Published in final edited form as:
J Biopharm Stat. 2012 ; 22(3): 485–495. doi:10.1080/10543406.2010.550701.

On Sample Size Calculation for Comparing Survival Curves under General Hypothesis Testing

Sin-Ho Jung and **Shein-Chung Chow**
Department of Biostatistics and Bioinformatics Duke University Durham, North Carolina

SUMMARY

The log-rank test is commonly used to test the equivalence of two survival distributions under right censoring. Jung et al. (2005) proposed a modified log-rank test for noninferiority trials and its corresponding sample size calculation. In this paper, we extend the use of the modified log-rank test for clinical trials with various types of non-conventional study objectives and propose its sample size calculation under general null and alternative hypotheses. The proposed formula is so flexible that we can specify any survival distributions and accrual pattern. The proposed methods are illustrated with designing real clinical trials. Through simulations, the modified log-rank test and the derived formula for sample size calculation are shown to have satisfactory small sample performance.

Keywords

Accrual rate; Non-inferiority trial; Proportional hazards

1 INTRODUCTION

In clinical trials, it is often of interest to compare the survival distribution of an experimental therapy, denoted by arm 2, with that of a standard (or control) therapy, denoted by arm 1. In practice, a proportional hazards model is assumed. Let $\Lambda_k(t)$ denote the cumulative hazard function of arm $k(= 1, 2)$ and $\Delta = \Lambda_2(t)/\Lambda_1(t)$ the hazard ratio. By the standard log-rank test (Peto and Peto, 1972), we usually test a null hypothesis that two arms have identical survival distributions against an alternative hypothesis that the experimental arm has a longer survival, i.e. $H_0 : \Delta = 1$ against $H_1 : \Delta < 1$ using the standard log-rank test. We will refer to such trials as superiority trials. Sample size for a superiority trial can be calculated under the alternative hypothesis that $H_1 : \Delta = \Delta_1$ for a pre-specified clinically significant hazard ratio $\Delta_1(< 1)$ (see e.g., George and Desu, 1973; Schoenfeld, 1983; Lakatos, 1988; Yateman and Skene, 1992).

In many cases, the experimental therapy may be less extensive, less toxic or less expensive than the standard therapy. In this case, the former may be acceptable so long as there is evidence that it is not worse than the latter. Thus, we may want to prove that an experimental therapy is stochastically equal or not inferior to the standard therapy. That is, with a given non-inferiority margin (the maximal hazard ratio of clinical insignificance) $\Delta_0(> 1)$, $H_0 : \Delta \geq \Delta_0$ against $H_1 : \Delta < \Delta_0$ using the non-inferiority log-rank test, e.g., Jung et al. (2005). In a non-inferiority trial, the experimental therapy can be hardly as efficacious as the control, so that we often want to show that the survival distributions are identical

Contact information for the first author: (phone) 919-668-8658; jung0005@mc.duke.edu.

between two arms. Thus, we calculate the sample size under $H_1 : \Delta = 1$, see e.g., Chow, Shao and Wang (2003) and Jung et al. (2005).

Generalizing superiority and non-inferiority trials, we consider calculating the sample size for $H_0 : \Delta = \Delta_0$ against $H_1 : \Delta = \Delta_1$ with $\Delta_0 \geq \Delta_1$. We have $\Delta_0 = 1$ for superiority trials and $\Delta_1 = 1$ for non-inferiority trials. In general, a non-inferiority margin is so small (e.g., $\Delta_0 \approx 1.1$ contrary to a usual superiority margin $\Delta_1 \approx 1.5$), so that the required sample size becomes very big. However, possibly due to lower adverse event, a moderate chemotherapy may be able to extend the survival of cancer patients compared to a standard aggressive chemotherapy. In this case, by using Δ_0 larger than 1 and Δ_1 smaller than 1, we can lower the sample size, and the effect size Δ_0/Δ_1 may be close to that of a superiority trial.

