Programs for Computing Group Sequential Boundaries Using the Lan-DeMets Method

Version 2

by

David M. Reboussin

Department of Public Health Sciences

Bowman Gray School of Medicine

Winston-Salem, NC 27157

David L. DeMets

Departments of Statistics and Biostatistics

University of Wisconsin

Madison, WI 53706

KyungMann Kim

Department of Biostatistics

Harvard School of Public Health and Dana-Farber Cancer Institute

Boston, MA 02115

K. K. Gordon Lan

Department of Statistics/Computer and Information Systems

George Washington University

Washington, DC 20052

July 11, 1996

Abstract

FORTRAN programs for the computation of boundaries and exit probabilities in the sequential analysis of clinical trials are described. The computations are appropriate for any trial based on normally distributed test statistics with independent increments, including those in which patients give a single continuous or binary response, survival studies, and certain longitudinal designs. Interim analyses need not be equally spaced, and their number need not be specified in advance. In addition to boundaries, power computations, probabilities associated with a given set of boundaries, and confidence intervals can be computed. An explanation of the input and output, along with some verbatim transcripts of interactive sessions are included. Some theoretical background and description of the numerical computations is provided in the appendix. The program and this documentation are updates of earlier versions.

KEY WORDS: Group sequential testing; Statistical computing; Type I error spending function; Use function.

1 Introduction

The design of many clinical trials includes some strategy for early stopping if an interim analysis reveals large differences between treatment groups. In addition to saving time and resources, such a design feature can reduce study participants' exposure to the inferior treatment. However, when repeated significance testing on accumulating data is done, some adjustment of the usual hypothesis testing procedure must be made to maintain an overall significance level (Armitage, McPherson & Rowe, 1969; McPherson & Armitage, 1971). The methods described by Pocock (1977) and O'Brien & Fleming (1979), among others, are popular implementations of group sequential testing for clinical trials. Sometimes interim analyses are equally spaced in terms of calendar time or the information available from the data, but this assumption can be relaxed to allow for unplanned or unequally spaced analyses. Lan & DeMets (1983) introduced type I error spending functions, denoted $\alpha^\star$, and determined boundaries by

$$\Pr(Z_1 \geq b_1 \text{ or } \cdots \text{ or } Z_k \geq b_k) = \alpha^\star(\tau) \tag{1}$$

where $b_1, \ldots, b_k$, are (upper) boundaries for the sequence of interim test statistics and τ is either the proportion of elapsed time to maximum duration or observed information to total information. That is, if the interim standardized test statistic at the k^{th} interim analysis is denoted by Z_k, we continue the trial as long as $|Z_k| < b_k$ (two-sided), otherwise termination is considered. The spending function $\alpha^\star(\tau) = 0$ for $\tau = 0$ and $\alpha^\star(\tau) = \alpha$ for $\tau = 1$. That is,

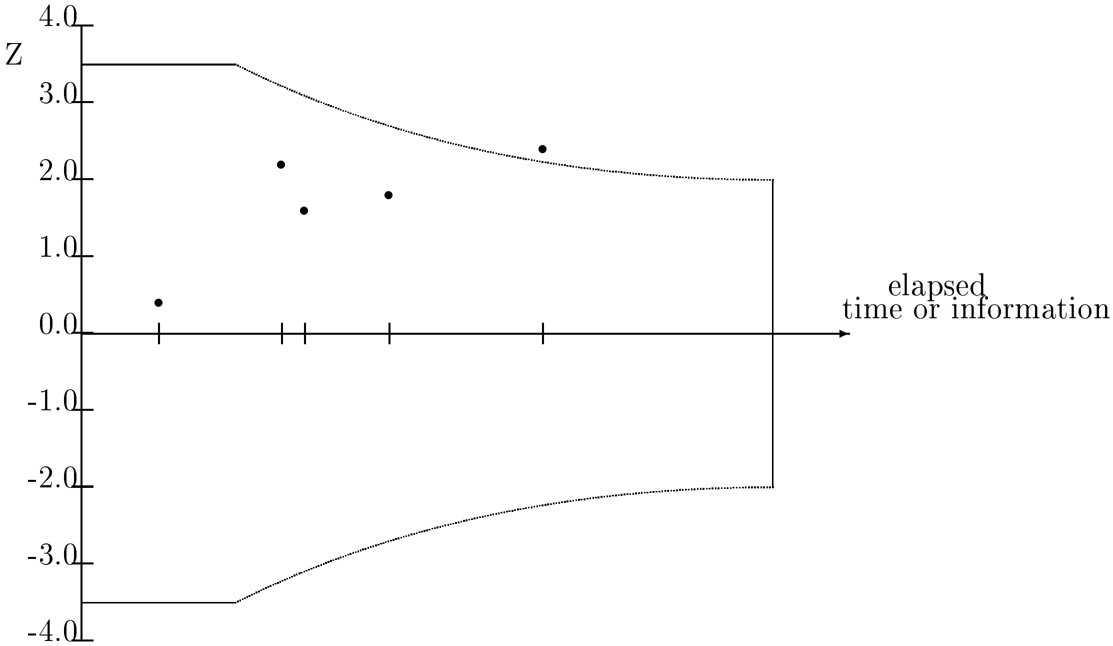

Figure 1: Sequential outcomes and boundaries for interim standardized test statistics from a clinical trial.

this flexible procedure guarantees a fixed α level when the trial is complete. Neither the time or the number of analyses needs to be specified in advance: only $\alpha^{\star}(\tau)$ must be specified. Issues surrounding the use of calendar time and information have been discussed by Lan & DeMets (1989) and Lan, Reboussin & DeMets (1994). Spending functions, which are also called use functions, are prespecified and correspond to those described by Lan & DeMets (1983) and Kim & DeMets (1987a). These are similar to commonly used group sequential boundaries proposed by Pocock (1977) and O'Brien & Fleming (1979). Additional spending functions may be found in Hwang, Shih & de Cani (1990).

Options

The program described here perform computations related to group sequential boundaries, such as the one illustrated in Figure 1. The program begins by prompting the user to specify whether it is being run interactively or not, and then to specify one of four options. It continues prompting based on the selected option. The options are:

- computation of boundaries for a specified spending function (including graphical presentation);

- power calculation for a specified set of boundary values and a drift parameter corresponding to the alternative hypothesis;
- computation of the exit probabilities for a specified spending function, analysis times, and drift parameter;
- computation of confidence intervals following termination of a trial.

For interim analysis of an ongoing clinical trial, the first option takes as input the times of the previous and current interim analyses, and the type I error spending function. The program then reports what boundaries should be used to determine whether or not to stop the trial. This is accomplished using Equation (1) and a searching routine which makes an initial choice of boundaries, computes stopping probabilities, and alters boundaries until the desired alpha level is obtained. The other options, in contrast, evaluate probabilities associated with a given set of boundaries. They require as input boundaries and times for the interim analyses. The package can be used to design sequential trials, determine boundary values while the trial is ongoing or compute confidence intervals when the trial is ended. We present examples for design or analyses using test statistics comparing mean, binomial, survival or repeated measures outcomes.

