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

respectively. For large n_1 and n, the distribution of (W_1, W) under H_0 is approximately bivariate normal with means 0, variances and covariance that can be approximated by

$$\hat{\sigma}_1^2 = n_1^{-1} \sum_{i=1}^{n_1} \int_0^\infty Y_{1i}(t)d\Lambda_0(t), \quad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \int_0^\infty Y_i(t)d\Lambda_0(t),$$

and $\mathrm{cov}(W, W_1) = \hat{\sigma}_1^2$, respectively, see Tsiatis [10]. So, the correlation coefficient between W_1 and W is given as $\hat{\rho} = \hat{\sigma}_1/\hat{\sigma}$.

Note that n_1 denotes the number of patients who have entered to the study before the interim analysis time τ, i.e. $n_1 = \sum_{i=1}^{n} I(e_i \leq \tau)$. At the interim analysis, the patients who have not entered to the study yet, i.e. $e_i > \tau$, have their survival times censored at time 0, i.e. $X_{1i} = 0$

and $\delta_{1i} = 0$, so that they make no contributions to W_1 and $\hat{\sigma}_1^2$. A two-stage trial using the one-sample log-rank test is conducted as follows.

At the design stage, we specify $\Lambda_0(t)$ and α, together with an interim analysis time and an early stopping value c_1. After the first stage of the trial, we reject the experimental therapy and stop the trial if $W_1/\hat{\sigma}_1$ c_1. Otherwise, we continue to accrue and treat patients for the second stage. After the second stage, we accept the experimental therapy if $W/\hat{\sigma} < c$. Here, critical value c satisfies

$$\alpha = P\left(\frac{W_1}{\hat{\sigma}_1} \leq c_1, \frac{W}{\hat{\sigma}} \leq c \mid H_0\right). \quad (8)$$

If (X, Y) is a bivariate normal random vector with means, μ_x and μ_y, variances σ_x^2 and σ_y^2, and correlation coefficient, ρ, then it is well known that the conditional distribution of X given $Y = y$ is normal with mean $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$ and variance $\sigma_x^2(1 - \rho^2)$. This result simplifies the calculation of type I error probability and power in our paper.

Let $Z_1 = W_1/\hat{\sigma}_1$ and $Z = W/\hat{\sigma}$. Noting that, conditioning on $Z = z$, Z_1 is approximately normal with mean $\hat{\rho}z$ and variance $1 - \hat{\rho}^2$, we approximate equation (8) by

$$\alpha = \int_{-\infty}^{c} \phi(z)\Phi\left(\frac{c_1 - \hat{\rho}z}{\sqrt{1 - \hat{\rho}^2}}\right)dz, \quad (9)$$

where $\hat{\rho} = \hat{\sigma}_1/\hat{\sigma}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and cumulative distribution functions of $N(0, 1)$ distribution, respectively.

3.2 P-value Calculation for Two-Stage Designs

Our testing rule just gives us the decision to accept the experimental therapy or not. However, often we may want to know how significant the evidence will be against $H_0 : \Lambda(t) = \Lambda_0(t)$. We obtain this information by calculating a p-value. An unbiased p-value should reflect the two-stage procedure of our testing. If a study is stopped early after observing $z_1 = w_1/\hat{\sigma}_1(> c_1)$, then we obtain a p-value by

$$\text{p-value} = \Phi(z_1).$$

If a study is continued to the second stage and observes $z = w/\hat{\sigma}$, then by modifying (9) we obtain a p-value by

$$\text{p-value} = P(Z_1 \leq c_1, Z \leq z) = \int_{-\infty}^{z} \phi(u)\Phi\left(\frac{c_1 - \hat{\rho}u}{\sqrt{1 - \hat{\rho}^2}}\right)du,$$

where (Z_1, Z) is a bivariate normal random vector with marginal means of 0 and variances of 1, and a correlation coefficient of $\hat{\rho} = \hat{\sigma}_1/\hat{\sigma}$. Note that the testing decision based on this p-value coincides with that based on the critical value as described in Section 3.1.

3.3 Sample Size Calculation for Two-Stage Design

At first, we derive a power function given τ and c_1 together with accrual period a, follow up period b, $\Lambda_h(t)$ for $h = 0, 1$ and (α, $1 - \beta$). The interim analysis time τ may be determined in terms of calendar time or number of events observed, but at the design stage we should specify it as a calendar time. If we want to specify it in terms of number of events at the design stage, we can convert it to a calendar time based on the expected accrual rate and the specified survival distribution.