For illustration, we consider an example concerning a randomized phase II trial for patients with locally advanced nasopharyngeal carcinoma. A high dose (arm 1) chemotherapy A concurrent with radiotherapy is a standard treatment for the patient population. However, the high dose chemotherapy is not well tolerated by most patients. So, we want to investigate a low dose (arm 2) concurrent chemotherapy A administered more frequently (weekly vs. every 3 weeks). It is known that the high dose chemotherapy has a 3-year progression-free survival (PFS) of $S_1(3) = 0.75$, corresponding to a hazard rate of $\lambda_1 = 0.096$ under an exponential model. In the new randomized phase II trial, we will not be interested in the low dose chemotherapy if its 3-year PFS is $S_2(3) = 0.65$ or lower, corresponding to $\Delta_0 = 1.4974$, and will be very much interested in it if its 3-year PFS is $S_2(3) = 0.8$ or higher, corresponding to $\Delta_1 = 0.7757$ under an exponential model. It is believed that the low expected adverse events and more frequent administration will possibly extend the PFS for arm 2 slightly beyond that of arm 1.

Note that the cumulative hazard function for the control arm $\Lambda_1(t)$ is identical under both H_0 and H_1. Assuming the proportional hazards model, the cumulative hazard function for the experimental arm is $\Lambda_2(t) = \Delta_0\Lambda_1(t)$ under H_0 and $\Lambda_2(t) = \Delta_1\Lambda_1(t)$ under H_1. Hence, the survival distributions for a sample size calculation can be specified by $(\Lambda_1(t), \Delta_0, \Delta_1)$.

In the next section, a generalized log-rank test is derived under general null and alternative hypotheses. A formula for sample size calculation is proposed in Section 3. The practical application of the proposed sample size calculation method is presented in Section 4. A brief discussion is given in the last section.

2 THE GENERALIZED LOG-RANK TEST

Let n_k denote the sample size in arm k and T_{ki} the survival time for subject i in arm k $(1 \leq i \leq n_k;\ k = 1, 2)$. Then we usually observe (X_{ki}, δ_{ki}), where X_{ki} is the minimum of T_{ki} and the censoring time and δ_{ki} is an event indicator taking 1 if the subject had an event and 0 otherwise.

For arm k, $T_{k1}, \ldots, T_{k,nk}$ are IID with hazard function $\lambda_k(t)$. Under the proportional hazards assumption, $\Delta = \lambda_2(t)\lambda_1(t)$ denotes the hazard ratio. By Cox (1972), the partial score function $W(\Delta)$ and the information function $\sigma_n^2(\Delta)$ are given as

$$W(\Delta) = \int_0^\infty \frac{Y_1(t)\,Y_2(t)}{Y_1(t) + \Delta Y_2(t)} \left\{\Delta d\widehat{\Lambda}_1(t) - d\widehat{\Lambda}_2(t)\right\},$$

and

$$\sigma_n^2(\Delta) = \Delta \int_0^\infty \frac{Y_1(t)\, Y_2(t)}{\{Y_1(t) + \Delta Y_2(t)\}^2} dN(t),$$

where $\widehat{\Lambda}_k(t) = \int_0^t Y_k^{-1}(t)\, dN_k(t)$ is the Aalen-Nelson estimator (Aalen, 1978; Nelson, 1969) for the cumulative hazard function $\Lambda_k(t) = \int_0^t \lambda_k(s)\, dS$, $Y_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \geq t)$ and $N_k(t) = \sum_{i=1}^{n_k} \delta_{ki} I(X_{ki} \leq t)$ are the at-risk process and the event process for group k, respectively, $N(t) = N_1(t) + N_2(t)$, and $I(\cdot)$ is the indicator function. Note that $W(1)$ is the standard log-rank test statistic (Peto and Peto, 1972). We call $W(\Delta)$ the generalized log-rank test in this paper.

As $n \to \infty$ and $n_k/n \to p_k \in (0, 1)$, $W(\Delta_0)/\sigma(\Delta_0)$ converges to the standard normal distribution under $H_0 : \Delta = \Delta_0$, see e.g. Fleming and Harrington (1991). Let $z_{1-\alpha}$ denote the $100(1 - \alpha)$ percentile for the standard normal distribution. Then we reject H_0, in favor of $H_1 : \Delta < \Delta_0$, if $W(\Delta_0)/\sigma_n(\Delta_0) > z_{1-\alpha}$ with one-sided type I error rate α.

The partial MLE $\widehat{\Delta}$ is obtained by solving $W(\Delta) = 0$. The MLE is a consistent estimator of the true hazard ratio. Furthermore, by the asymptotic linearity, the score-type test statistic $W(\Delta_0)/\sigma_n(\Delta_0)$ is asymptotically equivalent to the Wald-type test statistic $\sigma_n(\Delta_0)\left(\widehat{\Delta} - \Delta_0\right)$, so that our sample size formula derived in the following section is valid for both types of test statistics.