Summary of methodology

A detailed presentation of the methodology may be found in Lan & DeMets (1983), DeMets & Lan (1984), and Lan & Zucker (1993). Group sequential procedures for interim analyses are equivalent to discrete boundary crossing problems for a Brownian motion process $W(t)$ with drift parameter θ. We take advantage of this correspondence in both theoretical developments and in implementation. At each interim analysis, a standardized test statistic Z_k is computed. These normally distributed variates $Z_1, \ldots, Z_k$ have mean $\theta\sqrt{\tau_k}$, where θ is the "drift" parameter, and for $j \leq k$, $\mathrm{Cov}(Z_j, Z_k) = \sqrt{\tau_j/\tau_k}$ where τ_k is the information fraction (or information time) at the k^{th} analysis, e.g. $\tau_k = n_k/n_K$ if n_K is the maximum sample size (per arm). The drift parameter θ and the standardized difference δ_S are related by the equation $\mathrm{E}(Z_K) = \theta = \delta_S\sqrt{n_K}$.

To reiterate in more technical terms, the program uses Equation (1) to determine one of

- $b_1, \ldots, b_k$ for given $\alpha^\star(\tau)$, and $\tau_1, \ldots, \tau_k$,
- θ given p_k, $b_1, \ldots, b_k$ and $\tau_1, \ldots, \tau_k$

Table 1: Correspondence of notation for commonly used group sequential parameters.

Parameter		Kim and DeMets	Pocock
Standardized difference	δ_S	ζ	$\delta/\sqrt{2}\sigma$
Drift of Brownian motion	θ	ξ	$\Delta\sqrt{N}$
Pocock noncentrality	$\theta/\sqrt{K}$	$\xi/\sqrt{K}$	Δ
Accumulated sample size per arm	n_k	n_k	ni
Maximum sample size per arm	n_K or nK	n_K	nN

- $p_1, \ldots, p_K$, where $p_k = \Pr(Z_1 \geq b_1 \text{ or } \ldots \text{ or } Z_k \geq b_k)$, given $b_1, \ldots, b_k$ and $\tau_1, \ldots, \tau_k$,
- a confidence interval for θ given $b_1, \ldots, b_k$, $\tau_1, \ldots, \tau_k$, and Z_k

It may be useful to note correspondences between the notation used here and in some other references (see Table 1). To clarify notation for the sample size, let n_k be the number of subjects at the k^{th} look in each treatment arm. n_K is the maximum number of subjects per treatment arm and K is the maximum number of looks or interim analyses. If there are n subjects accumulated between interim analyses, $n_K = nK$. The drift parameter θ can be expressed in terms of the noncentrality parameter Δ in Pocock (1977) as $\theta = \Delta\sqrt{K}$.

2 Use of the Program for Study Design

Although spending functions provide flexibility in data monitoring and do not require analysis times to be prespecified, the anticipated number and timing of interim analyses must be specified for design purposes. This is not more restrictive than for the group sequential procedures proposed by Pocock (1977) or O'Brien & Fleming (1979). Deviation from the initial design, even substantially, does not cause a serious loss of power. Thus for design only, we shall assume $n_k = nK$, where K is the anticipated number of interim analyses and n is the anticipated number of subjects accrued between analyses.

Kim & DeMets (1992) provide a detailed discussion of sample size determination for group sequential testing. The relationship between sample size and power depends on two quantities: the drift parameter of the underlying Brownian motion and the standardized difference between control and treatment arms. Thus by determining θ and δ_S for a particular design problem, the required sample size can be computed. The value of θ depends on the desired power, the set of boundaries and analysis times, and the properties of Brownian motion. Exit or rejection probabilities for Brownian motion given a set of boundaries can be computed by the program or, for certain designs, found in the tables provided by Kim & DeMets (1992). The sequential boundaries are determined by the choice of spending function $\alpha^\star(\tau)$, the number and timing of interim analyses, the α level and whether the test is one or two sided. The standardized difference δ_S, on the other hand, depends on the type of data to be collected by the study. Several examples are detailed below for normal, binomial and survival data.

Kim & DeMets (1992) provide tables of drift parameters for spending functions producing O'Brien-Fleming type and Pocock type boundaries ($\alpha_1^\star(\tau)$ and $\alpha_2^\star(\tau)$, respectively). The program currently offers five choices for spending functions, but others can be added (see Appendix).

2.1 Normally distributed data

2.1.1 Kim-DeMets example

Kim & DeMets (1992) discuss the following example. Suppose that a normally distributed response has mean in controls of $\mu_C = 220$ with standard deviation $\sigma = 30$. The null hypothesis is $H_0 : \mu_C = \mu_E$, where μ_E is the mean in the experimental group, expected to be 200. The test statistic is

$$Z_k = \frac{\bar{x}_C - \bar{x}_E}{\sqrt{2\sigma^2/n_k}}.$$

Then the drift parameter is

$$\theta = \frac{\sqrt{n_K}(\mu_C - \mu_E)}{\sqrt{2\sigma^2}} = \sqrt{n_K}\,\delta_S.$$

So

$$n_K = \frac{2\sigma^2}{(\mu_C - \mu_E)^2} \times \theta^2.$$

For the program, we specify two-sided $\alpha = 0.05$ O'Brien-Fleming type ($\alpha_1^\star$) boundaries with $K = 5$ looks at 0.2, 0.4, 0.6, 0.8 and 1.0 (see Section 4.1). The output boundary values are

τ	O'Brien-Fleming type boundary values
0.2	$\pm$ 4.8769
0.4	$\pm$ 3.3569
0.6	$\pm$ 2.6803
0.8	$\pm$ 2.2898
1.0	$\pm$ 2.0310

Kim & DeMets (1992) indicate that $\theta = 3.28$ for 90% power, so

$$n_K = \frac{2(30)^2}{(220-200)^2} \times (3.28)^2 = 48.41.$$

The program can verify that $\theta = 3.28$ corresponds to 90% power, and that alternative timings of analyses does not greatly affect the power (see Section 4.1). The effect of alternative assumptions for μ_E on sample size can be determined without recomputing θ.

2.1.2 Kim-DeMets example with Pocock boundary

Suppose that in the previous example the O'Brien-Fleming type boundaries were replaced with Pocock type boundaries ($\alpha_2^\star$). The computations are identical except for the value of θ. Two-sided 0.05 Pocock type boundary values are

τ	Pocock type boundary values
0.2	$\pm$ 2.4380
0.4	$\pm$ 2.4268
0.6	$\pm$ 2.4101
0.8	$\pm$ 2.3966
1.0	$\pm$ 2.3859

Kim and DeMets (1992) indicate that for 90% power using these boundaries, $\theta = 3.55$, so

$$n_K = \frac{2(30)^2}{(220-200)^2} \times (3.55)^2 = 56.71.$$

2.1.3 An example using Pocock's notation

We duplicate an example from Pocock (1977). If we take $\alpha = 0.05$ and $N = 5$, corresponding boundaries are determined. For a desired power $1-\beta = 0.90$, we determine using the program that $\Delta\sqrt{N} = 3.28$ so that $\Delta = 1.59$. To compare two sample means, we compute

$$Z_i^* = (\bar{x}_E - \bar{x}_C)\sqrt{ni}/\sqrt{2\sigma^2}$$

where $i = 1, \ldots, N$ and $H_0 : \mu_C = \mu_E$ and from Pocock (1977)

$$\Delta = \sqrt{n}(\mu_E - \mu_C)/\sqrt{2\sigma^2}.$$

For $H_A : \mu_E - \mu_C = 0.5\sigma$,

$$n = \frac{2\sigma^2}{(\mu_E - \mu_C)^2} \times \Delta^2 = \frac{2\sigma^2}{(0.5\sigma)^2} \times (1.59)^2 = 20.22,$$

so $2nN = 2(20)(5) = 200$ subjects.

