



Published in final edited form as:

*J Stat Plan Inference*. 2013 February 1; 143(2): 283–295. doi:10.1016/j.jspi.2012.07.015.

## An efficient sequential design of clinical trials

Yi Cheng<sup>a,\*</sup> and Yu Shen<sup>b</sup>

<sup>a</sup>Department of Mathematical Sciences, Indiana University, South Bend, IN 46634

<sup>b</sup>Biostatistics Department, the M.D. Anderson Cancer Center, Houston, TX

### Abstract

We propose an efficient group sequential monitoring rule for clinical trials. At each interim analysis both efficacy and futility are evaluated through a specified loss structure together with the predicted power. The proposed design is robust to a wide range of priors, and achieves the specified power with a saving of sample size compared to existing adaptive designs. A method is also proposed to obtain a reduced-bias estimator of treatment difference for the proposed design. The new approaches hold great potential for efficiently selecting a more effective treatment in comparative trials. Operating characteristics are evaluated and compared with other group sequential designs in empirical studies. An example is provided to illustrate the application of the method.

### Keywords

Decision theory; Group sequential clinical trial design; Loss function; Predicted power; Reduced-bias estimator

## 1. Introduction

With large numbers of experimental agents in the developing process, there is an increasing need for innovative clinical trial designs that allow effective agents to be identified more efficiently compared with the standard designs. Group sequential designs have the flexibility to use cumulated information to modify the subsequent course of the trial. Monitoring a group sequential clinical trial is a dynamic process. The time of performing an interim analysis in a group sequential trial is a natural time to update and combine the information from prior experience and accumulated data using Bayesian methods. The decision-theoretic approach has the potential ability to concurrently consider efficacy, futility and cost (Berry, 1994). Although there is no explicit specification for the type I and II error rates when determining the maximum sample size, the decisions leading to false-positive and false-negative conclusions may be taken into account through a loss structure to satisfy regulatory settings (Berry and Ho, 1988; and Lewis and Berry, 1994).

In contrast to previous work in the area, Cheng and Shen (2005) proposed an adaptive design using the Bayesian decision-theoretic approach, which allowed the maximum sample size to be random. The design keeps the virtue of the self-designing trial in the frequentist framework (Fisher, 1998; Shen and Fisher, 1999). An algebraic expression that connects the overall type I error rate and the parameters of the loss function was established, so that the desired frequentist properties can be satisfied for regulatory settings. Later, the possibility of inefficient performance by the adaptive designs was raised by Jennison and Turnbull (2006).

---

\*Corresponding author. Tel.: +1 574 520 4519. Fax: +1 574 520 5589., ycheng@iusb.edu (Yi Cheng).

We recognized the limitations in terms of efficiency in the work of Cheng and Shen (2005), similar to the design of Shen and Fisher (1999). The realized type I error rate is often smaller than the specified significance level. Moreover, the procedures that are based on the non-sufficient statistics may increase the type II error rate with or without the decreased type I error rate. The monitoring procedures for continuing or stopping the trial based on the loss function alone are often not sensitive enough to fully reflect the evidence of operation characteristics at interim analysis. This issue motivated us to search for a more efficient group sequential design within this framework. Our goal of efficient clinical trial design is to minimize the overall sample size while maintaining the specified error rates under regulatory settings (Jennison & Turnbull, 2000, Ch.10). By combining the strengths of the Bayesian approach in establishing a procedure and the frequentist approach in evaluating the procedure, we propose a design that assesses the predicted power and the expected loss at each interim analysis and makes decisions based on these two factors. The maximum sample size is sequentially determined using interim data. The overall type I error rate can be controlled by choosing related designing parameters.

Similar to other sequential clinical trials, the standard estimator of the treatment difference, such as sample mean or posterior mean in this setting, is often biased. Methods in the literature for constructing unbiased or nearly unbiased estimators are not directly applicable to the estimator of parameter following the proposed design because of the random maximum sample size. In this article, we develop a new procedure of estimation aimed at minimizing the bias. The proposed estimation procedure is general and may be applied to other group sequential designs.

The remainder of this article is organized as follows. In Section 2 we present the proposed design, decision rules and the related properties. In Section 3 we derive the critical region and evaluate the frequentist properties with normal responses. We describe a new estimation procedure for the proposed design in Section 4. We conduct a simulation study and compare our inference with the existing adaptive and classical group sequential designs in Section 5. We apply the proposed design to an example in Section 6 and conclude in Section 7.