3 SAMPLE SIZE FORMULA

We want to estimate the sample size n under a specific alternative hypothesis that $H_1 : \Delta = \Delta_1 (< \Delta_0)$ with a desired power. Jung et al. (2005) proposed a sample size formula with $\Delta_0 > 1$ and $\Delta_1 = 1$ for designing non-inferiority trials. This paper is to extend their formula for general Δ_0 and Δ_1 with $\Delta_1 < \Delta_0$. Let $p_k = n_k/n$ denote the allocation proportion for arm k and $n = n_1 + n_2$.

The asymptotic results in this section are derived under H_1. Let $S_k(t)$ and $f_k(t) = -\partial S_k(t) \partial t$ denote the survival and probability density functions, respectively, for arm k under H_1. Note that $S_2(t) = S_1(t)^{\Delta 1}$ and $f_2(t) = \Delta_1 f_1(t) S_1(t)^{\Delta 1 - 1}$ under H_1. For a censoring variable C, let $G(t) = P(C \geq t)$ denote the survivor function of the censoring distribution which is common in two arms. By Jung et al. (2005), $\sigma_n^2(\Delta)$ is asymptotically equivalent to $n\sigma^2(\Delta)$, where

$$\sigma^2(\Delta) = \Delta p_1 p_2 \int_0^\infty \frac{G(t) S_1(t) S_2(t) \{p_1 f_1(t) + p_2 f_2(t)\}}{\{p_1 S_1(t) + \Delta p_2 S_2(t)\}^2} dt. \quad (1)$$

By the definition of $W(\Delta)$, we have

$$W(\Delta_0) - W(\Delta_1) = n^{-1/2}(\Delta_0 - \Delta_1) \int_0^\infty \frac{Y_1(t)\, Y_2(t)}{\{Y_1(t) + \Delta_0 Y_2(t)\}\{Y_1(t) + \Delta_1 Y_2(t)\}} dN(t),$$

which is asymptotically equivalent to $n\omega$, where

$$\omega = (\Delta_0 - \Delta_1)\, p_1 p_2 \int_0^\infty \frac{G(t) S_1(t) S_2(t) \{p_1 f_1(t) + p_2 f_2(t)\}}{\{p_1 S_1(t) + \Delta_0 p_2 S_2(t)\}\{p_1 S_1(t) + \Delta_1 p_2 S_2(t)\}} dt. \quad (2)$$

The integrals in (1) and (2) are calculated using a numerical method.

Under H_1, the generalized log-rank test statistic can be expressed as

$$\frac{W(\Delta_0)}{\sigma_n(\Delta_0)}=\frac{W(\Delta_1)}{\sigma_n(\Delta_1)}\times\frac{\sigma_n(\Delta_1)}{\sigma_n(\Delta_0)}+\frac{W(\Delta_0)-W(\Delta_1)}{\sigma_n(\Delta_0)},$$

which, from (1) and (2), can approximated by

$$\frac{W(\Delta_1)}{\sigma_1}\times\frac{\sigma_1}{\sigma_0}+\frac{\omega\sqrt{n}}{\sigma_0},$$

where $\sigma_l^2=\sigma^2(\Delta_l)$ for $l = 0, 1$.

Suppose that we want to estimate the sample size for detecting $H_1 : \Delta = \Delta_1$ with a power of $1 - \beta$ by the generalized log-rank test with a one-sided α at $H_0 : \Delta = \Delta_0$, i.e.

$$1-\beta=P\left(\frac{W(\Delta_0)}{\sigma_n(\Delta_0)}>z_{1-\alpha}|H_1\right)\approx P\left(\frac{W(\Delta_1)}{\sigma_1}\times\frac{\sigma_1}{\sigma_0}+\frac{\omega\sqrt{n}}{\sigma_0}>z_{1-\alpha}|H_1\right).$$

Since $W(\Delta_1)/\sigma_1$ is approximately $N(0, 1)$ under H_1, we have

$$-z_{1-\beta}=\left(z_{1-\alpha}-\frac{\omega\sqrt{n}}{\sigma_0}\right)\frac{\sigma_0}{\sigma_1}.$$

Hence the required sample size is obtained as

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

Note that ω, σ_0^2 and σ_1^2 are functions of the survival distributions $S_1(t)$ and $S_2(t)$ under H_1, and the common censoring distribution $G(t)$. They are calculated using numerical methods.