2.2 Binomially distributed data

In the binomial case, where we test $H_0 : p_E = p_C$, assume $p_C = 0.4$ and $p_E = 0.6$. The statistic

$$Z_k = (\hat{p}_E - \hat{p}_C)/\sqrt{\hat{p}_E(1-\hat{p}_E) + \hat{p}_C(1-\hat{p}_C)}$$

has asymptotically a normal distribution with a mean of 0 and a variance of 1 (under $H_0 : p_C = p_T$). The standardized difference is

$$\delta_S = (p_E - p_C)/\sqrt{p_E(1-p_E) + p_C(1-p_C)} \approx (p_E - p_C)/\sqrt{2\bar{p}(1-\bar{p})}$$

where $\bar{p} = (p_E + p_C)/2$.

2.2.1 Kim and DeMets example

Kim & DeMets (1992) show $\theta = \mathrm{E}(Z_K) = \sqrt{n_K}\delta_S = \sqrt{n_K}(p_C - p_E)/\sqrt{2\bar{p}(1-\bar{p})}$ so

$$n_K = \frac{2\bar{p}(1-\bar{p})}{(p_C - p_E)^2} \times \theta^2.$$

For example, if $p_C = 0.4$ and $p_E = 0.6$ under the alternative hypothesis, then $\bar{p} = 0.5$, and for a one sided $\alpha = 0.05$ test using five interim analyses and Pocock type boundaries ($\alpha_2^\star$), we have

τ	Pocock type Z values
0.2	+2.1762
0.4	+2.1437
0.6	+2.1132
0.8	+2.0895
1.0	+2.0709

For $H_A : p_E > p_C$ and 90% power, Kim & DeMets (1992) report $\theta = 3.21$ (or see Section 4.2), so

$$n_K = \frac{2(.5)(.5)}{(.2)^2} \times (3.21)^2 = 128.80.$$

2.2.2 O'Brien-Fleming example

As another binomial example, consider a two sided $\alpha = 0.05$ test with O'Brien-Fleming type ($\alpha_1^\star$) boundaries, and for design purposes only, assume $K = 5$ equally spaced analyses at 0.2, 0.4, 0.6, 0.8 and 1.0. As above, we take $H_0 : p_C = p_E$, but now let $p_C = 0.11$ and $p_E = 0.0825$ under the alternative hypothesis (a 25% reduction, $\bar{p} = 0.096$). The program produces

τ	O'Brien-Fleming type Z values
0.2	$\pm$ 4.8769
0.4	$\pm$ 3.3569
0.6	$\pm$ 2.6803
0.8	$\pm$ 2.2898
1.0	$\pm$ 2.0310

From Kim and DeMets (1992), $\theta = 3.28$ (see Section 4.1) so

$$n_K = \frac{2(0.096)(0.904)}{(0.028)^2} \times (3.28)^2 = 2381.78.$$

2.3 Survival data

Suppose we are interested in comparing the hazard rate of two populations. Let $\lambda_0(u)$ be the hazard function of the control group and $\lambda_1(u)$ the hazard function in the treatment group. Under the null hypothesis $\lambda_0 = \lambda_1$ and $\phi = \log(\lambda_1/\lambda_0) = 0$. The logrank statistic is

$$L(d) = \sum_{i=1}^{d} \left(x_i \frac{r_{ic}}{r_{ic} + r_{it}} \right)$$

where d is the number of events, x_i is 1 if the event at t_i is in the control group and 0 if it is in the treatment group, r_{ic} is the number of patients in the control group at risk just before t_i, and r_{it} is the number of patients in the treatment group at risk just before t_i. The expected value of $L(d)$ is approximately $\phi \times (d/4)$, and the estimated variance is

$$\sum_{i=1}^{d} \left(\frac{r_{ic}}{r_{ic} + r_{it}} \right) \left(1 - \frac{r_{ic}}{r_{ic} + r_{it}} \right) \approx d/4$$

These approximations are reasonable if $r_{it} \approx r_{ic}$ and ϕ is close to 0. If d_k is the number of events at analysis k, the statistic $Z_k = L(d_k)/\sqrt{d_k/4}$ has a $N(\phi\sqrt{d_k/4}, 1)$ distribution, so $\theta = \phi\sqrt{d_K/4}$. Then the maximum number of events required per arm is

$$d_K = 4\theta^2/\phi^2.$$

If we assume $\lambda_1/\lambda_0 = 0.2$ and $\theta = 3.261$ (see Section 4.3),

$$d_K = 4\,(3.261)^2/(\log 0.2)^2 = 16.42.$$

2.4 Repeated measures

Many clinical trials are designed to measure subjects repeatedly over the course of the trial, and define as the primary outcome the change or slope over time. For such trials, the difference between treatment groups can be tested using the estimated slopes from each group using

$$Z_k = (\bar{B}_T^{(k)} - \bar{B}_C^{(k)})/\sqrt{U_T^{(k)} + U_C^{(k)}}$$

where $\bar{B}_T^{(k)}$ and $\bar{B}_C^{(k)}$ are the average of the slopes estimated for patients in the treatment and control groups at the k^{th} interim analysis, and $U_T^{(k)}$ and $U_C^{(k)}$ are their variances. The sequentially computed Z_k have been shown to have the required Brownian motion structure when the variance parameters are known (Reboussin, Lan & DeMets, 1992; Wu & Lan, 1992). Lan, Reboussin & DeMets (1994) show

$$\mathrm{E}(Z_K) = \frac{B_T - B_C}{\sigma_B}\sqrt{\frac{\hat{I}}{4}}$$

where B_T and B_C are the mean population slopes, σ_B^2 is the between patient variance of the slopes, and $\hat{I}$ is the natural estimate of total information at the end of the trial. For the comparison of means and binomial proportions, $\hat{I} = 2n_K$, but in this case, the natural estimate of total information, denoted $\hat{\imath}$, is the sum of the natural estimates of information for each patient:

$$\hat{\imath} = \left[1 + R/\sum(t_j - \bar{t})^2\right]^{-1}$$

where R is the ratio of within to between patient variance. For design purposes, we may assume an identical number and timing of measurements for all patients, so that $\hat{I}$ is $2n_K\,\hat{\imath}$.

Then

$$\delta_S = \frac{B_T - B_C}{\sqrt{2\sigma_B^2}}$$

and

$$\theta = \frac{B_T - B_C}{\sqrt{2\sigma_B^2}}\sqrt{n_K \hat{i}}$$

so

$$n_K = \theta^2 \frac{2\sigma_B^2}{(B_T - B_C)^2}\,\frac{1}{\hat{i}} = \theta^2 \frac{2\sigma_B^2}{(B_T - B_C)^2}\left[1 + R/\sum(t_j - \bar{t})^2\right].$$

If a sufficient number of observations are taken on each patient, the term $1 + R/\sum(t_j - \bar{t})^2$ is nearly one (Lan, Reboussin & DeMets, 1994), so that the power computations are similar to the normal case.

