Published in final edited form as:

Stat Biopharm Res. 2017 ; 9(1): 35–43. doi:10.1080/19466315.2016.1189355.

Group Sequential Survival Trial Design and Monitoring Using the Log-Rank Test

Jianrong Wu and Xiaoping Xiong

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

Abstract

For randomized group sequential survival trial designs with unbalanced treatment allocation, the widely used Schoenfeld formula is inaccurate, and the commonly used information time as the ratio of number of events at interim look to the number of events at the end of trial can be biased. In this paper, a sample size formula for the two-sample log-rank test under the proportional hazards model is proposed that provides more accurate sample size calculation for unbalanced survival trial designs. Furthermore, a new information time is introduced for the sequential survival trials such that the new information time is more accurate than the traditional information time when the allocation of enrollments is unbalanced in groups. Finally, we demonstrate the monitoring process using the sequential conditional probability ratio test and compare it with two other well known group sequential procedures. An example is given to illustrate unbalanced survival trial design using available software.

Keywords

group sequential trial design; log-rank test; information time; proportional hazards model; sample size; time-to-event; unequal allocation

1 Introduction

Time-to-event is a common and important primary endpoint of phase III oncology clinical trials. The efficacy of trials can be tested by examining the difference in the survival distributions between treatment groups. The most popular test statistic is the nonparametric log-rank test, which is asymptotically efficient under the proportional hazards model. Randomized phase III trial designs with time-to-event endpoints have been studied extensively by many researchers. Sample size formulae for the log-rank test were derived by Schoenfeld (1981, 1983), Freedman (1982), Lachin and Foulkes (1986), Lakatos (1988), Collett (2003) and many others for two-arm trials; by Ahnn and Anderson (1995, 1998), Halabi and Singh (2004), Barthel et al. (2006) and others for multiple-arm trials. Among them, Schoenfeld formula is widely used. It is common practice in randomized trials to allocate equal sample sizes to the treatment groups to gain a higher statistical power. However, in some situations, despite the loss of power, investigators may still prefer an unbalanced design. For example, when the trial is designed to compare a new treatment to the standard therapy, and, based on a previous study, the investigators believe the new treatment to be at least as beneficial as the standard therapy, then it might be unethical to

randomize half of the patients to the potentially inferior standard treatment. Alternatively, the investigators may simply want to gain more experience or collect as much safety and efficacy data as possible on the new treatment and thus enroll more subjects in this group. For unbalanced trial designs, Schoenfeld formula can be inaccurate and tend to either underestimate or overestimate the sample size (Hsieh, 1992).

Survival trials often take a long time to finish, and data are accumulated gradually during the course of the study. For ethical reasons, such trials are often monitored for early stopping if a large difference in the survival distributions between two groups is observed during an interim analysis. Various group sequential methods have been developed in the past few decades, e.g., by Haybittle (1971), Pocock (1977), O'Brien and Fleming (1979), Lan and DeMets (1983), Whitehead and Stratton (1983), Xiong (1995), Jennison and Turnbull (1997), Lakatos (2002) and many others. Most of these methods have been comprehensively reviewed by Jennison and Turnbull (2000). The key information in survival trial monitoring is the determination of the information time, which is often defined as the ratio of the number of events to be observed at an interim analysis to the total number of events in the entire trial (Lan and Lachin, 1990). However, the information time tends to be biased either upward or downward for an unbalanced allocation design.

Motivated by these existing problems in the current common practice for unbalanced randomized survival trial design, a sample size formula for the log-rank test under the proportional hazards model is proposed that provides more accurate sample size calculation for an unbalanced allocation trial design. Furthermore, a new information time is proposed for the log-rank statistic so that this information time is nearly unbiased. The rest of the paper is organized as follows. In section 2, a sequential log-rank test statistic is introduced for group sequential trial design. A sample size formula for the log-rank test is proposed in section 3. In section 4, a new information time for the log-rank test is derived. A multistage group sequential procedure is discussed in section 5. Simulation studies are conducted in section 6 to evaluate the performance of the proposed sample size formula and information time. In section 7, an example is given to illustrate the proposed methods. Final discussion and concluding remarks are presented in section 8.

2 Sequential Log-rank Test

Suppose there are two groups, control and treatment groups which are designated groups 1 and 2, respectively. Assuming the survival distributions of the two groups satisfy the proportional hazards model

$$S_1(t)=[S_2(t)]^\delta, \quad (1)$$

where δ is the unknown hazard ratio. The hypothesis of improvement in the survival distribution of the treatment group compared to that of the control group can be expressed as

$$H_0:\delta \leq 1 \quad \text{vs.} \quad H_1:\delta > 1. \quad (2)$$

A well-known test statistic for testing the above hypothesis is the log-rank test. To consider the sequential trial design, we introduce a sequential log-rank test statistic as follows. Suppose there are r distinct event times in n subjects, $t_{(1)} < t_{(2)} \cdots < t_{(r)}$, among the two groups. At an event time $t_{(j)}$, d_{1j} and d_{2j} events occur for the control and treatment groups, respectively, with n_{1j} and n_{2j} subjects being at risk in the two groups just prior to $t_{(j)}$, for $1 \leq j \leq r$. Thus, there are $d_j = d_{1j} + d_{2j}$ events at $t_{(j)}$ among a total of $n_j = n_{1j} + n_{2j}$ subjects. Let τ be the study duration and $t (< \tau)$ be the calendar time from the start of study enrollment to the interim look, and let $n_{ij}(t)$ and $d_{ij}(t)$ represent, respectively, the parts of n_{ij} and d_{ij} with calendar times of entry no later than t , for $i = 1, 2$. Let $n_j(t) = n_{1j}(t) + n_{2j}(t)$, $d_j(t) = d_{1j}(t) + d_{2j}(t)$, and $e_{1j}(t) = n_{1j}(t)d_{1j}(t)/n_j(t)$, for $1 \leq j \leq r$. Then, the log-rank score at an interim look at calendar time t is

$$U(t) = \sum_{j=1}^r \{d_{1j}(t) - e_{1j}(t)\},$$

which is approximately normal with a mean $\log(\delta) V(t)$ and variance of $V(t)$, where

$$V(t) = \sum_{j=1}^r \frac{n_{1j}(t)n_{2j}(t)d_j(t)(n_j(t) - d_j(t))}{n_j^2(t)(n_j(t) - 1)}. \quad (3)$$

Then, the sequential log-rank test at calendar time t is given by

$$L(t) = U(t)/V^{1/2}(t), \quad (4)$$

which is approximately normal with a mean $\log(\delta) V^{1/2}(t)$ and unit variance. Thus, we reject null hypothesis at the end of trial if $L(\tau) > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1 - \alpha$ percentile of the standard normal distribution. Define $B_{t^*} = U(t)/V^{1/2}(\tau)$, then $B_{t^*} \sim N(\theta t^*, t^*)$ is approximately a Brownian motion with drift parameter $\theta = \log(\delta) V^{1/2}(\tau)$ and information time $t^* = V(t)/V(\tau)$, where $V(\tau)$ is the value of $V(t)$ at $t = \tau$. The Brownian motion property of the sequential log-rank test has been derived by Tsiatis (1982), Sellke and Siegmund (1983), and Slud (1984).