The power function is given as

$$1 - \beta = P\left(\frac{W_1}{\hat{\sigma}_1} \leq c_1, \frac{W}{\hat{\sigma}} \leq c \middle| H_1\right).$$

Before deriving a power function, we have to calculate c for a specified type I error rate α, i.e.

$$\alpha = P\left(\frac{W_1}{\hat{\sigma}_1} \leq c_1, \frac{W}{\hat{\sigma}} \leq c \middle| H_0\right).$$

So, for a power calculation, we need to derive the limits of $\hat{\sigma}_1^2$ and $\hat{\sigma}^2$ under H_0 and H_1, and $\omega_1 = E_{H_1}(W_1)$, $\omega = E_{H_1}(W)$, $\text{var}_{H_1}(W_1)$ and $\text{var}_{H_1}(W)$ under H_1.

Under H_0, we have $E_{H_0}(W_1) = E_{H_0}(W) = 0$, and $\hat{\sigma}_1^2$ and $\hat{\sigma}^2$ converge to

$$\upsilon_1 = -\int_0^\infty G_1(t) dS_0(t)$$

and

$$\upsilon = -\int_0^\infty G(t) dS_0(t)$$

respectively. Note that $\text{var}_{H_0}(W_1) = \upsilon_1$ and $\text{var}_{H_0}(W) = \upsilon$ under H_0. By independent increment of the one-sample log-rank statistic, $\text{corr}(W_1, W)$ is given as $\rho_0 = \sqrt{\upsilon_1/\upsilon}$.

Under H_1, we have $E_{H_1}(W_1) = \sqrt{n_1}\omega_1$ and $E_{H_1}(W) = \sqrt{n}\omega$, where

$$\omega_1 = \int_0^\infty G_1(t) S_1(t) d\{\Lambda_1(t) - \Lambda_0(t)\}$$

and

$$\omega = \int_0^\infty G(t) S_1(t) d\{\Lambda_1(t) - \Lambda_0(t)\}.$$

Furthermore, $\hat{\sigma}_1^2$ and $\hat{\sigma}^2$ converge to

$$\sigma_{01}^2 = -\int_0^\infty G_1(t)S_1(t)d\Lambda_0(t)$$

and

$$\sigma_0^2 = -\int_0^\infty G(t)S_1(t)d\Lambda_0(t)$$

respectively. The variances of W_1 and W are approximated by

$$\sigma_{11}^2 = -\int_0^\infty G_1(t)\bar{S}(t)d\bar{\Lambda}(t)$$

and

$$\sigma_1^2 = -\int_0^\infty G(t)\bar{S}(t)d\bar{\Lambda}(t)$$

respectively under H_1. By independent increment of the one-sample log-rank statistic, corr(W_1, W) is given as $\rho_1 = \sigma_{11}/\sigma_1$.

If the interim analysis time τ and the stopping value c_1 are reasonably chosen, the power of a two-stage design is not much lower than that of the corresponding single-stage design. So, when searching for the required accrual period (or sample size) of a two-stage design, we may start from the accrual period for the single-stage design. Assuming an accrual pattern with a uniform rate, the design procedure of two-stage designs can be summarized as follows.

- Given (α, $1 - \beta$, r, b, $\Lambda_0(t)$, $\Lambda_1(t)$), calculate the sample size n and accrual period a^* required for a single-stage design.

- Determine an interim analysis time τ during the accrual period a^* of the chosen single-stage design, i.e. $\tau < a^*$, and the stopping value c_1 at the interim analysis.

- The accrual period required for a two-stage design is obtained around a^* as follows: At $a = a^*$ (note that $n_1 \approx r\tau$ and $n \approx ra^*$),

  a.   Obtain c by solving equation

  $$\alpha = \int_{-\infty}^c \phi(z)\Phi\left(\frac{c_1 - \rho_0 z}{\sqrt{1 - \rho_0^2}}\right)dz.$$

  b.   Given (n_1, n, c_1, c, α), calculate

$$\text{power} = \int_{-\infty}^{\bar{c}} \phi(z)\Phi(\frac{\bar{c}_1 - \rho_1 z}{\sqrt{1 - \rho_1^2}})dz$$

where

$$\bar{c}_1 = \frac{\sigma_{01}}{\sigma_{11}}(c_1 - \frac{\omega_1 \sqrt{n_1}}{\sigma_{01}}) \text{ and } \bar{c} = \frac{\sigma_0}{\sigma_1}(c - \frac{\omega \sqrt{n}}{\sigma_0}).$$

- If the power is smaller than $1 - \beta$, increase a slightly, and repeat above procedure until the power is close enough to $1 - \beta$.