3 Using the Program to Sequentially Analyze a Trial

We describe how to run the program using data from the Beta-Blocker Heart Attack Trial or BHAT (Beta-Blocker Heart Attack Trial Research Group, 1982). BHAT, a study sponsored by the National Heart, Lung and Blood Institute, was designed to test whether long term use of propranolol by patients with recent heart attack reduced mortality. The following example does not correspond exactly to what was actually done for BHAT, though it is similar. From June 1978 to October 1980, 3837 patients were randomized to either propranolol (1916 patients) or placebo (1921 patients). Follow-up was originally scheduled to end in June 1982. The total information D (number of deaths by June 1982) was never observed since the trial was terminated early in October 1981. The value of D was estimated to be 628 when BHAT was designed, but with the data available in September 1982, was estimated to be around 400 (Lan & DeMets, 1989). In the six Policy and Data Monitoring Board meetings (May 1979, October 1979, March 1980, October 1980, April 1981, and October 1981), the observed number of deaths were (56, 77, 126, 177, 247, 318) and normalized log-rank statistics were (1.68, 2.24, 2.37, 2.30, 2.34, 2.82).

3.1 Computing boundaries with a spending function

Let t_c denote calendar time measured from the beginning of the trial, and T_c denote the maximum duration in calendar time. Let τ be the information fraction or "information time", which must often be estimated by $\hat{\tau}$, some function either of calendar time or number of observed patients or events. We begin with an example using only calendar time.

Example with calendar time

Set $t_c = 0$ in June 1978 and assume the maximum duration is $T_c = 48$ months, which corresponds to June 1982. Then the calendar times for interim analyses correspond to (11, 16, 21, 28, 34, 40) months after the start of the trial. We estimate τ as a function of calendar time by $\hat{\tau} = t_c/T_c = t_c/48$, so the information times are (0.2292, 0.3333, 0.4375, 0.5833, 0.7083, 0.8333), and adopt the spending function $\alpha^\star(\tau) = \alpha\tau$ to construct a data monitoring boundary. This corresponds to $\alpha_3^\star(\tau)$ in Lan & DeMets (1983) and Kim & DeMets (1987a). The original BHAT design had a two-sided significance level of 0.05.

When the data were monitored in May 1979, $t_{c1} = 11$, $\hat{\tau}_1 = 11/48 = 0.2292$ and $\alpha^\star(\hat{\tau}_1) = 0.025 \times 0.2292 = 0.0057$. The program produces a boundary value of $b_1 = 2.53$: if Z_1 is standard normal, $\Pr(Z_1 \geq 2.53) = 0.0057$. In October 1979, $t_{c2} = 16$, $\hat{\tau}_2 = 16/48 = 0.3333$, and $\alpha^\star(\hat{\tau}_2) = 0.0083$. Ignoring the observed number of deaths and using only calendar time, the calculation proceeds as follows. Suppose Z_1 and Z_2 are standard normal with correlation coefficient $\rho_{12} = (0.2292/0.3333)^{1/2} = 0.8293$. We wish to find b_2 such that $\Pr(Z_1 \geq 2.53 \,\text{or}\, Z_2 \geq b_2) = 0.0083$. This solution requires some numerical integration which the program performs. In fact, this equality is satisfied if $b_2 = 2.61$.

In this example, after specifying Option 1, the user is prompted for

- the number interim analyses (2),
- whether the analyses are equally spaced (no),
- times of the interim analyses (0.2292, 0.3333),
- whether a second time scale for information will be entered (Lan and DeMets, 1989) (no),
- the overall significance level (.05)
- whether the test was one-sided or two-sided symmetric (2),
- which $\alpha^\star$ function to apply ($\alpha_3^\star$)
- whether the boundary values should be truncated (no)

(see Section 4.4). The program returns $b_1 = 2.53$ and $b_2 = 2.61$. For the third analysis, we enter 0.2292, 0.3333, and 0.4375. The program returns b_1 and b_2 as before, plus $b_3 = 2.57$. Continuing in this manner, we obtain the boundary values (2.53, 2.61, 2.57, 2.47, 2.43, 2.38).

Example with information

We now repeat the above calculation using the information in the number of deaths. Assuming the total information is the number of expected events, $D = 628$, the information fractions are (56/628, 77/628, 126/628, 177/628, 247/628, 318/628), or (0.0892, 0.1226, 0.2006, 0.2818, 0.3933, 0.5064). Then at the second interim analysis, the program would ask for

- the number interim analyses (2),
- whether the analyses are equally spaced (no),
- times of the interim analyses (0.0892, 0.1226),
- whether a second time scale for information will be entered (Lan and DeMets, 1989) (no),
- the overall significance level (.05)
- whether the test was one-sided or two-sided symmetric (2),
- which $\alpha^\star$ function to apply ($\alpha_3^\star$).
- whether the boundary values should be truncated (no)

The information fractions are treated as times. Since we do not enter the information separately apart from the information fractions, the answer to the question on a second time scale is "no". The output boundary values are (2.84, 2.97). At the sixth analysis, when the additional times are input, the resulting boundary values are (2.84, 2.97, 2.79, 2.72, 2.61, 2.54).

Two time scales

Some users may be familiar with the use of both information and calendar time as described in Lan & DeMets (1989) and Lan, Reboussin & DeMets (1994). The program includes such an option. We will use the percent of elapsed calendar time to determine how much type I error probability is to be spent, but for the correlation of successive test statistics, we will use the information in the number of deaths. The first boundary is computed exactly as above. For the analysis in October 1979, at 16 months, $t_{c2} = 16$, $\hat{\tau}_2 = 16/48 = 0.3333$, and $\alpha^\star(\hat{\tau}_2) = 0.0083$ also just as before. To evaluate b_2, note that even though $\tau_2 = 77/D$ is unknown, $\tau_1/\tau_2 = 56/77 = 0.7273$ is observed. If Z_1 and Z_2 are standard normal then the correlation coefficient $\rho_{12} = \sqrt{0.7273} = 0.8528$, and the solution to $\Pr(Z_1 \geq 2.53 \,\text{or}\, Z_2 \geq b_2) = 0.0083$

is $b_2 = 2.59$. The program asks the same questions as before (see Section 4.6). Since the times entered were based on the percent of elapsed calendar time, it is desirable to use the information available in the number of deaths. When the question on a second time scale for information is asked, we answer "yes" and enter the information for each analysis, which is the number of deaths in this example. The resulting boundaries are (2.53, 2.59, 2.63, 2.50, 2.51, 2.47) for the six data monitoring points of BHAT, and this boundary is crossed at $t_{c6} = 40$ or in October of 1981. This is the same as the result given for the example in Lan & DeMets (1989).

3.2 Computing confidence intervals

Kim & DeMets (1987b) detail the theory for confidence intervals following early termination using group sequential tests. Suppose that a trial has been stopped at the k^{th} analysis with boundary values $b_1, \ldots, b_k$, and with final standardized estimate of treatment difference Z_k. The confidence interval is based on computing upper exit probabilities associated with $b_1, \ldots, b_{k-1}, Z_k$.