3 Sample Size Formula

A key step in designing a survival clinical trial is to calculate the required number of events or sample size based on testing the hypothesis (2) under the proportional hazards model (1). A widely used total number events formula for the log-rank test is given as follow

$$d(\tau) = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\omega_1 \omega_2 [\log(\delta)]^2}, \quad (5)$$

where α is the type I error (one-sided test), $1 - \beta$ is the power, and ω_1 and $\omega_2 = 1 - \omega_1$ are the probabilities of subjects assigned to treatment groups 1 and 2, respectively (Schoenfeld, 1981, 1983; Collett, 2003). Let $p_i(\tau)$ be the probability of a patient in the i^{th} group having an event during the study period $[0, \tau]$ for $i = 1, 2$, and let $P(\tau) = \omega_1 p_1(\tau) + \omega_2 p_2(\tau)$ be the probability of an event for a patient randomly chosen from the two groups with probabilities of ω_1 and ω_2 , respectively. Then, the total number of patients required for the trial is $n = d(\tau)/P(\tau)$, which yields the following sample size formula

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\omega_1 \omega_2 [\log(\delta)]^2 P(\tau)}. \quad (6)$$

This formula was first derived by Schoenfeld (1981) and has been widely used in randomized phase III survival trial designs. However, it tends to either underestimate or overestimate the sample size when the allocation of patients to treatment groups is unbalanced (Hsieh, 1982).

In an unpublished manuscript on the log-rank test, Xiong (2014) obtained a precise sample size formula over the study duration $[0, \tau]$ for the log-rank test that is given by

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{R(\tau)P(\tau)},$$

where $R(\tau)$ is a quantity for log-rank test similar to μ^2/σ^2 for the test for normal distribution and is a function of the survival distributions, the hazard ratio, the entry time distribution, the censoring distribution, and the allocation proportions of subjects in the two groups. $R(\tau)$ is an exact expression with components that could be integrations over interval $[0, \tau]$. Computation indicates that $[\log(\delta)]^2 \omega_1 \omega_2 \approx R(\tau)$ when the allocation is balanced (i.e., ω_1 (ω_2) is close to 0.5) and $|\log(\delta)| \gg 1$; this verifies the accuracy of the Schoenfeld formula (6) for this range of parameters in the application. The computation also indicates that $[\log(\delta)]^2 \omega_1 \omega_2 p_1(\tau) p_2(\tau) / [P(\tau)]^2 \approx R(\tau)$ for any $0 < \omega_1 < 1$ and $|\log(\delta)| \gg 2$, which leads to the following sample size formula:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 P(\tau)}{\omega_1 \omega_2 [\log(\delta)]^2 p_1(\tau) p_2(\tau)}. \quad (7)$$

This formula can be regarded as an extension of the Schoenfeld formula in the sense that equation (7) can be obtained by replacing the probability of entry assignment ω_i in (6) with the conditional probability of grouping a given failure, $\omega_i p_i(\tau)/P(\tau)$, for $i = 1, 2$. Conversely, equation (6) can be regarded as a special case of equation (7) just for the situation of the local alternative with $p_1(\tau) \approx p_2(\tau)$, in which we have $\omega_i \approx \omega_i p_i(\tau)/P(\tau)$. It is straightforward to show that formula (7) is mathematically equivalent to

$$\left(\frac{1}{n_1 p_1(\tau)} + \frac{1}{n_2 p_2(\tau)} \right)^{-1} = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{[\log(\delta)]^2}, \quad (8)$$

which can sometimes be used for application convenience as an alternative form of the formula (7) for the log-rank test statistic under the proportional hazards model. The equation (8) had been previously obtained as a sample size formula for the Wald test statistics under an exponential distribution (Rubinstein et al., 1981) or the Weibull distribution (Wu and Xiong, 2014). For the log-rank test statistic under the proportional hazards model, formula (7) is more accurate than the Schoenfeld formula (6), especially when the assignment of subjects is unbalanced for the two groups. Once the total sample size of two groups has been calculated based on formula (6) or (7), the sample size and number of events for i^{th} the group can be calculated as $n_i = n\omega_i$ and $d_i(\tau) = n_i p_i(\tau)$, $i = 1, 2$, respectively.

To calculate the sample size using formula (6) or (7), we have to calculate the probability of failure during $[0, \tau]$, $p_i(\tau) = P(T_i - \tau - Y, T_i < C)$, for a patient in group i ($i=1$ or 2), where Y is the calendar access time of enrollment, T_i is the time of event since enrollment for a patient from group i , and C is the time of lost to follow up or drop out since enrollment. Assuming Y , T_i and C are independent, $p_i(\tau)$ can be expressed as, for $i = 1, 2$,

$$p_i(\tau) = P(T_i < C \wedge (\tau - Y)) = \int_0^\infty G(t) S_i(t) d\Lambda_i(t),$$

where $S_i(t)$ and $\Lambda_i(t)$ are the survival distribution and cumulative hazard function of the i^{th} group, and $G(t) = [1 - K(t)]H(\tau - t)$ is the common distribution of overall censoring time for the two groups; $H(t)$ is the distribution of Y , and $K(t)$ is the distribution of C . If we assume that subjects are recruited with a uniform distribution $H(t)$ over the accrual period t_a and are followed for time t_f and we further assume that no subjects are lost to follow-up. Then $K(t) = 0$ for $t < \infty$ and the censoring distribution $G(t) = H(\tau - t)$ is uniform over interval $[t_f, t_a + t_f]$. Thus, for the situation of uniform entry and no lost to follow up and no drop out, the $p_i(\tau)$ in sample size formula (6) or (7) can be calculated by

$$p_i(\tau) = 1 - \frac{1}{t_a} \int_{t_f}^\tau S_i(t) dt, \quad i=1, 2, \quad (9)$$

where $S_2(t) = [S_1(t)]^{1/\delta}$ and $\tau = t_a + t_f$ (see Appendix 1). This integration can also be approximated numerically using Simpson's rule,

$$p_i(\tau) = 1 - \frac{1}{6} \{S_i(t_f) + 4S_i(0.5t_a + t_f) + S_i(t_a + t_f)\}, \quad (10)$$

where the survival distribution $S_1(t)$ can be a Kaplan-Meier curve, a parametric survival distribution, or a spline survival curve (Kooperberg and Stone, 1992) of the historical data

for the control group. According to our experiences, the approximation in (10) for most cases is practically accurate for being used in equations (6), (7), or (8).