## 2. New design and main results

Consider a clinical trial that compares a new treatment T with a control C, where the individual treatment response is  $X_T$  and the individual control response is  $X_C$ . For notation simplicity, we assume a design with equal randomization, which can be easily extended to more general two-arm trials. The block size at each stage is  $2B_i$ ,  $i = 1, 2, \dots$ , where  $B_i$  for each treatment arm is fixed before the trial starts, but the maximum number of blocks is not predetermined and is determined using interim data.

Let  $\bar{X}_{T_i}$  and  $\bar{X}_{C_i}$  be the observed mean responses in the  $i$ th block. Given the parameter of interest,  $\theta$ , which is a measure of the treatment difference, let

$$X_i = \bar{X}_{T_i} - \bar{X}_{C_i} \sim F_i(\cdot|\theta), \quad \text{and} \quad \theta = \int_{-\infty}^{\infty} x dF_i(x|\theta).$$

The one-sided hypothesis to be tested is

$$H_0: \theta \leq 0 \quad \text{versus} \quad H_1: \theta > 0.$$

Within a Bayesian framework, assume that  $\theta$  has a prior distribution  $\pi$  with  $E(\theta|\pi) = \delta$ .

### 2.1. Decision Rule

The loss structures may be devised from the perspective of the investigators, patients, and/or stockholders via the data monitoring committee (Gittins & Pezeshk, 2000). It has become a common practice to include patient advocates on data and safety monitoring committees for the conduct of clinical trials. We define a loss structure consisting of the respective cost of making a false-positive conclusion, a false-negative conclusion, and the cost of the total sample, in which the risk and benefit of the new agent under investigation should be balanced via these cost parameters.

Let  $D$  be the decision of either accepting (A) or rejecting (R) null hypothesis. At the  $j$ th interim analysis, the loss function is defined by

$$L(\theta, D, j) = L(\theta, D) + 2K_2 \sum_{i=1}^j B_i, \quad (2.1)$$

where

$$L(\theta, D) = K_0 h_0(\theta) I\{\theta \leq 0, D=R\} + K_1 h_1(\theta) I\{\theta > 0, D=A\},$$

$K_0$  is a constant and positive penalty for each unit of  $h_0(\theta)$  when  $H_0$  is incorrectly rejected, and  $K_1$  is a positive penalty for each unit of  $h_1(\theta)$  when  $H_0$  is incorrectly accepted. Here,  $h_0(\cdot)$  and  $h_1(\cdot)$  are positive and continuous functions, and the positive loss may change with the actual value of  $\theta$ . One sensible choice is to allow the loss function to change with the magnitude of  $\theta$ , such as  $h_0(\theta) = h_1(\theta) = |\theta|^w + c$  with  $c \geq 0$  and  $w > 0$ . In this case, the loss to accept the null hypothesis increases as the distance of a positive  $\theta$  to zero gets larger. Similarly, the loss to reject  $H_0$  increases as a negative  $\theta$  deviates from zero further. The unit cost of each sample,  $K_2$ , is relative to  $K_0$  and  $K_1$ , which should be much smaller than  $K_0$  and  $K_1$ . Another special case for the type of loss functions is the commonly used 0- $K_j$  loss, with  $h_0(\theta) = h_1(\theta) = 1$ . Following a common convention that making a false-positive conclusion leads to a more severe penalty than making a false-negative conclusion, we let  $K_0 > K_1$ .

Let  $\mathcal{X}_j = \{X_1, \dots, X_j\}$  be the accumulated data up to step  $j$ , where  $X_j$  represents data from the  $j$ th block. At each interim analysis the study will be terminated for futility or for efficacy if the evidence is strong enough. Otherwise, the study will continue to  $(j + 1)$ th block. The loss function related to the above two decisions at the  $j$ th interim analysis is defined as the expected loss of stopping the trial and accepting the null hypothesis, that is,

$$L_A(\mathcal{X}_j) = E \{L(\theta, A, j) | \mathcal{X}_j\},$$

and the predicted loss of continuing the trial to the  $(j + 1)$ th block,

$$L_{cont}(\mathcal{X}_j) = E \left[ \min_{D=A \text{ or } R} E\{L(\theta, D, j+1) | \mathcal{X}_{j+1}\} | \mathcal{X}_j \right]. \quad (2.2)$$

To better achieve the targeted power while minimizing the total sample size, we incorporate the predicted power into the monitoring rules. We consider this type of the group sequential designs motivated by internal reasons in view of the interim data on the primary endpoint.