Continuing with the previous example, the final observed standardized statistic was 2.82, and suppose that a 95 percent confidence interval is desired. The program prompts for

- the number of analyses (6),
- whether the analyses are equally spaced between 0 and 1 (no),
- the information times of the analyses (.2292, .3333, etc.),
- whether a spending function will be used (no),
- whether the boundary is one or two sided (2),
- whether the two sided boundary is symmetric (yes),
- the boundaries to be evaluated (2.53, 2.61, etc.).
- the value of the standardized statistic at the last analysis (2.82),
- the confidence level (0.95).

Computation of confidence intervals is the most time consuming of the four options since it involves a linear search. The program outputs the result (0.1881, 4.9347) (see Section 4.7).

Using the equation $\theta = \phi\sqrt{d/4}$ we can translate this interval into an interval for ϕ. The statistic is based on 318 events, so $0.1881 = \phi\sqrt{318/4}$, or $\phi = 0.021$ is the lower bound. Repeating this computation for the upper bound, we obtain (0.021, 0.553) as a 95% confidence interval for ϕ.

4 Examples with Output

This section contains examples of interactive sessions with the program, which were used for the examples considered in Sections 2 and 3.

4.1 Normally distributed data

This program output related to the first example in Section 2.1. For this example, we use 5 equally spaced interim analyses (0.2, 0.4, 0.6, 0.8, and 1.0) with two–sided O'Brien–Fleming boundaries and $\alpha = 0.05$. We first determine the boundaries and then for these boundaries, determining the drift parameter θ to calculate a sample size.

```
          PROGRAM PROMPTS                                                 USER INPUT
 Is this an interactive session? (1=yes,0=no)
                                                                              y
 interactive =            1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                              1
 Option 1: You will be prompted for a spending function.
 Number of interim analyses?
                                                                              5
         5 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                              y
Analysis times:    0.200    0.400    0.600    0.800    1.000
 Do you wish to specify a second time/information scale? (e.g.
 number of patients or number of events, as in Lan & DeMets 89?)  (1=yes, 0=no)
                                                                              n
 Overall significance level? (>0 and <=1)
                                                                              .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                              2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                              1
 Use function alpha-star           1
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)               n

 Bounds will not be truncated.

This program generates two-sided symmetric boundaries.
n =  5
alpha = 0.050
use function for the lower boundary = 1
use function for the upper boundary = 1
       Time                Bounds         alpha(i)-alpha(i-1)   cum alpha
    0.20           -4.8769     4.8769         0.00000         0.00000
```

```
   0.40           -3.3569     3.3569        0.00079         0.00079
   0.60           -2.6803     2.6803        0.00683         0.00762
   0.80           -2.2898     2.2898        0.01681         0.02442
   1.00           -2.0310     2.0310        0.02558         0.05000

Do you want to see a graph? (1=yes,0=no)
                                                                          y

        :
   5.00:        *
   4.60:
   4.20:
   3.80:
   3.40:                 *
   3.00:
   2.60:                         *
   2.20:                                  *       *
   1.80:
   1.40:
   1.00:
   0.60:
   0.20:
  -0.20:
  -0.60:
  -1.00:
  -1.40:
  -1.80:
  -2.20:                                  *       *
  -2.60:                         *
  -3.00:
  -3.40:                 *
  -3.80:
  -4.20:
  -4.60:
  -5.00:        *
        ..............................................
          0  .1  .2   .3  .4  .5  .6  .7   .8  .9  1

Done.
```

Once these initial boundaries are obtained, to compute the required sample size, we must find the drift parameter corresponding to the desired power. In the program, this is option 2. We enter the times and boundary values and select the desired power. Alternatively, drift parameters for some potential analysis scenarios are contained in Kim & DeMets (1992). In our example, a drift parameter of 3.2788 gives a power of 0.90.

```
          PROGRAM PROMPTS                                        USER INPUT
Is this an interactive session? (1=yes,0=no)
                                                                      y
interactive =           1
Enter number for your option:
(1) Compute bounds for given spending function.
(2) Compute drift for given power and bounds
(3) Compute probabilities for given bounds.
(4) Compute confidence interval.
                                                                      2
Option 2: You will be prompted for bounds and a power level.
Number of interim analyses?
                                                                      5
         5 interim analyses.
```

```
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                          y
Analysis times:    0.200    0.400    0.600    0.800    1.000
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                          y
 Spending function will determine bounds.
 Overall significance level? (>0 and <=1)
                                                                          .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                          2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                          1
 Use function alpha-star           1
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)           n

 Bounds will not be truncated.
      Time                Bounds
   0.20            -4.8769     4.8769
   0.40            -3.3569     3.3569
   0.60            -2.6803     2.6803
   0.80            -2.2898     2.2898
   1.00            -2.0310     2.0310
 Desired power? (>0 and <=1)
                                                                          .9
 Power is 0.900

   n =  5, drift =  3.2788

look     time      lower    upper     exit  probability    cum exit pr

 1       0.20    -4.8769   4.8769        0.00032           0.00032
 2       0.40    -3.3569   3.3569        0.09939           0.09971
 3       0.60    -2.6803   2.6803        0.34658           0.44629
 4       0.80    -2.2898   2.2898        0.29966           0.74595
 5       1.00    -2.0310   2.0310        0.15405           0.90000

 Done.
```

A drift of 3.28 was used in Section 2.1.1 to compute the required sample size for 90% power, which was 48.44 patients per arm.

Consider another sample size determination based on a different initial analysis plan. This set of analyses will be planned for unequally spaced time points 0.1, 0.4, 0.75, 1.0, but other features of the test are the same. The program determines the corresponding drift parameter.

```
          PROGRAM PROMPTS                                             USER INPUT
 Is this an interactive session? (1=yes,0=no)
                                                                          y
 interactive =           1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
```

```
 (4) Compute confidence interval.
                                                                           2
 Option 2: You will be prompted for bounds and a power level.
 Number of interim analyses?
                                                                           4
         4 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                           n
 Times of interim analyses: (>0 & <=1)
                                                                  .1 .4 .75 1.0
Analysis times:    0.100    0.400    0.750    1.000
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                           y
 Spending function will determine bounds.
 Overall significance level? (>0 and <=1)
                                                                         .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                           2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                           1
 Use function alpha-star          1
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)           n

 Bounds will not be truncated.
      Time               Bounds
    0.10            -6.9914     6.9914
    0.40            -3.3569     3.3569
    0.75            -2.3449     2.3449
    1.00            -2.0125     2.0125
 Desired power? (>0 and <=1)
                                                                          .9
 Power is 0.900

    n =  4, drift =  3.2696

look     time      lower    upper     exit  probability    cum exit pr

 1       0.10    -6.9914   6.9914       0.00000           0.00000
 2       0.40    -3.3569   3.3569       0.09871           0.09871
 3       0.75    -2.3449   2.3449       0.58876           0.68746
 4       1.00    -2.0125   2.0125       0.21254           0.90000

 Done.
```

The sample size is computed

$$n_K = \left(\frac{\theta}{\delta_S}\right)^2 = \left(\frac{3.27}{0.4714}\right)^2 = 48.11.$$

Notice that the different timing of interim analyses has little impact on the sample size needed to achieve 90% power.