4 Information Time

In the monitoring process of a sequential trial, the statistically appropriate measure of how far a trial has been progressed is the amount of statistical information accumulated, which is often measured by information time. Information time plays an important role in trial monitoring. It is the key information for determining rejection and acceptance regions at each interim analysis. Therefore, it is important to have an accurate estimate of the information time at each interim analysis.

Different test statistics have different information times. For the log-rank test, several information times have been proposed in the literature (Lan and Lachin, 1990). For example, the exact information time for the log-rank test is defined as the ratio of variance at calendar time t to that at the end of the trial, that is $t_V = V(t)/V(\tau)$, where $V(t)$ is given by (3). In practice, however, t_V cannot be used for early stopping, because $V(\tau)$ is unknown until the end of the study. One can replace the final variance $V(\tau)$ with its projected final variance $\tilde{V} = (z_{1-\alpha} + z_{1-\beta})^2 / [\log(\delta)]^2$ (Collett, 2003) to define the information time $\tilde{t}_V = V(t)/\tilde{V}$, which is often used in the earlier stages of trial monitoring; at a later stage, it could be that $V(t) > \tilde{V}$, in which case, this information time cannot be used either. A convenient information time that is commonly used in practice is defined as the ratio of expected (observed) number of events $d(t)$ ($d^0(t)$) up to time t in two groups to the expected (observed) total number of events $d(\tau)$ ($d^0(\tau)$) at the end of the trial in the two groups, that is, $t_d = d(t)/d(\tau)$ ($t_d^0 = d^0(t)/d^0(\tau)$), where $d(\tau)$ and $d^0(\tau)$ can be estimated from the projected total number events $(z_{1-\alpha} + z_{1-\beta})^2 / \omega_1 \omega_2 [\log(\delta)]^2$. This information time can be biased if the patient allocations between the two groups are unbalanced.

Lan and Zucker (1993) proposed an information for the log-rank test at calendar time t that is given by

$$I(t) = \left(\frac{1}{N_1(t)} + \frac{1}{N_2(t)} \right)^{-1} \frac{d(t)}{N(t)}, \quad (11)$$

where $d(t)$ is the expected total number of events in two groups at time t under the null hypothesis, and $N_i(t)$ is the sample size at time t in the i^{th} group, where $i = 1, 2$ and $N(t) = N_1(t) + N_2(t)$. Only when a subject has had an event by time t is the information of one's survival complete, and this subject can be taken as a complete sample by t . Thus, we can modify $I(t)$ in (11) by replacing $N_i(t)$ with $d_i(t)$ and $N(t)$ with $d(t) = d_1(t) + d_2(t)$. Then,

$$I(t) \simeq \left(\frac{1}{d_1(t)} + \frac{1}{d_2(t)} \right)^{-1}, \quad (12)$$

where $d_i(t)$ is the expected number of events by time t in the i^{th} group, where $i = 1, 2$. Thus, the information time $t_V = V(t)/V(\tau) \simeq I(t)/I(\tau)$ at calendar time t can be approximated by

$$t_D = D(t)/D(\tau), \quad (13)$$

where $D(t) = (d_1^{-1}(t) + d_2^{-1}(t))^{-1}$ and $D(\tau) = (d_1^{-1}(\tau) + d_2^{-1}(\tau))^{-1}$.

To calculate the information time defined by (13) under the uniform accrual and no loss to follow up, as discussed in section 3, the expected number of events of in the i^{th} group up to time t can be calculated by $d_i(t) = n\omega_i p_i(t)$, where

$$p_i(t) = (t \wedge t_a)/t_a - \int_{(t-t_a)^+}^t S_i(x) dx, \quad (14)$$

and $S_2(t) = [S_1(t)]^{1/\delta}$ (see Appendix 1).

If the k^{th} interim look is planned at the time when d_{ik} events have occurred, then the information time at the k^{th} interim look can be calculated by

$$t_k = D_k/D_\tau, \quad (15)$$

where $D_k = (d_{1k}^{-1} + d_{2k}^{-1})^{-1}$ and $D_\tau = (z_{1-\alpha} + z_{1-\beta})^2 / [\log(\delta)]^2$. Obviously, we will have more precise information by knowing the number of events in each group than by knowing the only total number of events in the two groups combined. The simulation conducted in section 6 showed that, in an unbalanced design, the proposed information time t_D given in (13) does reduce the biases and provides more accurate information time for sequential trial design.

Let n be the total sample sizes for the fixed sample test calculated based on formula (6) or (7), then the traditional information time at calendar time t ($0 < t < \tau$) can be calculated by $t_d = d(t)/d(\tau)$, where

$$d(t) = n(\omega_1 p_1(t) + \omega_2 p_2(t)), \quad (16)$$

and $d(\tau) = (z_{1-\alpha} + z_{1-\beta})^2 / \omega_1 \omega_2 [\log(\delta)]^2$ and $p_i(t)$ by (14). The proposed new information time can be calculated by $t_D = D(t)/D(\tau)$, with

$$D(t) = n \left(\frac{1}{\omega_1 p_1(t)} + \frac{1}{\omega_2 p_2(t)} \right)^{-1}, \quad (17)$$

where $D(\tau) = (z_{1-\alpha} + z_{1-\beta})^2 / [\log(\delta)]^2$ and $p_i(t)$ by (14).

5 Group Sequential Procedure

In this section, we will apply a sequential conditional probability ratio test (SCPRT) procedure (Xiong, 1995) to the log-rank test $L(t)$ by using (4). The SCPRT has two unique features: (i) the maximum sample size of the sequential test is not greater than the size of the reference fixed sample test; and (ii) the probability of discordance, or the probability that the conclusion of the sequential test would be reversed if the experiment were not stopped according to the stopping rule but continued to the planned end, can be controlled to an arbitrarily small level (Xiong et al., 2003). Furthermore, the power function of the SCPRT is virtually the same as that of the fixed sample test (Xiong, 2015). The SCPRT boundaries derived in this paper have analytical solutions. All these features make the SCPRT attractive and simple to use.