As discussed by Jennison & Turnbull (2006), it is proper to increase the remaining sample size if continuing as planned would give low conditional power under  $\theta = \delta$ ; an interim estimate of  $\hat{\theta}$  still represents a worthwhile clinical improvement. In this case, extending the study could ensure the desired power under the estimated effect size. Our goal is to achieve a desired power at an anticipated effective size with the flexibility that the desired power may be preserved at the true effective size ( $\theta$ ) when the true effective size deviates from the anticipated size within a certain extension. Of course, there is a realistic limitation to such a deviation. In our numerical study (Section 5), the values of  $\theta$  are mostly limited to the set  $\{\theta: |\theta - \delta| < 0.5\delta\}$ . Since the proposed design is intended to achieve the desired power at  $\theta$ , the true efficacy value in a given range, it is natural to use the current posterior mean as an estimator of  $\theta$  to evaluate the conditional power.

Given data observed up to the  $j$ th block, letting  $\delta_j$  be the posterior mean of  $\theta$  given  $x_j$ , the predicted power related to the critical region  $R_{j+1}$ , is defined as  $P(R_{j+1}|\theta = \delta_j, x_j)$ , where

$$R_{j+1} = \{ \mathcal{X}_{j+1} : E\{L(\theta, A, j+1)|\mathcal{X}_{j+1}\} > E\{L(\theta, R, j+1)|\mathcal{X}_{j+1}\} \}. \quad (2.3)$$

**Definition 2.1**—The proposed design follows a two-step decision rule described below, where  $M$  is the total number of interim analysis and the procedure starts from  $j = 1$ .

1. If  $L_A(x_j) \leq L_{con}(x_j)$ , terminate the trial and accept  $H_0$  at the  $j$ th step, let  $M = j$ .
2. If  $L_A(x_j) > L_{con}(x_j)$ , evaluate the predicted power.
  - a. If the predicted power  $P(R_{j+1}|\theta = \delta_j, x_j) \leq 1 - \beta$ , terminate the trial, let  $M = j$ . If  $E\{L(\theta, A, M)|x_M\} > E\{L(\theta, R, M)|x_M\}$ , reject  $H_0$ , otherwise accept  $H_0$ .
  - b. If  $P(R_{j+1}|\theta = \delta_j, x_j) > 1 - \beta$ , continue the trial to the  $j + 1$ th analysis and repeat 1 and 2.

**Remark 2.1**—In our earlier work (Cheng and Shen, 2005), the decision at block  $j$  to stop the trial and reject null hypothesis for efficacy is only based on  $L_R(x_j) = E\{L(\theta, R, j)|x_j\}$ , if it is no greater than  $L_{con}(x_j)$ . The design parameters, such as penalties for making wrong decisions in the typical decision-theoretic approaches are often searched empirically to match with the specified error rates by trial and error. As a result, the design may not be efficient regarding the sample size. One advantage of using the predicted power directly in the course of procedure is the ability to stop the trial whenever evidence favoring the experimental treatment is sufficient. Otherwise, we continue the trial to gain more information about the treatments. Further comparison is presented in the numerical study section of this paper.

## 2.2. Properties

Some practical concerns need to be addressed, specifically, whether the trial will be terminated with a finite number of interim analysis and whether the decision rules lead to a correct recommendation with the specified probabilities. The following Theorems 2.1 and 2.2 confirm that these desirable properties hold asymptotically with the proposed design. The proofs of the theorems are given in Appendix A.

**Theorem 2.1**—Let  $M$  be the total number of blocks following the proposed decision procedures. Assuming that the prior density of  $\theta$ ,  $\pi(\theta)$ , satisfies certain regularity

conditions,  $h_0(\theta)$  and  $h_1(\theta)$  are positive and continuous functions of  $\theta$ , there exists a positive  $w$  such that

$$\lim_{\theta \rightarrow -\infty} h_0(\theta)/(1+|\theta|^w) = 0, \quad \lim_{\theta \rightarrow \infty} h_1(\theta)/(1+|\theta|^w) = 0$$

almost surely, and that  $K_0$ ,  $K_1$  and  $K_2$  are positive, then

$$P_\theta(M < +\infty) = 1$$

if  $\theta = 0$ ; When  $\theta = 0$ ,  $P_0(M < +\infty) = 1$  if  $K_1 h_1(0) < 2BK_2$ , where  $2B$  is the minimum block size.