4.2 Binomially distributed data

In much the same manner as was done to compare two means from a normal population, we can compare two proportions from a binomial population. Recall the example from Section 2.2.1. We use option 2 to determine the drift parameter for a power of 90% given one sided 0.05 Pocock boundaries and five equally spaced analyses:

```
          PROGRAM PROMPTS                                                  USER INPUT
 Is this an interactive session? (1=yes,0=no)
                                                                             y
 interactive =           1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                             2
 Option 2: You will be prompted for bounds and a power level.
 Number of interim analyses?
                                                                             5
         5 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                             y
Analysis times:    0.200    0.400    0.600    0.800    1.000
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                             y
 Spending function will determine bounds.
 Overall significance level? (>0 and <=1)
                                                                             .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                             1
1.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                             2
 Use function alpha-star           2
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)              n

 Bounds will not be truncated.
       Time               Bounds
    0.20            -8.0000     2.1762
    0.40            -8.0000     2.1437
    0.60            -8.0000     2.1132
    0.80            -8.0000     2.0895
    1.00            -8.0000     2.0709
 Desired power? (>0 and <=1)
                                                                             .9
 Power is 0.900

    n =  5, drift =  3.2055

look     time      lower    upper     exit  probability    cum exit pr

 1       0.20    -8.0000   2.1762       0.22884           0.22884
 2       0.40    -8.0000   2.1437       0.25845           0.48729
 3       0.60    -8.0000   2.1132       0.19989           0.68718
 4       0.80    -8.0000   2.0895       0.13238           0.81956
```

```
5       1.00    -8.0000   2.0709        0.08044           0.90000

Done.
```

4.2.1 The impact of changing frequency

Even if the interim analyses actually performed during the study are not equally spaced, the power is not greatly affected. This can be seen in the following example. Recall our original plan had looks at 0.2, 0.4, 0.6, 0.8 and 1.0 and a target power of 90%. Suppose instead the looks occur at 0.2, 0.5, 0.6, 0.8, and 1.0. Option 3 generates appropriate boundaries and computes the power for a drift of 3.21. As shown, the power is not seriously affected.

```
          PROGRAM PROMPTS                                            USER INPUT
 Is this an interactive session? (1=yes,0=no)
                                                                        y
 interactive =           1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                        3
 Option 3: You will be prompted for bounds or a spending
 function to compute them.
 Number of interim analyses?
                                                                        5
         5 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                        n
 Times of interim analyses: (>0 & <=1)
                                                                   .2 .5 .6 .8 1.0
Analysis times:    0.200    0.500    0.600    0.800    1.000
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                        y
 Spending function will determine bounds.
 Overall significance level? (>0 and <=1)
                                                                       .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                        1
1.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                        2
 Use function alpha-star          2
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)         n

 Bounds will not be truncated.
      Time              Bounds
   0.20           -8.0000     2.1762
   0.50           -8.0000     2.0435
   0.60           -8.0000     2.1609
   0.80           -8.0000     2.0866
   1.00           -8.0000     2.0680
 Do you wish to use drift  parameters? (1=yes, 0=no)                    y
```

```
 How many drift parameters do you wish to enter?
                                                                                               1
         1 drift parameters.
 Enter drift parameters:
                                                                                             3.21
Drift parameters:    3.210
 Drift is equal to the standard treatment difference times the square
 root of total information per arm.

    n =  5, drift =  3.2100

look      time      lower    upper     exit  probability    cum exit pr

 1       0.20    -8.0000   2.1762       0.22945           0.22945
 2       0.50    -8.0000   2.0435       0.38289           0.61234
 3       0.60    -8.0000   2.1609       0.07757           0.68991
 4       0.80    -8.0000   2.0866       0.13220           0.82211
 5       1.00    -8.0000   2.0680       0.07941           0.90152

 Done.
```

4.3 Survival data

Referring to the previous survival example in Section 2.3, assume that three equally spaced analyses were initially planned for this study, and that test was to have 90% power. The following output from the program illustrates the Brownian motion drift parameter of 3.261 will give the desired power.

```
        PROGRAM PROMPTS                                                        USER INPUT

    Is this an interactive session? (1=yes,0=no)
                                                                                    y
 interactive =            1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                                    2
 Option 2: You will be prompted for bounds and a power level.
 Number of interim analyses?
                                                                                    3
         3 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                                    y
Analysis times:    0.333    0.667    1.000
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                                    y
 Spending function will determine bounds.
 Overall significance level? (>0 and <=1)
                                                                                   .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                                    2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
```

```
 (5) alpha * t^2
                                                                           1
 Use function alpha-star           1
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)               n

 Bounds will not be truncated.
       Time                Bounds
    0.33            -3.7103     3.7103
    0.67            -2.5114     2.5114
    1.00            -1.9930     1.9930
 Desired power? (>0 and <=1)
                                                                          .90
 Power is 0.900

    n =  3, drift =  3.2608

look     time      lower    upper     exit  probability    cum exit pr

 1       0.33    -3.7103   3.7103       0.03380           0.03380
 2       0.67    -2.5114   2.5114       0.52651           0.56031
 3       1.00    -1.9930   1.9930       0.33969           0.90000

 Done.
```

4.4 Computing bounds during analysis of a trial

This is an interactive session using the BHAT data and calendar time as the only time scale. The input sequence is described in Section 3.

```
          PROGRAM PROMPTS                                                USER INPUT

 Is this an interactive session? (1=yes,0=no)
                                                                              y
 interactive =           1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                              1
 Option 1: You will be prompted for a spending function.
 Number of interim analyses?
                                                                              2
          2 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                              n
 Times of interim analyses: (>0 & <=1)
                                                                        .2292 .3333
Analysis times:    0.229    0.333
 Do you wish to specify a second time/information scale? (e.g.
 number of patients or number of events, as in Lan & DeMets 89?)  (1=yes, 0=no)
                                                                             no
 Overall significance level? (>0 and <=1)
                                                                            .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                             2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
```

```
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                              3
 Use function alpha-star           3
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)                n

 Bounds will not be truncated.

This program generates two-sided symmetric boundaries.
n =  2
alpha = 0.050
use function for the lower boundary = 3
use function for the upper boundary = 3
       Time                Bounds         alpha(i)-alpha(i-1)   cum alpha
    0.23            -2.5284     2.5284         0.01146         0.01146
    0.33            -2.6098     2.6098         0.00520         0.01667

 Do you want to see a graph? (1=yes,0=no)
                                                                              n

 Done.
```

In this case, the program outputs the number of analyses so far, the type I error specified, the use function chosen, the times, the computed boundaries, and the type I error "spent" at each analysis so far.