For a candidate design (n, τ, c_1, c), the probability of early termination (PET) under H_0, $P(W_1/\hat{\sigma}_1 > c_1|H_0)$, is approximated by PET $= \bar{\Phi}(c_1)$, where $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. Let $n_2 = n - n_1$ denote the number of patients who are accrued after the interim analysis. Note that, given the maximal accrual period a, we have $n_2 = \{(a - \tau) \vee 0\}r$. So, the expected sample size (EN) under H_0 is given as

$$\text{EN} = n_1 \times \text{PET} + n \times (1 - \text{PET}) = n - n_2 \times \text{PET} = r[a - \{(a - \tau) \vee 0\} \times \text{PET}].$$

Under Uniform Accrual and Exponential Survival Models—Suppose that the survival distribution is exponential with hazard rate λ_0 under H_0 and λ_1 under H_1. If patients are accrued at a constant rate during period a and followed for an additional period of b, and the interim analysis is taken place before completion of patient accrual (i.e. $\tau < a$), then the censoring distribution at the interim analysis is $U(0, \tau)$ and that at final analysis is $U(b, a + b)$, for which the survivor functions are given as

$$G_1(t) = \begin{cases} 1 & \text{if } t \leq 0 \\ 1 - t/\tau & \text{if } 0 < t \leq \tau \\ 0 & \text{if } t > \tau \end{cases}$$

and

$$G(t) = \begin{cases} 1 & \text{if } t \leq b \\ -t/a + (a+b)/a & \text{if } b < t \leq a+b \\ 0 & \text{if } t > a+b \end{cases}$$

respectively. Note that we assume administrative censoring only. If loss to follow up is expected, then we may incorporate it in the calculation if its distribution can be modeled or we may increase the final sample size by the expected proportion of loss to follow up.

Under these assumptions, we can show that

$$v_1 = 1 - (1 - e^{-\lambda_0 \tau})/(\tau \lambda_0),$$
$$v = 1 - (1 - e^{-\lambda_0 a})e^{-\lambda_0 b}/(a\lambda_0),$$
$$\sigma_{11}^2 = 1 - (1 - e^{-\bar{\lambda}\tau})/(\tau\bar{\lambda}),$$
$$\sigma_1^2 = 1 - (1 - e^{-\bar{\lambda}a})e^{-\bar{\lambda}b}/(a\bar{\lambda}),$$
$$\sigma_{01}^2 = \Delta\{1 - (1 - e^{-\lambda_1 \tau})/(\tau\lambda_1)\},$$
$$\sigma_0^2 = \Delta\{1 - (1 - e^{-\lambda_1 a})e^{-\bar{\lambda}b}/(a\bar{\lambda})\},$$
$$\omega_1 = (1 - \Delta)\left\{1 - \frac{1}{\tau\lambda_1}(1 - e^{-\lambda_1 \tau})\right\},$$
$$\omega = (1 - \Delta)\left\{1 - \frac{e^{-\lambda_1 b}}{a\lambda_1}(1 - e^{-\lambda_1 a})\right\}.$$

Example 2: We consider the design parameter values of Example 1, $(a, \lambda_0, \lambda_1, r, b) = (0.1,$ 0.693, 0.462, 30, 1). Then, we can show that the two-stage design defined by $(\tau, c_1, c, n) =$ (1.27, 0.610, −1.275, 60) has 90% power like the single-stage of Example 1. This two-stage has PET = 0.27 and EN = 54.0. By accruing only one more patient in maximum, the two-stage design saves 6 patients in expectation compared to the size of the single-stage design. From 10,000 simulations on the design setting, the two-stage design is shown to have an empirical type I error of 9.3% and a power of 88%.

4 Optimal Two-Stage Designs

In this section, we propose some optimal two-stage designs for given $(a, 1-\beta, r, b, \Lambda_0(t),$ $\Lambda_1(t))$. Given (r, b), a candidate two-stage design specified by (n, τ, c_1, c) has a type I error rate of a for $\lambda(t) = \lambda_0(t)$ and a power no smaller than $1 - \beta$ for $\lambda(t) = \lambda_1(t)$. We consider two optimality criteria, one to minimize the expected sample size under H_0 and the other to minimize the expected study period under H_0. By specifying an accrual rate r, we assume a uniform accrual trend, but we can extend the following results to any accrual pattern.