Now we apply the SCPRT to the test statistic $B_{t^*} = U(t)/V^{1/2}(\tau) \sim N(\theta t^*, t^*)$, which is a Brownian motion in information time $t^* = V(t)/V(\tau)$ over $[0, 1]$, and the drift parameter $\theta = \log(\delta) V^{1/2}(\tau)$. Note that, for convenience, we use the notation t^* to represent a general information time in this section. Therefore, the conditional density $f(B_{t^*} | B_1)$ is the normal density of $N(B_1 t^*, (1-t^*)t^*)$. Let $s_0 = z_{1-\alpha}$ be the critical value of B_1 to reject the null for the fixed sample test. Then, the conditional maximum likelihood ratio for the stochastic process on information time t^* is as follows (Xiong et al., 2003):

$$L(t^*, B_{t^*} | z_{1-\alpha}) = \frac{\max_{\{s > s_0\}} f(B_{t^*} | B_1 = s)}{\max_{\{s \leq s_0\}} f(B_{t^*} | B_1 = s)}.$$

Taking the logarithm, the log-likelihood ratio can be simplified as

$$\log(L(t^*, B_{t^*} | z_{1-\alpha})) = \pm \frac{(B_{t^*} - z_{1-\alpha} t^*)^2}{2(1-t^*)t^*},$$

which has a positive sign if $B_{t^*} > z_{1-\alpha} t^*$ and a negative sign if $B_{t^*} < z_{1-\alpha} t^*$. Suppose the k^{th} interim looks are planned at calendar time t_k , $k = 1, \dots, K$. Then, based on the SCPRT procedure presented above, the lower and upper boundaries for $B_{t_k^*}$ at the k^{th} look are given by

$$a_k = z_{1-\alpha} t_k^* - \{2at_k^*(1-t_k^*)\}^{1/2}; \quad b_k = z_{1-\alpha} t_k^* + \{2at_k^*(1-t_k^*)\}^{1/2}, \quad (18)$$

for $k = 1, \dots, K$, where $t_k^* = V(t_k)/V(\tau)$ is the information time at the k^{th} look at calendar time t_k . The a in (18) is the boundary coefficient, and it is crucial to choose an appropriate a for the design such that the probability of the conclusion from the sequential test being reversed by the test at the planned end is small but not unnecessarily so. The larger the value of a , the smaller the discordance probability and the further apart the upper and lower boundaries, making it harder for the sample path to reach boundaries and stop early and resulting in larger expected sample sizes. Thus, an appropriate a can be determined by

choosing an appropriate discordance probability (Xiong et al., 2003). The nominal critical p -values for testing H_0 are

$$P_{a_k} = 1 - \Phi(a_k / \sqrt{t_k^*}); \quad P_{b_k} = 1 - \Phi(b_k / \sqrt{t_k^*}). \quad (19)$$

The observed p -value at the k^{th} look is

$$P_{B_{t_k^*}} = 1 - \Phi(B_{t_k^*} / \sqrt{t_k^*}).$$

The stopping rule for monitoring the trial can be executed by stopping the trial when, for the first time, $P_{B_{t_k^*}} \geq P_{a_k}$ (accept H_0 and stop for futility) or $P_{B_{t_k^*}} \leq P_{b_k}$ (reject H_0 and stop for efficacy). Since $L(t_k)$ has the same asymptotic distribution under the null hypothesis as $B_{t_k^*} / \sqrt{t_k^*}$, the lower and upper boundaries for $L(t_k)$ at the k^{th} look are given by

$$\bar{a}_k = a_k / \sqrt{t_k^*} \quad \text{and} \quad \bar{b}_k = b_k / \sqrt{t_k^*}, \quad (20)$$

respectively, and the observed p -value at the k^{th} stage can be calculated from the test statistic $L(t_k)$ by applying all observations up to stage k . In practice, if the distribution under the null is estimated from historical data by a parametric survival distribution or a spline version of the survival distribution, then one can use the information time $t_D = D(t)/D(\tau)$ to replace the exact but unknown information time $t_V = V(t)/V(\tau)$, or one can calculate the information time at the k^{th} interim look by $t_D^* = D_k/D_\tau$, as defined in equation (15) for a maximum information trial where the trial is stopped after a planned number of events are observed in two groups.

6 Simulation Studies

Simulations were conducted to study the accuracies of the Schoenfeld formula (6) and the new formula (7) for the log-rank test statistic under the proportional hazards model. The Weibull distribution $S(t) = e^{-\lambda t^\kappa}$ was used to simulate the time-to-event data for both treatment groups with the same shape parameter κ to satisfy the proportional hazards model assumption, λ is the hazard parameter. The Weibull shape parameter κ was taken as 0.5, 1, and 2 to reflect cases of decreasing, constant, and increasing hazard function. The hazard parameter for the control group was taken as $\lambda = 0.5$, and the hazard ratio δ was set to 1.3, 1.5, 1.8, and 2. Furthermore, we assumed that patients were recruited with a uniform distribution over the accrual period $t_a = 3$ (years) and followed for $t_f = 1$ (years), so that the study period $\tau = t_a + t_f = 4$ (years), and that no subject was lost to follow-up during the study. The total sample sizes required for the study were calculated by formulae (6) and (7) under the proportion of allocation to control group $\omega_1 = 0.3, 0.5$ and 0.7 , with nominal type I error of 0.05 for a one-sided test and power of 90%. Simulations were also performed for the log-logistic distribution $S(t) = 1/(1 + \lambda t^\rho)$, where $\lambda = 0.5$ for the control group, and the

shape parameter $p = 0.5, 1$ and 2 . Other design parameters were the same as the simulation for the Weibull model above. In each design parameter configuration, 10,000 trials were generated under the Weibull or log-logistic distribution to calculate the empirical power (Tables 1–2). The results showed that the Schoenfeld formula (6) overestimate the sample size when $\omega_1 = 0.3$ and underestimate the sample size when $\omega_1 = 0.7$. The formula (7) estimated the sample size accurately when $\omega_1 = 0.3$ and slightly overestimated when $\omega_1 = 0.7$. For a balanced design ($\omega_1 = 0.5$), the formula (7) is almost identical to the Schoenfeld formula (6), and both are very accurate.

We also conducted a simulation study to evaluate the performance of the proposed new information time t_D from Section 4. The exact information time $t_V = V(t)/V(\tau)$ is a random variable and its sample mean t_V was obtained by calculating the variances of the log-rank test at calendar time t and at the end of the study $t = \tau$ based on 10,000 simulated trials. In the simulations, the shape parameter k (or p) was set to $0.5, 1$ and 2 for the Weibull (log-logistic) model and the scale parameter λ was set to $1/\log(2)$ and 0.5 for the Weibull and log-logistic model, respectively. Assume a uniform accrual with an accrual period $t_a = 5$ (years), and a follow-up duration $t_f = 2$ (years). The proposed new information time at calendar time t can be calculated by t_D from formula (13) with $p_\lambda(t)$ in (14). We compared the new and exact information times with the commonly used information time t_d (Tables 3–4). The results showed that the proposed information time t_D was closer than t_d to the mean of exact information time t_V , and t_d is biased either downward or upward, depending on the sample size allocation ratio.

Overall, sample size formula (7) and the new information time t_D performed better than the commonly used methods for designing trials with unbalanced treatment allocation.