The power of the generalized log-rank test roughly depends on the number of events, rather than the number of patients, so that one may want to calculate the expected number of events at the final data analysis. The number of events D is calculated as in the standard log-rank test, i.e. $D = n(p_1d_1 + p_2d_2)$, where $d_k=1+\int_0^\infty S_k(t)\,dG(t)$, see e.g. Schoenfeld (1983). Note that if the survival distributions are shorter, then the required sample size is smaller.

3.1 SOME PRACTICAL INPUT PARAMETER SETTINGS

In this section, we will illustrate the application of the proposed methods when designing a clinical trial for comparing survival distributions with right censoring as follows.

(A) An exponential distribution can be uniquely specified by a single parameter, such as the hazard rate, the median or the survival probability at a chosen time point. Furthermore, the family of exponential distributions fit real survival data relatively well. So, we often specify the survival distributions using exponential distributions with hazard rates λ_k, $S_k(t) = \exp(-\lambda_k t)$.

(B) When a trial is open, patients are usually uniformly recruited during an accrual period a and additional follow-up period b. In this case, we have

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

Note that these assumptions can be easily extended to non-exponential survival models and a non-uniform censoring (accrual) distribution. Further, we assume that there is no loss to follow-up by (B), but it can be easily extended to account for possible loss to follow-up, see Jung, Kim and Chow (2008).

3.2 WHEN AN ACCRUAL RATE IS SPECIFIED

In Section 3.1, we assume that the accrual period is known. In designing a clinical trial, however, we can estimate the accrual pattern, rather than an accrual period, based on the experience from previous studies on the same patient population. For example, we may assume that patients are expected to be entered to the study at an rate of r during accrual period based on the number of patients treated by the member sites recently. In this case, (B) is replaced by:

(B′) Patients are accrued following a Poisson distribution with rate r, and are followed for a period b after the completion of accrual.

With (λ_1, Δ_0, Δ_1, α, $1-\beta$, p_1, b) specified, $\omega=\omega(a)$ and $\sigma_1=\sigma_1(a)$ are functions of a. Hence, under (A) and (B), (3) is expressed as

$$n=\frac{\left\{\sigma_0(a)z_{1-\alpha}+\sigma_1(a)z_{1-\beta}\right\}^2}{\omega^2(a)}. \tag{4}$$

On the other hand, under the Poisson accrual distribution (B′), we have

$$n=a\times r. \tag{5}$$

By equating the right hand sides of (4) and (5), we obtain an equation on a,

$$a\times r=\frac{\left\{\sigma_0(a)z_{1-\alpha}+\sigma_1(a)z_{1-\beta}\right\}^2}{\omega^2(a)}. \tag{6}$$

Equation (6) is solved using a numerical method, such as bisection method. Let a^* denote the solution to equation (6). Then, given an accrual rate r, instead of an accrual period a, we obtain the sample size by $n=a^*\times r$. The procedure for a sample size calculation may be summarized as follows.

[1] Specify the input variables:

- Type I and II error probabilities, (α, β)
- Allocation proportions, p_1, p_2
- Hazard rate λ_1 for the control arm under exponential survival model, and hazard ratios Δ_0 and Δ_1 under H_0 and H_1, respectively
- Accrual rate r, and follow-up period b

[2] Solve

$$a \times r = \frac{\left\{\sigma_0(a)\, z_{1-\alpha} + \sigma_1(a)\, z_{1-\beta}\right\}^2}{\omega^2(a)}$$

with respect to a using the bisection method, where

$$\sigma_0^2(a) = \Delta_0 \lambda_1 p_1 p_2 \int_0^{a+b} \frac{G(t)\, e^{-\lambda_1(1+\Delta_1)t}\left(p_1 e^{-\lambda_1 t} + \Delta_1 p_2 e^{-\Delta_1 \lambda_1 t}\right)}{\left(p_1 e^{-\lambda_1 t} + \Delta_0 p_2 e^{-\Delta_1 \lambda_1 t}\right)^2} dt$$

$$\sigma_1^2(a) = \Delta_1 \lambda_1 p_1 p_2 \int_0^{a+b} \frac{G(t)\, e^{-\lambda_1(1+\Delta_1)t}}{p_1 e^{-\lambda_1 t} + \Delta_1 p_2 e^{-\Delta_1 \lambda_1 t}} dt$$

$$\omega(a) = (\Delta_0 - \Delta_1)\, \lambda_1 p_1 p_2 \int_0^{a+b} \frac{G(t)\, e^{-\lambda_1(1+\Delta_1)t}}{p_1 e^{-\lambda_1 t} + \Delta_0 p_2 e^{-\Delta_1 \lambda_1 t}} dt$$

and

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

[3] For the solution $a = a^*$ to the equation in [2], the required sample size is given as $n = a^* \times r$.