4.1 Optimal Designs

Among the candidate two-stage designs, we define the *optimal* design as the one minimizing the expected sample size, or the expected accrual period (EA), given as EA = $a - \{(a - \tau)\vee$ $0\} \times$ PET, under H_0. We also define the *minimax* design as the one minimizing the maximal sample size n, or equivalently the maximal accrual period a. For a given n, there may be multiple two-stage designs satisfying the $(a, 1 - \beta)$ condition. The minimax design has the smallest EA among them.

Through our experience from numerical studies, we have found that the maximal accrual period of the minimax design is not very different from the accrual period of the corresponding single-stage design, a reasonable interim analysis should be conducted when the survival data are somewhat matured, and the rejection value of stage 1, c_1, is not very different from 0. Given $(a, 1 - \beta, r, b, \Lambda_0(t), \Lambda_1(t))$, let a^* denote the accrual period of the single-stage design. An efficient computational procedure to identify the minimax and optimal designs can be summarized as follows.

Search for Optimal Designs

A. Specify $(\alpha, 1 - \beta, r, b, \Lambda_0(t), \Lambda_1(t))$.

B. Find the accrual period a^* for the single-stage design requiring the smallest sample size.

C. Changing the values of a, τ and c_1, calculate c and power. As reasonable range and increment for these design parameters, we chose $a \in [0.80a^*, 1.5a^*]$ and $\tau \in [0.20a^*, 1.2a^*]$ with an increment of $1/r$ for each of them, and $c_1 = [-0.2, 1]$ with an increment of 0.005.

  i. If the power is smaller than $1 - \beta$, then go to the next combination of (a, τ, c_1).

  ii. If the power is larger than or equal to $1 - \beta$, then calculate PET and EA.

   - If EA is smaller than the minimum of the expected accrual periods among the all candidate designs we have gone through, then save the current $(a, \tau, c_1, c, \text{EA}, \text{PET}, \text{power})$.

   - Otherwise, go to the next combination of (a, τ, c_1).

D. Among the designs saved during procedure (C), the one with the smallest a is the minimax design and the one with the smallest EA is the optimal design.

Table 1 shows single-stage design and minimax and optimal two-stage designs for various settings defined by the design parameter values. We consider an annual accrual rate of $r = 30$ or 60 patients, a follow-up period of $b = 1$ year and an exponential survival distribution with an annual hazard rate $\lambda_0 = 0.7$ under H_0, corresponding to a median survival of about one year. Under H_1, we assume a hazard rate of $\Delta = \lambda_0/\lambda_1 = 1.4, 1.5, 1.6$ or 1.7. We also consider $(\alpha, 1 - \beta) = (0.05, 0.9), (0.1, 0.9)$ or $(0.05, 0.85)$. Under each design setting, we report the sample size for the single-stage design n^*. For the minimax and the optimal two-stage designs, we report expected sample size n_1, analysis time τ and rejection value c_1 at the first stage, maximal sample size n, maximal accrual period a, second stage rejection value c, expected sample size under H_0 EN, and probability of early termination under H_0 PET. For example, under $(r, \lambda_0, \Delta, \alpha, 1 - \beta) = (30, 0.7, 1.4, 0.05, 0.9)$, the optimal design will conduct an interim analysis at $\tau = 1.9$ years when about $n_1 = 57$ patients are expected to be entered, to reject the experimental treatment and terminate the study early if $W_1/\hat{\sigma}_1$ is larger than $c_1 = -0.130$. Otherwise, the trial will continue to accrue a total of 107 patients. The expected sample size is 79.2 and the probability of early termination is 0.55 under H_0. On the other hand, the required numbers at the first and the second stages of the minimax design are 62 patients and 98 patients, respectively. The sample size increases in $1 - \beta$ and r, and decreases in α and Δ. Although not covered by our numerical study, it is easy to show that sample size decreases in b and increases in λ_0. We observe that the maximal sample size n of the minimax two-stage design is identical to the sample size n^* of the single-stage design under most of the design settings. In order to minimize the maximal sample size n, the minimax designs are similar to the corresponding single-stage designs in the sense that

n_1 is close to n and PET is low. Compared to the minimax designs, however, the optimal designs conduct interim analyses earlier with a larger PET to minimize EN.