Theorem 2.1 shows that with probability 1 the trial will be terminated with a finite number of interim analysis. The regularity conditions are specified by Schervish (1995, p. 429).

**Theorem 2.2**—Under the conditions of Theorem 2.1, let  $\theta^*$  denote the true treatment difference, then

$$\lim_{j \rightarrow \infty} [E\{L(\theta, A, j) | \mathcal{X}_j\} - E\{L(\theta, R, j) | \mathcal{X}_j\}] = \begin{cases} -K_0 h_0(\theta^*), & \text{if } \theta^* < 0; \\ K_1 h_1(\theta^*), & \text{if } \theta^* > 0 \end{cases} \quad (2.4)$$

almost surely. When  $\theta^* > 0$ ,  $\lim_{j \rightarrow \infty} P_{\theta^*}(R_{j+1} | \theta = \delta_j, \mathcal{X}_j) = 1$  almost surely. When  $\theta^* = 0$ ,

$$\lim_{j \rightarrow \infty} P_0 [E\{L(\theta, A, j) | \mathcal{X}_j\} - E\{L(\theta, R, j) | \mathcal{X}_j\} < 0] = \frac{K_0 h_0(0)}{K_1 h_1(0) + K_0 h_0(0)}.$$

The condition on  $K_2 > 0$ , i.e., a positive cost for enrolling each patient, is critical to avoid unnecessary extension of the trial under  $H_0$ . Theorem 2.2 ensures that with the proposed monitoring strategies, it is expected that  $E\{L(\theta, A) | \mathcal{X}_j\} > E\{L(\theta, R) | \mathcal{X}_j\}$  (since  $K_1 h_1(\theta^*) > 0$ ) and the predicted power  $P_{\theta^*}(R_{j+1} | \theta = \delta_j, \mathcal{X}_j)$  will be near 1 asymptotically under the alternative hypothesis. When the true treatment difference  $\theta^* < 0$ , we should expect that  $E\{L(\theta, A) | \mathcal{X}_j\} < E\{L(\theta, R) | \mathcal{X}_j\}$  (since  $-K_0 h_0(\theta^*) < 0$ ) and the predicted power remains small, so that the trial will conclude futility with enough data. When  $\theta^* = 0$ , where zero is the boundary value for the two hypotheses, the probability to conclude futility approximates to  $K_0 h_0(0) / (K_1 h_1(0) + K_0 h_0(0))$ , which is connected to the type I error rate. For a special case in which  $h_0(0) = h_1(0)$ , this limit becomes  $K_0 / (K_0 + K_1)$ . We will give more details of this matter in Section 3.2.

### 2.3. Computation

There are three key elements in the decision rule,  $L_A(\mathcal{X}_j)$ ,  $L_{con}(\mathcal{X}_j)$ , and the predicted power  $P(R_{j+1} | \theta = \delta_j, \mathcal{X}_j)$ . We derive the computational formulation for each as follows. Given data observed up to the  $j$ th block and loss function defined in (2.1), the expected loss of accepting  $H_0$  can be expressed as

$$L_A(\mathcal{X}_j) = E\{L(\theta, A, j) | \mathcal{X}_j\} = 2K_2 \sum_{i=1}^j B_i + K_1 \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_j) d\theta, \quad (2.5)$$

where  $\pi(\cdot | \mathcal{X}_j)$  is the posterior density function of  $\theta$  given  $\mathcal{X}_j$ . Similarly,

$$L_R(\mathcal{X}_j) = E\{L(\theta, R, j) | \mathcal{X}_j\} = 2K_2 \sum_{i=1}^j B_i + K_0 \int_{-\infty}^0 h_0(\theta) \pi(\theta | \mathcal{X}_j) d\theta. \quad (2.6)$$

Expressions (2.5) and (2.6) provide the computational formulation for the two components,  $E\{L(\theta, A, j+1) | \mathcal{X}_{j+1}\}$  and  $E\{L(\theta, R, j+1) | \mathcal{X}_{j+1}\}$ , of  $L_{cont}(\mathcal{X}_j)$  defined in (2.2). The outside expectation on the right-hand side of (2.2) is determined with respect to the conditional distribution of  $X_{j+1}$  given  $\mathcal{X}_j$ . When  $F(\cdot | \theta)$  has a density function  $f(\cdot | \theta)$ , the conditional density of  $X_{j+1}$  given  $\mathcal{X}_j$  can be expressed as:

$$g(x | \mathcal{X}_j) = \int_{-\infty}^\infty \pi(\theta | \mathcal{X}_j) F(x | d\theta) = \int_{-\infty}^\infty f(x | \theta) \pi(\theta | \mathcal{X}_j) d\theta.$$