3.3 UNDER GENERAL ACCRUAL PATTERNS

In (B′), we assume a constant accrual rate over the whole accrual period. Usually in a multi-center trial, however, it takes a while (e.g. 1 to 2 years) until the study is approved by the institutional review boards of the study centers and the accrual rate is stabilized. Let $h(t)$ for $t \ge 0$ be the function representing the pattern of patient accrual over time period. For example, if we expect that the accrual will be linearly increasing for the first a_0 years, called a run-in time, and maintain a constant accrual of r per year after then, then we have

$$h(t) = \begin{cases} rt & \text{for } 0 \le t \le a_0 \\ r & \text{for } t > a_0 \end{cases} \tag{7}$$

This is similar to the piecewise linear accrual pattern considered by Yateman and Skene (1992). Given an accrual period of a and an accrual function $h(t)$, (5) is extended to

$$n = \int_0^a h(t)\, dt. \tag{8}$$

The accrual function $h(t)$ is related to the censoring distribution function $G(t)$ as follows. Let E denote the entry time of a patient in the study. The probability density function of E is expressed as

$$\begin{cases} h(t)/\int_0^a h(s)\,ds & \text{if } 0 \le t \le a \\ 0 & \text{otherwise} \end{cases}$$

Since $C = a + b - E$, the survivor function of C, $G(t) = P(C \ge t)$, is given as

$$G(t) = P(E \le a+b-t) = \begin{cases} 1 & \text{if } t \le b \\ \int_0^{a+b-t} h(s)\,ds / \int_0^a h(s)\,ds & \text{if } b < t \le a+b \\ 0 & \text{if } t > a+b \end{cases} \quad (9)$$

By calculating $\sigma_0^2(a)$, $\sigma_1^2(a)$ and $\omega(a)$ using this $G(t)$ and equating the right hand sides of (4) and (8), we obtain an equation on a for a general accrual pattern. Using the solution a^* to this equation, we obtain the required sample as $n = \int_0^{a^*} h(t)\,dt$.

Combining (8) and (9), we calculate the probability to observe an event from a patient in arm k by $d_k = 1 - n^{-1}\int_0^\infty S_k(t)\,h(a+b-t)\,dt$.

For example, for the piecewise linear accrual pattern of (7), we have

$$\int_0^a h(s)\,ds = \begin{cases} ra^2/2 & \text{if } 0 \le a \le a_0 \\ ra_0^2/2 + r(a-a_0) & \text{if } a > a_0 \end{cases}$$

and

$$\int_0^{a+b-t} h(s)\,ds = \begin{cases} ra_0^2/2 + r(a-a_0) & \text{if } t < b \\ ra_0^2/2 + r(a+b-a_0-t) & \text{if } b \le t < a+b-a_0 \\ r(a+b-t)^2/2 & \text{if } a+b-a_0 \le t < a+b \\ 0 & \text{if } t \ge a+b \end{cases}$$

The sample size for a general accrual pattern can be summarized as follows. Note that the term $\int_0^a h(s)\,ds$ is cancelled out in the equation for sample size calculation.

[1] Specify the input variables:

- Type I and II error probabilities, (α, β)
- Allocation proportions, p_1, p_2
- Hazard rate λ_1 for the control arm under exponential survival model, and hazard ratios Δ_0 and Δ_1 under H_0 and H_1, respectively
- Accrual pattern $h(t)$, and follow-up period b