When $= 1.5$ and the accrual rate is 30 patients per year, the sample size savings are as large as 33% and 38% for the minimax and the optimal designs respectively compared to the corresponding single-stage design if we stop the trial at the interim analysis. However, we only need 1% and 10% more patients for the minimax and the optimal designs when we continue and finish the trial at the second stage. With accrual rate 60 patients per year, the sample size savings are as large as 32% and 38% for the minimax and the optimal designs respectively compared to the corresponding single-stage designs if we stop the trial at the interim analysis, whereas we need as little as 1% and 7% more patients of the single-stage design for the minimax and the optimal designs when we continue and finish the trial at the second stage.

4.2 Implementation of Optimal Two-Stage Designs

The minimax and optimal two-stage designs are based on the accrual rate expected at the design stage. Once the study is open, however, the realized accrual pattern may be different from that specified at the design. In this case, the analysis times and the variances of W_1 and W and their correlation coefficient may be different from those specified at the design stage. Noting that the power of the log-rank test depends on the number of events, rather than the number of patients or calendar time, we propose to choose the interim and the final analysis times when we have the expected numbers of events by the chosen two-stage design. Let D_1 and D denote the expected number of events under H_1 at the interim and final analysis times based on the design parameters, respectively, i.e.

$$D_1 = n_1 \{ 1 - (1 - e^{-\lambda_1 \tau})/\lambda_1 \tau \}$$

and

$$D = n \{ 1 - e^{-\lambda_1 b}(1 - e^{-\lambda_1 a})/\lambda_1 a \},$$

where $n_1 = r \min(\tau, a)$. If $\lambda_0 \approx \lambda_1$, the asymptotic distribution of $(W_1/\hat{\sigma}_1, W/\hat{\sigma})$ is a bivariate normal with means 0, variances 1 and correlation coefficient $\hat{\rho} = \sqrt{\hat{\sigma}_1/\hat{\sigma}} = \sqrt{nD_1/(n_1 D)}$. Hence, as far as the real accrual trend is close to the expected one at the design, so that n_1 patients are entered at τ and a total of n patients are entered during the expected accrual period a, the two-stage design specified by $(\alpha, \lambda_0, c_1, \tau, r, b)$ will be identical to that specified by $(\alpha, \lambda_0, c_1, D_1, D)$. Recall that c_1 is fixed by the design, but c is recalculated for a type I error rate of α reflecting the realized accrual trend. So, τ, c, a and b will be used as just references during the trial.

In summary we propose to conduct a study with a two-stage design as follows.

- Specify $(\lambda_0, \lambda_1, \alpha, 1 - \beta)$.

- Choose a two-stage design $(n_1, n, \tau, c_1, c, a, b)$.

- Calculate D_1 and D based on the selected two-stage design.

- Accrue n patients and follow them until D events are observed unless the study is stopped early by the stage 1 analysis.

  Stage 1: When D_1 events are observed, conduct the stage 1 analysis by calculating W_1 and $\hat{\sigma}_1^2$, and stop the trial rejecting the study therapy if $W_1/\hat{\sigma}_1 > c_1$. Otherwise, proceed to stage 2.

  Stage 2: Conduct the final analysis when D events are observed by calculating W, $\hat{\sigma}^2$ and critical value c' based on $\hat{\rho} = \hat{\sigma}_1/\hat{\sigma}$. We reject the study therapy if $W/\hat{\sigma} > c'$.

4.3 Simulation Study

The one-sample log-rank test is based on large sample approximation. To demonstrate the small sample performance of our two-stage design, we conducted simulations under the minimax and the optimal two-stage designs listed in Table 1. Given (λ_h, r, b), we generated 10,000 simulation samples of size n, and conduct two-stage analysis defined by (τ, c_1). By calculating the p-value for testing, we did not have to calculate c for each simulation sample. We calculated the empirical size (for $h = 0$) and power (for $h = 1$) by the proportion of simulation samples rejecting H_0 (i.e. those with p-value$< \alpha$) among 10,000 samples. In our simulations, the stage 1 sample size may not be exactly n_1 for each simulation sample since the interim test is conducted at τ regardless of the sample size. And, n patients are accrued during an accrual period of a and the final test is conducted at $a + b$.

Table 2 reports the empirical size and power of the single-stage and minimax and optimal two-stage designs that are listed in Table 1. The one-sample log-rank test seems to be slightly conservative under H_0 and the designs are slightly underpowered under H_1. However, the empirical size and power are very close to their nominal levels overall.