7 Example

The proposed new calculations of sample size and information time are applicable and beneficial to any group sequential designs. To illustrate new approaches with a randomized phase III survival trial design with unbalanced treatment allocation, we compare it with the commonly used methods implemented using a commercially available software EAST 5 (Cytel Software Corporation, 2007) which implement the traditional calculation of sample size and information time. For this example, we assume that the overall survival is exponentially distributed for both the control and treatment groups. Assume the 3-year survival probability of the control group is 50% and that the investigators expect the new treatment may improve the 3-year survival probability to 60%. We also assume a uniform accrual with accrual period $t_a = 5$ (years), follow-up time $t_f = 3$ (years), and study duration $\tau = 8$ (years), that no subjects are lost to follow-up, and that 30% ($\omega_1 = 0.3$) of patients are randomized to the control group and 70% ($\omega_2 = 0.7$) of patients to the treatment group. Then, the trial can be designed to detect a hazard ratio $\delta = \log(0.5)/\log(0.6) = 1.357$ or $1/\delta = 0.737$ as a two-sided test for the hypothesis (2) with type I error of 0.05 and power of 90%. Thus, the total number of deaths and the sample size calculated for a fixed sample test using the Schoenfeld formula (the same as calculated in EAST 5) are $d(\tau) = 438$ and $n = 697$, respectively; and using the sample size formula (7), those are $d(\tau) = 413$ and $n = 656$, respectively, calculated using the R function SIZEinfTime (Appendix 3). Now suppose that

the trial will be monitored for both futility (accept H_0) and efficacy (reject H_0) based on a 7-stage group sequential design at calendar time $t = 2, 3, 4, 5, 6, 7$ and 8 years. Then, the traditional information time is $t_d = (0.106, 0.225, 0.378, 0.559, 0.735, 0.880, 1.000)$. Inputting this information time into EAST 5, the total number of deaths and sample size required for this 7-stage group sequential design increase to 499 and 794, respectively, by using O'Brien-Fleming boundaries; and to 664 and 1057, respectively, by using Pocock boundaries. The new information time is $t_D = (0.109, 0.231, 0.386, 0.568, 0.743, 0.885, 1.000)$, and the required number of deaths and sample size for the SCPRT are the same as for the fixed sample test. The symmetric Pocock and O'Brien-Fleming boundaries and boundary-crossing probabilities calculated in EAST 5 using Lan-DeMets error spending functions are given in Table 5. On using the SCPRT, implemented in a user-friendly software SCPRTinfWin (Xiong, 2007), a maximum conditional probability of discordance $\rho = 0.02$ is given, with maximum probability of discordance $\rho_{max} = 0.0056$, to calculate the boundary coefficient $a = 3.496$ for a 7-stage group sequential design with an unequally spaced information time (Xiong et al., 2003). After inputting the number of interim looks $K = 7$ and the information time t_D at each stage in SCPRTinfWin, the symmetric group sequential futility and efficacy boundaries and boundary crossing probabilities calculated by SCPRTinfWin are obtained (Table 5). The actual significance level and power for this 7-stage group sequential design based on the SCPRT are 0.051 and 0.899, respectively, which are very close to the nominal levels of type I error of 5% and power of 90%. The expected sample size (Appendix 2), study duration, and number of events under the alternative H_1 calculated from EAST 5 and SCPRTinfWin are recorded in Table 5. The operating characteristics of the three designs can be summarized as follows. First, the SCPRT requires the smallest maximum sample size and the maximum number of deaths, which is the same for the fixed sample test. Second, the SCPRT also requires the smallest expected sample size. However, Pocock's method requires the smallest expected study duration and number of deaths, and the O'Brien-Fleming method requires the expected study duration to be larger than for Pocock's method but smaller than for the SCPRT. This is because Pocock's method has the highest stopping probabilities at early stages, the O'Brien-Fleming method has the highest stopping probabilities in the middle stages, and the SCPRT has the highest stopping probabilities in the later stages of the trial.

For a group sequential design, the spending function specifies the relation of the cumulative type I errors and the information times at interim and final looks. For some group sequential methods such as O'Brien-Fleming or Pocock's methods, the spending function is a simple analytical function and with which one may determine easily information times for interim looks according to desired type I errors to be spent at those interim looks. The group sequential boundaries of the design are then calculated according to these type I errors and information times. The case is different for the SCPRT that the form of its boundaries is an analytical form but its spending function cannot be obtained as a simple analytical function. However, the values of its spending function can be calculated as the cumulative probabilities of rejecting H_1 at the information times of interim looks under the setting of H_0 using SCPRTinfWin. For examples, for the design with $\alpha = 0.05$, the spending function equals 0.001, 0.0017, 0.0021, 0.0026, 0.003, 0.0036, 0.0045, 0.0081, 0.01, or 0.051, respectively, when the information time equals 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.

For the design with $\alpha = 0.025$, the spending function equals 0.0007, 0.001, 0.0013, 0.0016, 0.0018, 0.0021, 0.0026, 0.0038, 0.0051, or 0.026, respectively, when the information time equals 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1. With these values of spending function for the SCPRT, one may determine information times for desired type I errors for interim looks in a design, or may control the type I errors at interim looks during a test.

8 Conclusion

For randomized survival trials with unequal sample size allocation, the commonly used Schoenfeld formula is inaccurate and tends to either underestimate or overestimate the sample size, depending on the sample size allocation ratio of the two groups. The formula (7) provides more accurate sample size estimation. Furthermore, the widely used information time tends to be biased either upward or downward. The proposed new information time is more accurate, particularly for an unbalanced allocation. An R function, `SIZEinfTime`, is given for the sample size and information time calculations using the proposed methods. Overall, the proposed method gives accurate sample size and information time estimations. For the group sequential design based on the SCPRT, one can input the information time calculated by the R function into `SCPRTinfWin` to calculate the sequential boundaries for the purpose of trial monitoring. The SCPRT design reduces the number of patients required to be enrolled on the study compared to methods implemented in EAST 5. Furthermore, the SCPRT has two advantages compared to other sequential methods. First the maximum sample size of the SCPRT is the same as for the fixed sample test, whereas other methods require the sample size to be increased. Second, the probability of discordance, can be controlled to an arbitrarily small level, whereas it cannot be prespecified in the design stage when using other sequential procedures.

Acknowledgments

This work was supported in part by National Cancer Institute (NCI) support grant P30CA021765 and ALSAC.

Appendix 1

Derivation of the probability of having an event before calendar time t

Assume subjects are accrued over an accrual period of length t_a , with an additional follow-up time t_f so that the study duration $\tau = t_a + t_f$ and that the entry time Y is uniformly distributed with $H(t)$ over $[0, t_a]$. With no patient loss to follow-up or drop out, the censoring distribution $G(t) = H(\tau - t)$ is a uniform distribution over interval $[t_f, t_a + t_f]$, that is, $G(x) = 1$ if $x - t_f = (t_a + t_f - x)/t_a$ if $t_f - x - t_a + t_f = 0$ otherwise. Let T be the event time with survival distribution $S_i(x)$ under the alternative for $i = 1, 2$. The probability of a subject in the i^{th} group having an event before calendar time t can be calculated by

$$\begin{aligned} p_i(t) &= P(T_i + Y \leq t) = \int_0^\infty P(T_i + Y \leq t | Y = x) dH(x) \\ &= \int_0^\infty P(T_i \leq t - x) dH(x) = \frac{1}{t_a} \int_0^{t \wedge t_a} \{1 - S_i(t - x)\} dx \\ &= (t \wedge t_a) / t_a - \frac{1}{t_a} \int_{(t - t_a)^+}^t S_i(x) dx. \end{aligned}$$

By letting $t = \tau$, we have the equation (9).