[2] Solve

$$1 = \frac{\left\{\tilde{\sigma}_0(a)\,z_{1-\alpha} + \tilde{\sigma}_1(a)\,z_{1-\beta}\right\}^2}{\tilde{\omega}^2(a)}$$

with respect to a using the bisection method, where

$$\tilde{\sigma}_0^2(a) = \Delta_0 \lambda_1 p_1 p_2 \int_0^{a+b} \frac{\tilde{G}(t)\, e^{-\lambda_1(1+\Delta_1)t} \left(p_1 e^{-\lambda_1 t} + \Delta_1 p_2 e^{-\Delta_1 \lambda_1 t}\right)}{\left(p_1 e^{-\lambda_1 t} + \Delta_0 p_2 e^{-\Delta_1 \lambda_1 t}\right)^2} dt$$

$$\sigma_1^2(a) = \Delta_1 \lambda_1 p_1 p_2 \int_0^{a+b} \frac{\tilde{G}(t)\, e^{-\lambda_1(1+\Delta_1)t}}{p_1 e^{-\lambda_1 t} + \Delta_1 p_2 e^{-\Delta_1 \lambda_1 t}} dt$$

$$\omega(a) = (\Delta_0 - \Delta_1)\, \lambda_1 p_1 p_2 \int_0^{a+b} \frac{\tilde{G}(t)\, e^{-\lambda_1(1+\Delta_1)t}}{p_1 e^{-\lambda_1 t} + \Delta_0 p_2 e^{-\Delta_1 \lambda_1 t}} dt$$

and

$$\tilde{G}(t) = \int_0^{a+b-t} h(s)\, ds$$

[3] For the solution $a = a^*$ to the equation in [2], the required sample size is given as $n = \int_0^{a^*} h(t)\, dt$.

4 EXAMPLES

We consider some real examples of cancer trials that can be designed using the proposed sample size method.

Example 1

We consider the locally advanced nasopharyngeal carcinoma study that is discussed in the introduction section. Under an equal allocation $p_1 = p_2 = 0.5$, an annual accrual rate of $r = 55$ patients and $b = 3$ years of additional follow-up after completion of accrual, we need $n = 138$ patients ($n_k = 69$ per arm) for $1 - \beta = 80\%$ power for detecting $H_1 : \Delta_1 = 0.7757$ by the generalized log-rank test with one-sided $\alpha = 10\%$ for $H_0 : \Delta_0 = 1.4974$. We expect about 42 events (progressions or deaths) at the final data analysis.

In order to validate this sample size, 10,000 samples of size $n = 138$ are generated from the design setting of (p_1, λ_1, Δ_1, r, b) = (0.5, 0.096, 0.7757, 55, 3) as in the above sample size calculation, and the generalized log-rank test with one-sided $\alpha = 0.1$ is applied to each simulated sample. The empirical power is obtained as 0.7920, which is very close to the nominal $1 - \beta = 0.8$. We also conduct simulations to check the small sample property of the generalized log-rank test by generating 10,000 samples of size $n = 138$ under $H_0 : \Delta_0 = 1.4974$. The empirical type I error rate is 0.0923, which is very close to the nominal $\alpha = 0.1$.

Extending this example, suppose that the accrual is expected to linearly increase for the first year ($a_0 = 1$) and maintain a constant accrual rate of $r = 55$ per year from the beginning of the second year. Then, we obtain $a^* = 3$ and $n = 136$. With the slightly longer accrual period (3 years vs. 138/55 = 2.5 years), this design setting requires a slightly smaller sample size than the setting with a constant accrual rate which was considered above.

Example 2

Chemotherapy B has been a standard regimen for patients with non-bulky stage I and II Hodgkin lymphoma. In a previous study on 6 cycles of B, each patient had a FDG-PET (fluorodeoxyglucose positron-emission tomography) imaging after 2 cycles of B. It was found that the patients with a negative PET image (group 1) and those with a positive PET image (group 2) had a 3-year PFS, defined as the time to disease progression or death, of $S_1(3) = 0.86$ and $S_2(3) = 0.52$, respectively, and the hazard ratio λ_2/λ_1, was estimated as $\Delta_0 = 4.3$.