5 Discussion

While a phase I trial of a new anticancer drug aims to gain information about the maximum tolerated dose of a therapy by treating only three to six patients per dose level, the purpose of a phase II trial is to determine whether the drug has sufficient activity against a specified type of tumor to warrant its further development. Further development may mean combining the drug with other drugs, evaluation in patients with less advanced disease, or initiation of randomized controlled phase III studies in which a larger number of patients are treated and followed for longer period of time. Oftentimes, a phase II study of a cancer treatment is uncontrolled for obtaining an initial estimate of the degree of antitumor effect of the treatment. In this paper, we have considered single-arm phase II clinical trials where right censored failure time is the primary endpoint. We restricted our attention to two-stage designs because in a multi-institution setting more than two stages are difficult to manage in practice with no additional gain.

We have tabulated minimax and optimal phase II designs for $(\alpha, \beta) = (0.05, 0.1)$, $(0.1, 0.1)$, and $(0.05, 0.15)$ with a variety of hazard rates for alternative hypothesis, and through simulations showed that they maintain the type I error rate and power closely to their

nominal levels. In a phase II clinical trial of efficacy of a new treatment α is a less serious error from a drug discovery viewpoint, but it is serious from a cost perspective since it leads to unnecessary follow-up of a futile drug. The optimal designs achieve reductions in EN by having a smaller first stage sample size than the minimax designs with a higher probability early stopping when the study treatment is not efficacious A smaller first stage exposes relatively fewer patients to the new treatment if it turns out to be inactive. However, the minimax design may be more attractive than the optimal design when the difference in expected sample size is small and the patient accrual rate is low.

The proposed two-stage designs are not only to save the sample size, but also to save the study period. If the observed accrual of a study is much faster that expected, then we already may have n patients accrued at the planned interim analysis time τ. In this case, we do not save any patients by the chosen two-stage design, so that we may consider changing it to a single-stage design by skipping the planned interim test. In this case, however, the required follow-up period for the final analysis may be too long. As such, we may want to conduct an interim analysis during the follow-up period, so that we can stop the study early and save the valuable resources when we do not observe any promising evidence from the experimental therapy. On the other hand, if the accrual of a study is too slow, then we may also want to stop the trial early if the experimental therapy is not very efficacious compared to a historical control, rather than dragging the study to complete the target accrual. Whatever is the case, the maximal sample size n of a minimax design is very similar to that of the corresponding single-stage design, so that we do not lose much by conducting an interim analysis while saving sample size and study time.

Owzar and Jung [11] proposed to use the survival probability at a fixed time point in phase II trials. The binary endpoints are to control the type I error rate exactly by using the binomial distributions. However, by dichotomizing the survival data at a fixed time point, we lose too much of efficiency. For example, suppose that a historical control therapy has a one year PFS rate of $p_0 = 50\%$ and we will be interested in the experimental therapy if its one year PFS is $p_1 = 65\%$ or higher. For $(\alpha, 1 - \beta) = (0.05, 0.9)$, by Simon [3], two-stage design based on the binary outcome requires $(n_1, n) = (57, 93)$ with $EN = 75.0$ by the minimax design and $(n_1, n) = (42, 105)$ with $EN = 62.3$ by the optimal design. Assuming exponential PFS models, the corresponding hazard rates are $\lambda_0 = 0.693$ and $\lambda_1 = 0.438$. Assuming an annual accrual rate of 60 patients and an additional follow-up period of one year, two-stage one-sample logrank test requires $(n_1, n) = (50, 72)$ with EN = 67 for the minimax design and $(n_1, n) = (44, 76)$ with EN = 65 for the optimal design. Note that the two-stage one-sample log-rank test requires much smaller sample size than the two-stage binary test. Furthermore, if there are censored observations before the cutoff time point, then the binary test will have to exclude them in analysis while the one-sample log-rank test can include them.

We note that large sample tests and sample size formula tend to be conservative and underpowered with small sample sizes. When we used the exact variance formulas $\sigma_{11}^2 = - \int_0^\infty G_1(t) dS_1(t)$ and $\sigma_1^2 = - \int_0^\infty G(t) dS_1(t)$, simulation results seemed to be underpowered. After trying various approximations, we found that the formulas given in Section 3.3 give the best performance.