4.5 Using the program noninteractively

Some users may want to use the program noninteractively. This can be done by preparing an input file with the appropriate format. Each question is answered on its own line in the input file, and the answer to the first question must be "no" or "0". Here is an input file which reproduces the above interactive session:

```
0                                      # noninteractive
1                                      # option 1: bounds
2                                      # number of analyses
0                                      # equally spaced? (0=no)
.2292 .3333                            # times of analyses
0                                      # second time scale? (0=no)
.05                                    # alpha
2                                      # 1 or 2 sided test
3                                      # use function (1-5)
0                                      # truncate boudaries (0=no)
0                                      # show graph? (0=no)
0                                      # start again? (0=no)
```

The resulting output is

```
 Is this an interactive session? (1=yes,0=no)
 interactive =           0
          2 interim analyses.
Analysis times:    0.229    0.333
 alpha = 0.050
2.-sided test
 Use function alpha-star           3
```

```
This program generates two-sided symmetric boundaries.
n =  2
alpha = 0.050
use function for the lower boundary = 3
use function for the upper boundary = 3
       Time                Bounds         alpha(i)-alpha(i-1)   cum alpha
    0.23            -2.5284     2.5284         0.01146         0.01146
    0.33            -2.6098     2.6098         0.00520         0.01667

 Do you want to see a graph? (1=yes,0=no)

 Done.
```

4.6 Using information to compute boundaries during analysis

For this session, the numbers of events were entered as information, as described in Section 3.1.

```
           PROGRAM PROMPTS                                              USER INPUT

 Is this an interactive session? (1=yes,0=no)
                                                                           y
 interactive =            1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                           1
 Option 1: You will be prompted for a spending function.
 Number of interim analyses?
                                                                           6
          6 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                           n
 Times of interim analyses: (>0 & <=1)
                                                 .2292 .3333 .4375 .5833 .7083 .8333
Analysis times:    0.229    0.333    0.438    0.583    0.708    0.833
 Do you wish to specify a second time/information scale? (e.g.
 number of patients or number of events, as in Lan & DeMets 89?)  (1=yes, 0=no)
                                                                           y
 Second scale will estimate covariances.
 Information:
                                                               56 77 126 177 247 318
Information  56.000  77.000 126.000 177.000 247.000 318.000
 Overall significance level? (>0 and <=1)
                                                                          .05
 alpha = 0.050
 One(1) or two(2)-sided symmetric?
                                                                           2
2.-sided test
 Use function? (1-5)
 (1) OBrien-Fleming type
 (2) Pocock type
 (3) alpha * t
 (4) alpha * t^1.5
 (5) alpha * t^2
                                                                           3
 Use function alpha-star           3
 Do you wish to truncate the standardized bounds? (1=yes, 0=no)            n
```

```
 Bounds will not be truncated.

This program generates two-sided symmetric boundaries.
n =  6
alpha = 0.050
use function for the lower boundary = 3
use function for the upper boundary = 3
    Time    Information     Bounds      alpha(i)-alpha(i-1)  cum alpha
  0.23       56.00    -2.5284   2.5284       0.01146       0.01146
  0.33       77.00    -2.5905   2.5905       0.00520       0.01667
  0.44      126.00    -2.6327   2.6327       0.00521       0.02187
  0.58      177.00    -2.5036   2.5036       0.00729       0.02916
  0.71      247.00    -2.5073   2.5073       0.00625       0.03542
  0.83      318.00    -2.4655   2.4655       0.00625       0.04166

 Do you want to see a graph? (1=yes,0=no)
                                                                              n

 Done.
```

In addition to the output described previously, the information is also reported.

4.7 Computing a confidence interval at the end of a trial

In addition to the information needed to compute probabilities associated with a set of boundaries, computing a confidence interval also requires the last value of the standardized test statistic.

```
          PROGRAM PROMPTS                                              USER INPUT

 Is this an interactive session? (1=yes,0=no)
                                                                               y
 interactive =            1
 Enter number for your option:
 (1) Compute bounds for given spending function.
 (2) Compute drift for given power and bounds
 (3) Compute probabilities for given bounds.
 (4) Compute confidence interval.
                                                                               4
 Option 4: You will be prompted for bounds and a confidence level.
 Number of interim analyses?
                                                                               6
         6 interim analyses.
 Equally spaced times between 0 and 1? (1=yes,0=no)
                                                                               n
 Times of interim analyses: (>0 & <=1)
                                                            .2292 .3333 .4375 .5833 .7083 .8333
Analysis times:    0.229    0.333    0.438    0.583    0.708    0.833
 Are you using a spending function to determine bounds? (1=yes,0=no)
                                                                              no
 You must enter a set of bounds.
 One(1)- or two(2)-sided?
                                                                               2
           2-sided test
 Symmetric bounds? (1=yes,0=no)
                                                                               y
 Two sided symmetric bounds.
 Enter upper bounds (standardized):
```

```
                                                                        2.53 2.61 2.57 2.47 2.43 2.38
Bounds entered.
      Time                 Bounds
   0.23            -2.5300     2.5300
   0.33            -2.6100     2.6100
   0.44            -2.5700     2.5700
   0.58            -2.4700     2.4700
   0.71            -2.4300     2.4300
   0.83            -2.3800     2.3800
Enter the standardized statistic at the last analysis:
                                                                                    2.82

  Last value:                  2.8200
Enter confidence level (>0 and <1):
                                                                                     .95
95. percent confidence interval

Starting computation for lower limit . . .
Lower limit computed, starting on upper limit . . .


95. percent confidence interval: ( 0.1881, 4.9347)
Drift is equal to the standard treatment difference times the square
root of total information per arm.

Done.
```

Translation of the standardized parameter back to an estimate of the difference between treatment groups is done in Section 3.2

Acknowledgements

The authors wish to acknowledge of Kris Erlandson and Bill Ladd for assistance in constructing examples, and Wen Wei for assistance in programming.

References

[1] Armitage, P., McPherson, C. K. & Rowe, B. C. (1969), 'Repeated significance tests on accumulating data', *Journal of the Royal Statistical Society, Series A* **132**, 235–244.

[2] Beta-Blocker Heart Attack Trial Research Group (1982), 'A randomized trial of propranolol in patients with acute myocardial infarction. I, Mortality results.', *Journal of the American Medical Association* **246**, 1707–1714.

[3] DeMets, D. L., & Lan, K. K. G. (1984), 'An overview of sequential methods and their applications in clinical trials', *Communications in Statistics, Theory and Methods*, **13**, 2315–2338.

[4] Hwang, I. K., Shih, W. J. & deCani, J. S. (1990), 'Group sequential designs using a family of type I error probability spending functions', *Statistics in Medicine*, **9**, 1439–1445.

[5] Kim, K. & DeMets, D. L. (1987*a*), 'Design and analysis of group sequential tests based on the type I error spending rate function', *Biometrika* **74**, 149–154.

[6] Kim, K. & DeMets, D. L. (1987*b*), 'Confidence intervals following group sequential tests in clinical trials', *Biometrics* **43**, 857–864.

[7] Kim, K. & DeMets, D. L. (1992), 'Sample size determination for group sequential clinical trials with immediate response', *Statistics in Medicine* **11**, 1391–1399.

[8] Lan, K. K. G. & DeMets, D. L. (1983), 'Discrete sequential boundaries for clinical trials', *Biometrika* **70**, 659–663.

[9] Lan, K. K. G. & DeMets, D. L. (1989), 'Group sequential procedures: calendar versus information time', *Statistics in Medicine* **8**, 1191–1198.

[10] Lan, K. K. G. and Zucker, D. M., (1993) 'Sequential monitoring of clinical trials: the role of information and Brownian motion', *Statistics in Medicine* **12**, 753–765.

[11] Lan, K. K. G., Reboussin, D. M. & DeMets, D. L. (1994), 'Information and information fractions for design and sequential monitoring of clinical trials', *Communications in Statistics, Part A—Theory and Methods* **23**, 403–420.