Appendix 2

Formulae for calculating expected sample size, number of events, and study time for sequential tests with information time for survival data

For a group sequential test with information time for survival data, let $\hat{s}$ be the stopping time for the calendar time t over $[0, \tau]$, and let $\hat{s}^*$ be the corresponding stopping time for the corresponding information time t^* over $[0, 1]$. Let $N_a(t)$ be the number of patients enrolled by calendar time t , then for a study with uniform enrollment over $[0, t_a]$, $N_a(t) = (t/t_a)N_{max}$ for $0 < t < t_a$ and $N_a(t) = N_{max}$ for $t_a \leq t \leq \tau$, where N_{max} is the maximum sample size for the group sequential trial. For a group sequential test with K looks, assume $t_1, \dots, t_K$ are calendar times and $t_1^*, \dots, t_K^*$ are the corresponding information times. Assume the probability that the sequential test statistic W_t on information time stops at time t_k^* is denoted as $P_{t_k^*} = P(\hat{s}^* = t_k^*)$, which is usually available from the distribution of W_t . Under the alternative hypothesis, the expected sample size is $E[N_a(\hat{s})] = \sum_{k=1}^K N_a(t_k) P_{t_k^*}$, and the expected study duration is $E(\hat{s}) = \sum_{k=1}^K t_k P_{t_k^*}$. For the traditional method, the expected number of events is $E[d(\hat{s})] = d_{max} E_{t^*}$ with $E_{t^*} = \sum_{k=1}^K t_k^* P_{t_k^*}$, where d_{max} is the maximum number of events for the group sequential design. For the new method, the expected number of events is $E[d(\hat{s})] = \sum_{k=1}^K d_a(t_k) P_{t_k^*}$ with $d_a(t_k) = N_a(t_k) P(t_k)$, where $P(t_k) = \omega_1 p_1(t_k) + \omega_2 p_2(t_k)$ with $p(t_k)$ by (14).

Appendix 3

R code for sample size and information time calculation

```
SIZEinfTime=function(s1, s2, x, pi, ta, tf, alpha, beta, t)
{
  ### s1 and s2 are the survival probabilities at x for groups 1 and 2 ####
  ### pi is the proportion of patients assign to group 1 #####
  ### ta and tf are the accrual time and follow-up time #####
  ### alpha and beta are the type I and II error, study power=1-beta #####
  ### for a two-sided test, input half of alpha as alpha in this function ###
  ### t is the calendar time at which to calculate the information time####
  ### one can calculate the sample size and information time for any #####
  ### distribution by change S1 and S2 to the corresponding distributions##
  z0=qnorm(1-alpha); z1=qnorm(1-beta)
  lambda1=-log(s1)/x; HR=log(s1)/log(s2)
  S1=function(t){ans=exp(-lambda1*t); return(ans)}
  S2=function(t){ans=S1(t)^(1/HR);return(ans)}
  dSC=ceiling((z0+z1)^2/(pi*(1-pi)*(log(HR))^2))##expected total number of
```

```

events#
p1=1-integrate(S1, tf, ta+tf)$value/ta
p2=1-integrate(S2, tf, ta+tf)$value/ta
P0=pi*p1+(1-pi)*p2
NSC=ceiling((z0+z1)^2/(pi*(1-pi)*(log(HR))^2*P0))##Schoenfeld formula (6)#
NXW=ceiling((z0+z1)^2*P0/(pi*(1-pi)*(log(HR))^2*p1*p2))##New formula (7)##
dXW=ceiling(NXW*P0)
a=t-min(t,ta); b=min(t,ta)
pt1=b/ta-integrate(S1, a, t)$value/ta
pt2=b/ta-integrate(S2, a, t)$value/ta
dt=NSC*(pi*pt1+(1-pi)*pt2)
dtau=(z0+z1)^2/(pi*(1-pi)*log(HR)^2)
Dt=NXW*(1/(pi*pt1)+1/((1-pi)*pt2))^(−1)
Dtau=(z0+z1)^2/log(HR)^2
td=round(dt/dtau,3) ### information time td based on traditional method #
tD=round(Dt/Dtau,3) ### information time tD based on new method #
ans=data.frame(dSC=dSC, NSC=NSC, dXW=dXW, NXW=NXW, td=td, tD=tD)
return(ans)
}
SIZEinfTime(s1=0.5, s2=0.6, x=3, pi=0.3, ta=5, tf=3, alpha=0.05, beta=0.1,
t=2)
dSC NSC dXW NXW      td      tD
438 697 413 656 0.106 0.109

```

References

- Barthel FMS, Babiker AG, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for nonuniform accrual, non-proportional hazards, loss to follow-up and crossover. *Statistics in Medicine*. 2006; 25:2521–2542. [PubMed: 16479563]
- Collett, D. Modeling survival data in medical research. 2. London: Chapman and Hall; 2003.
- Cytel Software Corporation. EAST version 5.4. Cytel; Cambridge MA: 2007.
- Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*. 1982; 1:121–129. [PubMed: 7187087]
- Halabi S, Singh B. Sample size determination for comparing several survival curves with unequal allocations. *Statistics in Medicine*. 2004; 23:1793–1815. [PubMed: 15160409]
- Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *British Journal of Radiology*. 1971; 44:793–797. [PubMed: 4940475]
- Hsieh FY. Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Statistics in Medicine*. 1992; 11:1091–1098. [PubMed: 1496196]
- Jennison, C., Turnbull, BW. Group sequential methods with applications to clinical trials. New York: Chapman and Hall; 2000.
- Kooperberg C, Stone CJ. Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*. 1992; 1:301–328.
- 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. *Biometrics*. 1986; 42:507–519. [PubMed: 3567285]
- Lakatos E, Lan KKG. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine*. 1992; 11:179191.