Fortran programs to search for minimax and optimal designs and simulations for a specific two-stage designs are provided in the Web Supplementary Materials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Gehan EA. The determination of the number of patients required in a follow-up trial of a new chemotherapeutic agent. Journal of Chronic Diseases. 1961; 13:346–353. [PubMed: 13704181]

2. Fleming TR. One sample multiple testing procedure for phase II clinical trials. Biometrics. 1982; 38:143–151. [PubMed: 7082756]

3. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials. 1989; 10:1–10. [PubMed: 2702835]

4. Brunstein CG, Fuchs EJ, Carter SL, Karanes C, Costa LJ, Wu J, Devine SM, Wingard JR, Aljitawi OS, Cutler CS, Jagasia MH, Ballen KK, Eapen M, O'Donnell PV. Alternative donor transplantation after reduced intensity conditioning: results of parallel phase 2 trials using partially HLA-mismatched related bone marrow or unrelated double umbilical cord blood grafts. Blood. 2011; 118:282–288. [PubMed: 21527516]

5. Case LD, Morgan TM. Duration of accrual and follow-up for two-stage clinical trials. Lifetime Data Analysis. 2001; 7:21–37. [PubMed: 11280845]

6. Berry G. The analysis of mortality by the subject-years methods. Biometrics. 1983; 39:173–184. [PubMed: 6871346]

7. Finkelstein DM, Muzikansky A, Schoenfeld DA. Comparing survival of a sample to that of a standard population. Journal of the National Cancer Institute. 2003; 95:1434–1439. [PubMed: 14519749]

8. Aalen OO. Nonparametric inference for a family of counting processes. Annals of Statistics. 1978; 6:701–726.

9. Nelson W. Hazard plotting for incomplete failure data. Journal of Quality Technology. 1969; 1:27–52.

10. Tsiatis AA. Repeated significance testing for a general class of statistics used in censored survival analysis. Journal of the American Statistical Association. 1982; 77:855–861.

11. Owzar K, Jung SH. Designing phase II trials in cancer with time-to-event endpoints (with discussion). Clinical Trials. 2008; 5:209–221. [PubMed: 18559409]

Table 1

Single-stage (n^*), minimax and optimal two-stage designs with $\lambda_0 = 0.7$ and a follow-up time of $b = 1$ year

r		n^*	Minimax Design						Optimal Design					
			$n_1(\tau)$	c_1	n	c	EN	PET	$n_1(\tau)$	c_1	n	c	EN	PET
							$(\alpha, 1-\beta) = (.05, .9)$							
30	1.4	97	62(2.07)	0.215	98	−1.643	82.8	.41	57(1.90)	−0.130	107	−1.633	79.2	.55
	1.5	73	49(1.60)	0.520	74	−1.643	65.6	.30	45(1.47)	0.010	80	−1.631	61.7	.50
	1.6	59	46(1.53)	0.500	59	−1.643	54.9	.31	35(1.17)	0.130	64	−1.630	50.9	.45
	1.7	50	34(1.13)	0.490	51	−1.640	45.1	.31	31(1.00)	0.230	54	−1.630	43.7	.41
							$(\alpha, 1-\beta) = (.1, .9)$							
	1.4	78	58(1.93)	0.610	78	−1.280	72.3	.27	49(1.60)	0.185	84	−1.265	68.4	.43
	1.5	59	45(1.47)	0.830	59	−1.280	55.7	.20	37(1.20)	0.350	63	−1.265	53.0	.36
	1.6	48	33(1.07)	0.705	48	−1.275	44.0	.24	31(1.00)	0.455	50	−1.265	43.4	.32
	1.7	40	31(1.03)	1.080	40	−1.280	38.5	.14	25(0.83)	0.560	42	−1.265	36.9	.29
							$(\alpha, 1-\beta) = (.05, .85)$							
	1.4	85	54(1.80)	0.330	86	−1.643	73.9	.37	49(1.60)	−0.075	94	−1.628	69.5	.53
	1.5	65	43(1.43)	0.320	66	−1.640	56.9	.37	38(1.27)	−0.020	72	−1.624	54.3	.51
	1.6	53	34(1.13)	0.585	53	−1.641	47.3	.28	31(1.00)	0.105	58	−1.622	44.9	.46
	1.7	44	27(0.90)	0.555	45	−1.638	39.5	.29	27(0.87)	0.190	48	−1.622	38.5	.42
							$(\alpha, 1-\beta) = (.05, .9)$							
60	1.4	113	78(1.30)	0.665	113	−1.643	103.7	.25	70(1.17)	0.105	122	−1.628	97.9	.46
	1.5	85	58(0.97)	0.670	86	−1.641	78.7	.25	53(0.87)	0.320	91	−1.630	76.2	.37
	1.6	69	49(0.82)	0.710	69	−1.640	64.2	.24	42(0.70)	0.420	73	−1.628	62.5	.34
	1.7	58	41(0.67)	1.090	58	−1.643	55.1	.14	37(0.60)	0.525	61	−1.630	53.2	.30
							$(\alpha, 1-\beta) = (.1, .9)$							
	1.4	90	65(1.07)	0.900	91	−1.278	85.4	.18	57(0.95)	0.480	96	−1.265	83.2	.32
	1.5	68	45(0.73)	1.050	69	−1.277	64.8	.15	45(0.73)	0.695	71	−1.268	64.0	.24
	1.6	55	41(0.67)	1.085	55	−1.277	52.6	.14	35(0.58)	0.775	57	−1.266	51.9	.22
	1.7	46	29(0.48)	1.110	47	−1.274	43.8	.13	29(0.47)	0.920	48	−1.268	43.7	.18