[12] McPherson, C. K. & Armitage, P. (1971), 'Repeated significance tests on accumulating data when the null hypothesis is not true', *Journal of the Royal Statistical Society, Series A* **134**, 15–25.

[13] O'Brien, P. C. & Fleming, T. R. (1979), 'A multiple testing procedure for clinical trials', *Biometrics* **35**, 549–556.

[14] Pocock, S. J. (1977), 'Group sequential methods in the design and analysis of clinical trials', *Biometrika* **64**, 191–199.

[15] Reboussin92a Reboussin, D. M., DeMets, D. L., Kim, K. & Lan, K. K. G. (1992), Programs for computing group sequential boundaries using the Lan-DeMets method, Technical Report 60, Department of Biostatistics, University of Wisconsin-Madison.

[16] Reboussin, D. M., Lan, K. K. G. & DeMets, D. L. (1992). Group sequential testing of longitudinal data. Technical Report 72, Department of Biostatistics, University of Wisconsin-Madison.

[17] Wu, M. C. & Lan, K. K. G., (1992), 'Sequential monitoring for comparison of changes in a response variable in clinical studies', *Biometrics* **48**, 765–779.

Appendix

Theory related to the computations.

Consider a Brownian motion process in continuous time, $W(t)$, $t \in [0,1]$, having unknown drift parameter θ, which may be inspected at times t_i, $i = 1, \dots, k$. We wish to test the hypothesis $H_0 : \theta \leq 0$ at each inspection time t_i and proceed only if the test fails to reject; that is, if $W(t_i)$ does not exceed some value, so that the sequential test rejects if $Z(t_i) = W(t_i)/\sqrt{t_i} > b_i$. Consider a sequence of boundaries, $b_1, \dots, b_k$ applied at times $t_1, \dots, t_k$. Let g denote the standard normal density function, $g(u) = (2\pi)^{-1/2} \exp(-u^2/2)$. The probability distribution for W at analysis i is determined recursively by $f_1(x) = g(x)$ and

$$f_i(x) = \int_{-\infty}^{b_{i-1}} g\left((x-u)/\sigma_i\right) \sigma_i^{-1} f_{i-1}(u)\, du$$

where σ_i^2 is the variance of $[W(t_i) - W(t_{i-1})]$, that is, $\sigma_i^2 = t_i - t_{i-1}$. Integrating f_i from $-\infty$ to $+\infty$ gives the probability that the trial continues past the $(i-1)^{th}$ analysis.

Computations at the first analysis involve only the standard normal density and distribution function, but for the second and beyond, numerical integration is necessary. By applying Fubini's theorem, we have the continuation probability at analysis i

$$\begin{aligned} P_i &= \int_{-\infty}^{b_i} \int_{-\infty}^{b_{i-1}} g\left((x-u)/\sigma_i\right) \sigma_i^{-1} f_{i-1}(u) du\, dx \\ &= \int_{-\infty}^{b_{i-1}} \int_{-\infty}^{b_i} g\left((x-u)/\sigma_i\right) \sigma_i^{-1} f_{i-1}(u) dx\, du \\ &= \int_{-\infty}^{b_{i-1}} \Phi\left((b_i-u)/\sigma_i\right) f_{i-1}(u) du. \end{aligned}$$

Note that only a single numerical integration is now required. This manipulation allows the use of simple, accurate approximations to the normal distribution function to be used for computing P_i. Extension of the above to two sided tests is straightforward: if a_i is the lower bound, it can be substituted for $-\infty$ in the above integrals.

Description of computations.

For the first analysis, which uses only the cumulative normal distribution, we have $W(\tau_1) \sim N(\theta\tau_1, \tau_1)$. The probability calculated for exceeding the first upper boundary is

$$
\begin{aligned}
P\left(\frac{W(\tau_1)}{\sqrt{\tau_1}} > b_1\right) &= P\left(W(\tau_1) > b_1\sqrt{\tau_1}\right) \\
&= P\left(\frac{W(\tau_1)-\theta\tau_1}{\sqrt{\tau_1}} > \frac{b_1\sqrt{\tau_1}-\theta\tau_1}{\sqrt{\tau_1}}\right) \\
&= P\left(Z > b_1 - \theta\sqrt{\tau_1}\right) \\
&= 1 - \Phi\left(b_1 - \theta\sqrt{\tau_1}\right).
\end{aligned}
$$

In the programs, given (a_i, b_i), $i > 1$, separate subroutines are called to compute the exit probability, denoted P_i and, if there are more analyses to come, to compute f_i. For the routine computing P_i, a grid of values of $f_{i-1}(u)$ for $u \in (a_{i-1}, b_{i-1})$, saved from the previous step, is needed. The grid size is standardized, so that it is finer when the increment has a smaller standard deviation. At each grid point u, the quantity $[\Phi((b_i-u)/\sigma_i)-\Phi((a_i-u)/\sigma_i)]\, f_{i-1}(u)$ is computed and stored in an array. This array is then passed to a numerical integration routine along with a_{i-1}, b_{i-1} and the grid size, and $1 - P_i$ is returned. The other subroutine computes f_i for a grid of values between a_i and b_i. For each grid point, the grid of values of f_{i-1} is needed. Letting u denote a point in the grid from a_{i-1} to b_{i-1}, and x denote a point in the grid from a_i to b_i, the quantity $f_{i-1}(u)\, g((u-x)/\sigma_i)/\sigma_i$ is computed and stored in an array. As before, this array is passed to a numerical integration routine, along with a_{i-1}, b_{i-1}, and the grid size, and $f_i(x)$ is obtained and stored for the next step. Currently, the numerical integration routine is a composite trapezoidal rule, which appears to produce fairly accurate results. Reboussin, DeMets, Kim & Lan (1992) present testing of the programs for computational accuracy and simulations results for validity. Their appendices contain listings of the code.

Programming for spending functions.

Boundaries and information fractions are related by the type I error spending function. The program contains five choices for these functions in a single subroutine called `alphas`. The critical source code is:

```
c     Calculate probabilities according to use function.
      do 50 i=1,nn
         if (iuse .eq. 1) then
            pe(i)=2.d0*
     .        (1.d0-pnorm(znorm(1.d0-(alpha/side)/2.d0)/dsqrt(t(i))))
```

```
        else if (iuse .eq. 2) then
           pe(i)=(alpha/side)*dlog(1.d0 + (e-1.d0)*t(i))
        else if (iuse .eq. 3) then
           pe(i)=(alpha/side)*t(i)
        else if (iuse .eq. 4) then
           pe(i)=(alpha/side)*(t(i) ** 1.5d0)
        else if (iuse .eq. 5) then
           pe(i)=(alpha/side)*(t(i) ** 2.0d0)
c     Add other spending function options here: e.g.
c       else if (iuse.eq.6) then . . .
        else
           write(6,*) ' Warning: invalid use function.'
        end if
```

Additional spending functions can be added as “silent” options by editing this section of code. For example, here is the code for a spending function which does not allow stopping until the trial is half over. Once half the information has accumulated, the type I error is spent uniformly until the end of the trial.

```
        else if (iuse .eq. 6) then
           if (t(i).le.0.0) then
              pe(i)=0.0d0
           else
              pe(i)=(alpha/side)*(t(i) * 2.0d0 - 1.d0)
           end if
```

This could also be added to the input routine with some additional programming effort.