In a new (single-arm) phase II trial, the patients with a negative PET image after 2 cycles of B will be treated by additional 4 cycles of the standard chemotherapy B, whereas those with a positive PET image after 2 cycles of B will be treated by 4 cycles of a more aggressive chemotherapy C. By the PET-guided chemotherapy strategy, it is believed that the 3-year PFS of the PET-positive patients can be increased to $S_2(3) = 0.74$, from 0.52, resulting in a hazard ratio of $\Delta_1 = 2$. Note that the group 1 patients will receive the same treatment as that of the previous study. Based on an exponential distribution model for PFS, the annual hazard rate for group 1, corresponding to $S_1(3) = 0.86$, is $\lambda_1 = 0.05$. The previous study observed about $p_2 = 20\%$ of PET-positivity. Assuming an annual accrual rate of $r = 60$ patients and $b = 3$ years of additional follow-up after completion of accrual, we need $n = 195$ patients for $1 - \beta = 90\%$ power for detecting $H_1 : \Delta_1 = 2$ by the generalized log-rank test with one-sided $\alpha = 10\%$ for $H_0 : \Delta_0 = 4.3$. Under this specific alternative hypothesis, we expect about 47 events (progressions or deaths) at the data analysis. Note that this is not a randomized trial, so that the resulting allocation proportions may be slightly different from the specified $(p_1, p_2) = (0.8, 0.2)$. If the observed allocation proportions are closer to 0.5, the planned sample size has enough power. However, if they are farther from 0.5, we may consider checking the statistical power of the study based on the observed power while other design parameters are fixed at the values used when designing the study. In this power checking, we may use the estimated censoring distribution $\widehat{G}(t)$ based on the accrual times of patients too.

Simulation studies are conducted to evaluate the calculated sample size under the above design settings of H_0 and H_1, respectively. Using 10,000 simulation samples of size $n = 132$ under each hypothesis, the empirical type I error rate and power are observed as 0.0984 (to be compared to $\alpha = 0.1$) and 0.8749 (to be compared to $1 - \beta = 0.9$), respectively.

5 DISCUSSIONS

We have proposed a sample size calculation method of the generalized log-rank test for hypotheses $H_0 : \Delta = \Delta_0$ vs. $H_0 : \Delta = \Delta_1$ $(\Delta_0 > \Delta_1)$. The proposed sample size formula is identical to that of the log-rank test by Schoenfeld (1983) and Lakatos (1988) if $\Delta_0 = 1$, and that of the futility log-rank test by Jung et al. (2005) if $\Delta_1 = 1$.

From (1)–(3), we observe that n is complicatedly dependent on the allocation proportions p_k, but the total sample size n is approximately minimized when $p_1 = p_2 = 1/2$. The sample size increases in λ_1, r and $1 - \beta$, and decreases in Δ_0/Δ_1, b and α.

Our sample size method is so flexible that we can assume any survival distributions for two arms and any accrual pattern during study period. As a typical setting, we specify type I error rate α, power $1 - \beta$, allocation proportion for arm 1 p_1, accrual rate r, follow-up period b, and survival distributions for two arms under H_0 and H_1 (or the survival distribution of the control arm and the hazard ratios Δ_0 and Δ_1 under H_0 and H_1. respectively); and we obtain the required sample size using our formula. However, we can obtain any of these input and output parameters given other parameter values. Furthermore, accrual rate r can be

replaced by an expected accrual rate *a* in these calculations. The general accrual pattern approach will be very useful when checking the statistical power in the middle of a study with the realized accrual times.

Acknowledgments

This research was supported by a grant from the National Cancer Institute (CA142538-01).

REFERENCES

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

Chow, SC.; Shao, J.; Wang, H. Sample Size Calculation in Clinical Research. 2nd Edition. Chapman and Hall; CRC Press; Taylor & Francis; New York: 2008.

Cox DR. Regression models and life tables (with discussion). Journal of Royal Statistical Society B. 1972; 34:187–220.

Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. Wiley; New York: 1991.

George SL, Desu MM. Planning the size and duration of a trial studying the time to some critical event. Journal of Chronic Disease. 1973; 27:15–24.

Jung SH, Kang SJ, McCall L, Blumenstein B. Sample size computation for noninferiority log-rank test. Journal of Biopharmaceutical Statistics. 2005; 15:957–967. [PubMed: 16279354]

Jung SH, Kim C, Chow SC. Sample size calculation for the log-rank tests for multi-arm trials with a control. Journal of Korean Statistical Society. 2008; 37:11–22.

Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. Biometrics. 1988; 44:229–241. [PubMed: 3358991]

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

Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). Journal of the Royal Statistical Society, Series A. 1972; 135:185–206.

Schoenfeld DA. Sample size formula for the proportional hazards regression model. Biometrics. 1983; 39:499–503. [PubMed: 6354290]

Yateman NA, Skene AM. Sample size for proportional hazards survival studies with arbitrary patient entry and loss to follow-up distributions. Statistics in Medicine. 1992; 11:1103–1113. [PubMed: 1496198]