r	n*	Minimax Design						Optimal Design					
		$n_1(\tau)$	c_1	n	c	EN	PET	$n_1(\tau)$	c_1	n	c	EN	PET
						$(\alpha, 1-\beta) = (.05, .85)$							
1.4	99	83(1.37)	0.585	99	−1.643	93.9	.28	60(0.98)	0.130	108	−1.624	85.8	.45
1.5	75	57(0.93)	0.725	75	−1.641	70.5	.23	46(0.75)	0.270	81	−1.622	66.8	.39
1.6	61	46(0.75)	1.100	61	−1.643	58.1	.14	37(0.60)	0.370	66	−1.621	54.8	.36
1.7	51	34(0.57)	0.870	51	−1.638	47.7	.19	29(0.48)	0.535	54	−1.624	46.5	.30

Table 2

Empirical size and power for the single-stage and minimax and optimal two-stage designs listed in Table 1

	r	Single-Stage		Minimax		Optimal	
		Size	Power	Size	Power	Size	Power
		$(\alpha, 1-\beta) = (0.05, 0.9)$					
30	1.4	0.043	0.881	0.044	0.876	0.046	0.874
	1.5	0.043	0.878	0.043	0.879	0.045	0.883
	1.6	0.039	0.880	0.041	0.889	0.043	0.882
	1.7	0.042	0.890	0.041	0.895	0.043	0.889
		$(\alpha, 1-\beta) = (0.1, 0.9)$					
	1.4	0.091	0.878	0.091	0.882	0.091	0.880
	1.5	0.092	0.880	0.091	0.881	0.091	0.881
	1.6	0.093	0.893	0.091	0.887	0.087	0.879
	1.7	0.091	0.890	0.087	0.886	0.091	0.890
		$(\alpha, 1-\beta) = (0.05, 0.85)$					
	1.4	0.044	0.827	0.044	0.834	0.044	0.827
	1.5	0.045	0.840	0.049	0.825	0.043	0.832
	1.6	0.043	0.841	0.043	0.838	0.043	0.835
	1.7	0.043	0.838	0.047	0.842	0.043	0.839
		$(\alpha, 1-\beta) = (0.05, 0.9)$					
60	1.4	0.044	0.889	0.044	0.883	0.047	0.889
	1.5	0.043	0.890	0.045	0.888	0.043	0.887
	1.6	0.043	0.892	0.041	0.891	0.046	0.896
	1.7	0.043	0.900	0.045	0.899	0.043	0.891
		$(\alpha, 1-\beta) = (0.1, 0.9)$					
	1.4	0.092	0.894	0.094	0.890	0.096	0.888
	1.5	0.094	0.890	0.094	0.894	0.098	0.886
	1.6	0.090	0.895	0.094	0.888	0.096	0.892
	1.7	0.092	0.901	0.092	0.896	0.093	0.891

	Single-Stage		Minimax		Optimal	
r	Size	Power	Size	Power	Size	Power
	$(\alpha, 1 - \beta) = (0.05, 0.85)$					
1.4	0.045	0.838	0.046	0.847	0.045	0.844
1.5	0.045	0.844	0.046	0.843	0.045	0.846
1.6	0.041	0.851	0.043	0.845	0.043	0.846
1.7	0.041	0.849	0.040	0.843	0.045	0.847