- Lakatos E. Designing complex group sequential survival trials. *Statistics in Medicine*. 2002; 21:1969–1989. [PubMed: 12111882]
- Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983; 70:659–663.
- Lan KKG, Lachin JM. Implementation of group sequential logrank tests in a maximum duration trial. *Biometrics*. 1990; 46:759–770. [PubMed: 2242413]
- Lan KKG, Zucker DM. Sequential monitoring of clinical trials: the role of information and Brownian motion. *Statistics in Medicine*. 1993; 12:753–765. [PubMed: 8516592]
- Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977; 64:191–199.
- O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35:549–556. [PubMed: 497341]
- Rubinstein LV, Gail MH, Santner TJ. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *Journal of Chronic Diseases*. 1981; 34:469–479. [PubMed: 7276137]
- Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981; 68:316–319.
- Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983; 39:499–503. [PubMed: 6354290]
- Sellke T, Siegmund D. Sequential analysis of the proportional hazards model. *Biometrika*. 1983; 79:315–326.
- Slud EV. Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics*. 1984; 12:551–571.
- 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.
- Whitehead J, Stratton I. Group sequential clinical trial with triangular continuation regions. *Biometrics*. 1983; 39:227–236. [PubMed: 6871351]
- Wu J, Xiong X. Group sequential design for randomized phase III trials under the Weibull model. *Journal of Biopharmaceutical Statistics*. 2014; doi: 10.1080/10543406.2014.971165
- Xiong X. A class of sequential conditional probability ratio tests. *Journal of American Statistical Association*. 1995; 15:1463–1473.
- Xiong, X. A computer program for SCPRT on information time, version 1.0. 2007. <http://www.stjuderesearch.org/site/depts/biostats/scprt>
- Xiong X. A precise approach for sequential test design on comparing survival distributions by log-rank test. 2014 *Un-published Manuscript*.
- Xiong X, Tan M, Boyett J. Sequential conditional probability ratio tests for normalized test statistic on information time. *Biometrics*. 2003; 59:624–631. [PubMed: 14601763]

Table 1

Sample sizes (n), empirical type I error ($\hat{\alpha}$) and power ($1 - \hat{\beta}$) for the Schoenfeld formula (6) and the new formula (7) under the Weibull (WB) model with scale parameter $\lambda = 0.5$, where sample sizes were calculated with a two-sided type I error of 5% and power of 90%; and the corresponding type I errors and empirical powers were estimated by 10,000 simulation runs.

WB	$\kappa = 0.5$					$\kappa = 1$					$\kappa = 2$				
	Formula	ω_1	δ	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$
(6)	0.5	1.3	1.243	953	.051	.901	953	.050	.902	725	.049	.902	1.5	546	.898
				415	.048	.898	415	.050	.894	309	.053	.898			
		1.8	276	208	.050	.895	208	.056	.898	151	.053	.897	2.0	205	.896
				154	.051	.898	154	.049	.897	111	.053	.896			
	0.3	1.3	1535	1167	.048	.918	1167	.053	.917	873	.051	.906	1.5	689	.905
				517	.052	.920	517	.051	.918	375	.048	.905			
		1.8	358	266	.050	.925	266	.052	.924	186	.054	.912	2.0	271	.915
				200	.048	.932	200	.051	.927	137	.055	.915			
	0.7	1.3	1428	1104	.051	.886	1104	.050	.888	853	.054	.892	1.5	614	.883
				473	.053	.875	473	.052	.881	361	.052	.883			
(7)	0.5	1.3	1253	958	.048	.904	958	.051	.902	725	.049	.902	1.5	557	.898
				421	.050	.904	421	.051	.900	710	.053	.898			
		1.8	288	214	.052	.904	214	.052	.904	152	.053	.897	2.0	219	.896
				161	.050	.915	161	.048	.907	112	.053	.896			
	0.3	1.3	1438	1108	.047	.896	1108	.050	.904	853	.048	.898	1.5	625	.899
				478	.050	.889	478	.051	.897	362	.053	.899			
		1.8	314	238	.050	.893	238	.052	.899	176	.057	.890	2.0	234	.899
				176	.049	.900	176	.054	.891	128	.053	.899			
	0.7	1.3	1546	1172	.053	.908	1172	.048	.904	873	.049	.898	1.5	701	.896
				523	.051	.916	523	.050	.909	376	.048	.896			
(8)	1.8	372	.050	272	.049	.923	186	.052	.898	2.0	286	.050	.931	207	.052
				207	.052	.923	138	.052	.901						

Table 2

Sample sizes (n), empirical type I error ($\hat{\alpha}$) and power ($1 - \hat{\beta}$) for the Schoenfeld formula (6) and the new formula (7) under the log-logistic (LG) model with scale parameter $\lambda = 0.5$, where sample sizes were calculated with a two-sided type I error of 5% and power of 90%; and the corresponding type I errors and empirical powers were estimated by 10,000 simulation runs.

LG	$p = 0.5$					$p = 1$					$p = 2$				
	Formula	ω_1	δ	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$	n	$\hat{\alpha}$	$1 - \hat{\beta}$
(6)	0.5	1.3	1.549	.049	.902	.1237	.048	.906	.917	.047	.906				
			1.5	.682	.054	.891	.543	.051	.896	.399	.052	.898			
		1.8	345	.047	.892	.274	.053	.894	.199	.049	.898				
			2.0	.257	.052	.888	.204	.048	.896	.148	.052	.896			
	0.3	1.3	1921	.050	.918	.1527	.048	.910	.1120	.048	.911				
			1.5	.867	.048	.922	.685	.050	.921	.495	.053	.916			
		1.8	453	.051	.928	.356	.049	.928	.253	.052	.922				
			2.0	.343	.051	.929	.269	.054	.932	.190	.054	.924			
	0.7	1.3	1773	.052	.880	.1422	.052	.887	.1063	.051	.890				
		1.5	764	.051	.866	.612	.052	.878	.455	.048	.884				
(7)	0.5	1.3	377	.051	.851	.301	.050	.864	.223	.049	.870				
			2.0	.276	.049	.848	.220	.055	.858	.163	.050	.865			
	1.5	1565	.051	.907	.1247	.048	.905	.921	.047	.900					
		700	.052	.899	.554	.048	.906	.403	.052	.902					
	1.8	365	.051	.910	.286	.053	.909	.205	.052	.905					
		2.0	.278	.051	.917	.217	.051	.913	.153	.050	.909				
	0.3	1.3	1788	.053	.897	.1431	.050	.894	.1067	.047	.900				
			1.5	.780	.052	.892	.622	.050	.894	.459	.048	.896			
		1.8	394	.054	.890	.312	.054	.894	.228	.054	.893				
			2.0	.294	.055	.889	.233	.049	.894	.168	.053	.897			
(7)	0.7	1.3	1937	.048	.911	.1537	.048	.905	.1124	.050	.902				
		1.5	.885	.049	.919	.696	.050	.916	.500	.053	.908				
		1.8	474	.049	.929	.369	.048	.927	.259	.050	.919				
		2.0	.366	.049	.935	.284	.050	.931	.197	.046	.927				

Table 3

The exact information time t_V was calculated as the mean (se) based on 10,000 simulated samples and the traditional information time t_I and new information time t_D were calculated under the Weibull (WB) model

WB	$\omega_1=0.2$			$\omega_1=0.3$			$\omega_1=0.7$			$\omega_1=0.8$		
	κ	t	t_V (se)	t_D	t_d	t_V (se)	t_D	t_d	t_V (se)	t_D	t_d	t_d
0.5	1	1	.093(.024)	.093	.087	.092(.023)	.091	.088	.087(.018)	.086	.094	
	2	2	.242(.038)	.240	.230	.240(.035)	.238	.232	.232(.028)	.228	.242	
	3	3	.417(.044)	.414	.400	.415(.040)	.411	.403	.404(.032)	.397	.416	
	4	4	.608(.042)	.604	.588	.606(.039)	.601	.592	.594(.032)	.586	.607	
	5	5	.812(.029)	.808	.792	.810(.029)	.804	.795	.798(.025)	.789	.810	
	6	6	.930(.016)	.927	.919	.929(.017)	.926	.921	.923(.016)	.917	.929	
1	1	1	.058(.020)	.057	.050	.057(.019)	.055	.051	.052(.015)	.049	.058	
	2	2	.193(.039)	.189	.172	.191(.035)	.185	.175	.179(.026)	.170	.191	
	3	3	.368(.050)	.362	.336	.365(.045)	.357	.341	.349(.033)	.334	.365	
	4	4	.564(.053)	.557	.527	.561(.047)	.551	.532	.543(.034)	.524	.559	
	5	5	.770(.043)	.763	.733	.768(.039)	.757	.739	.750(.029)	.731	.765	
	6	6	.923(.024)	.918	.900	.921(.023)	.914	.903	.911(.018)	.898	.919	
2	1	1	.036(.016)	.035	.030	.035(.015)	.034	.031	.031(.011)	.029	.036	
	2	2	.187(.044)	.183	.163	.184(.038)	.179	.167	.171(.027)	.162	.185	
	3	3	.383(.062)	.379	.354	.379(.053)	.374	.358	.364(.036)	.353	.380	
	4	4	.584(.067)	.580	.556	.581(.058)	.576	.560	.566(.037)	.555	.581	
	5	5	.787(.057)	.782	.758	.784(.050)	.778	.762	.770(.031)	.758	.782	
	6	6	.951(.030)	.947	.931	.950(.027)	.944	.934	.941(.017)	.931	.947	

Table 4

The exact information time t_V was calculated as the mean (se) based on 10,000 simulated samples and the traditional information time t_d and new information time t_D were calculated under the log-logistic (LG) model

LG	p	$\omega_1=0.2$			$\omega_1=0.3$			$\omega_1=0.7$			$\omega_1=0.8$		
		t	t_V (se)	t_D	t_d	t_V (se)	t_D	t_d	t_V (se)	t_D	t_d	t_V (se)	t_D
0.5	1		.095(.020)	.095	.092	.094(.020)	.094	.093	.092(.016)	.092	.095	.091(.016)	.091
	2		.243(.030)	.242	.237	.242(.029)	.241	.238	.238(.024)	.237	.242	.237(.023)	.236
	3		.416(.034)	.415	.408	.415(.033)	.413	.410	.409(.028)	.408	.415	.408(.027)	.407
	4		.607(.033)	.605	.598	.606(.032)	.603	.599	.599(.028)	.597	.605	.597(.026)	.596
	5		.809(.023)	.808	.800	.809(.025)	.806	.802	.802(.022)	.800	.808	.801(.020)	.799
	6		.927(.014)	.926	.922	.927(.016)	.925	.923	.923(.014)	.922	.926	.922(.013)	.921
1	1		.055(.015)	.054	.051	.054(.015)	.054	.052	.051(.012)	.051	.054	.051(.011)	.051
	2		.179(.028)	.178	.170	.178(.026)	.176	.172	.172(.022)	.170	.178	.170(.021)	.168
	3		.342(.034)	.340	.328	.341(.033)	.337	.331	.331(.027)	.328	.340	.329(.027)	.326
	4		.530(.036)	.527	.514	.528(.035)	.524	.517	.517(.029)	.514	.527	.515(.028)	.512
	5		.735(.029)	.732	.719	.733(.029)	.729	.722	.723(.025)	.719	.732	.720(.024)	.716
	6		.898(.018)	.895	.887	.896(.019)	.894	.889	.890(.017)	.887	.895	.888(.016)	.886
2	1		.028(.011)	.028	.025	.028(.011)	.027	.025	.025(.009)	.025	.028	.025(.008)	.024
	2		.144(.029)	.141	.130	.142(.026)	.139	.132	.133(.020)	.130	.141	.130(.021)	.128
	3		.314(.041)	.309	.290	.311(.037)	.305	.294	.297(.028)	.291	.308	.293(.030)	.288
	4		.508(.045)	.502	.480	.505(.040)	.498	.485	.490(.031)	.482	.501	.485(.033)	.478
	5		.715(.039)	.709	.688	.711(.036)	.705	.692	.697(.027)	.689	.707	.692(.029)	.686
	6		.899(.022)	.895	.880	.897(.022)	.892	.883	.887(.018)	.882	.894	.884(.018)	.879

Table 5

Operating characteristics of a 7-stage group sequential design based on Pocock and O'Brien-Fleming boundaries using the Lan-DeMets error spending function implemented in EAST and SCPRT methods for the example in section 7.

Pocock				O'Brien-Fleming				SCPRT			
Boundaries		Stopping prob.		Boundaries		Stopping prob.		Boundaries		Stopping prob.	
t	Reject	Under	Under	Reject	Reject	Reject	Reject	Reject	Under	Under	Under
	H_0	H_1	H_0	H_1	H_0	H_1	H_0	H_1	H_0	H_1	H_1
2	2.393	-0.953	0.179	0.128	5.907	-3.901	0.000	0.000	1.003	-0.645	0.027
3	2.316	-0.313	0.247	0.203	3.970	-1.795	0.036	0.007	1.494	-0.735	0.050
4	2.219	0.315	0.261	0.255	2.984	-0.522	0.267	0.144	1.922	-0.652	0.096
5	2.150	0.889	0.185	0.217	2.396	0.384	0.362	0.356	2.244	-0.376	0.169
6	2.124	1.351	0.087	0.122	2.064	0.997	0.208	0.287	2.378	0.067	0.218
7	2.125	1.716	0.032	0.053	1.887	1.400	0.090	0.143	2.299	0.612	0.204
8	2.129	2.129	0.010	0.021	1.777	1.777	0.037	0.062	1.645	1.645	0.236

N_{max}	D_{max}	d_{max}	N_{max}	D_{max}	d_{max}	N_{max}	D_{max}	d_{max}
1057	8.0	664	794	8.0	499	656	8.0	413
$E[N_d(\hat{s})]$	$E(\hat{s})$	$E[d(\hat{s})]$	$E[N_d(\hat{s})]$	$E(\hat{s})$	$E[d(\hat{s})]$	$E[N_d(\hat{s})]$	$E(\hat{s})$	$E[d(\hat{s})]$
836	4.25	289	769	5.61	326	628	6.4	323

N_{max} , D_{max} and d_{max} are the maximum sample size, study duration, and number of deaths for 7-stage group sequential designs. $E[N_d(\hat{s})]$, $E(\hat{s})$ and $E[d(\hat{s})]$ are the expected sample size, study duration, and number of deaths under H_1 for the sequential designs. t is the calendar time for each interim look.

Abbreviations: Prob., probability.