The last element is the predicted power  $P(R_{j+1} | \theta = \delta_j, \mathcal{X}_j)$ , where  $R_{j+1}$  is defined in (2.3), and can be expressed as

$$R_{j+1} = \left\{ \mathcal{X}_{j+1} : \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_{j+1}) d\theta > \frac{K_0}{K_1} \int_{-\infty}^0 h_0(\theta) \pi(\theta | \mathcal{X}_{j+1}) d\theta \right\}, \quad (2.7)$$

for  $j = 0, 1, \dots, M - 1$ .

### 3. Applications with normal responses

To illustrate the design procedures in application, we assume that the outcomes follow normal distributions. For simplicity of notation, let

$$\begin{aligned} \bar{X}_{T_i} &\sim N(\mu_T, \sigma^2 / (2B_i)), & \bar{X}_{C_i} &\sim N(\mu_C, \sigma^2 / (2B_i)), \\ X_i = \bar{X}_{T_i} - \bar{X}_{C_i} &\sim N(\theta, \sigma^2 / B_i), \end{aligned}$$

where  $\sigma^2$  is known. (It is not conceptually different with an unknown variance.) The prior distribution for  $\theta$  is assumed to be  $N(\delta, \sigma^2 / B_0)$ , where  $B_0$  can be interpreted as the equivalent sample size on which the prior information is based (Spiegelhalter et al, 1994). Typically,  $B_0$  is much smaller than  $B_j$  so that the prior information plays a minor role in the decision process. We let  $X_0 = \delta$  to keep the notation for prior information coherent with that from the interim analysis. After data from the first  $j$  blocks are observed, letting  $n_j = \sum_{i=0}^j B_i$ , the posterior distribution of  $\theta$  is  $N(\delta_j, s_j^2)$ , where

$$\delta_j = \frac{\sum_{i=0}^j B_i X_i}{n_j} \text{ and } s_j^2 = \frac{\sigma^2}{n_j}$$

are the posterior mean and variance, respectively.

Consider the following function in the loss structure:  $h_0(\theta) = h_1(\theta) = h(\theta) = |\theta| + c$ , where  $c$  is a nonnegative constant. This function increases linearly with the distance of  $\theta$  from zero, when an incorrect action is taken. It satisfies the conditions of Theorems 2.1 and 2.2 when  $cK_1 < 2BK_2$ .

### 3.1. Expected losses and predicted power

For  $h(\theta) = |\theta| + c$  and normal responses, we provide explicit expressions for the three key elements,  $L_A(x_j)$ ,  $L_{con}(x_j)$  and the predicted power  $P(R_{j+1}|\theta = \delta_j, x_j)$ , in the decision rule through the following three propositions, respectively.

**Proposition 3.1**—Let  $h_0(\theta) = h_1(\theta) = |\theta| + c$ . We have

$$L_A(\mathcal{X}_j) = 2K_2 \sum_{i=1}^j B_i + K_1 \left\{ \delta_j \Phi(\delta_j/s_j) + s_j \varphi(\delta_j/s_j) + c \Phi(\delta_j/s_j) \right\}, \quad (3.1)$$

$$L_R(\mathcal{X}_j) = 2K_2 \sum_{i=1}^j B_i + K_0 \left\{ -\delta_j \Phi(-\delta_j/s_j) + s_j \varphi(\delta_j/s_j) + c \Phi(-\delta_j/s_j) \right\}. \quad (3.2)$$

where  $\varphi(\cdot)$  and  $\Phi(\cdot)$  are the standard normal probability density function and cumulative distribution function, respectively.

**Proof**—Since the posterior distribution of  $\theta$  given  $x_j$  follows  $N(\delta_j, s_j^2)$ , it can be verified that

$$\int_0^\infty \theta \pi(\theta|\mathcal{X}_j) d\theta = \delta_j \Phi(\delta_j/s_j) + s_j \varphi(\delta_j/s_j), \quad (3.3)$$

$$\int_0^\infty \pi(\theta|\mathcal{X}_j) d\theta = \Phi(\delta_j/s_j).$$

Applying the result of (3.3) to the right-hand side of (2.5). The expression of (3.1) is obtained. By a similar argument, noting that  $1 - \Phi(x) = \Phi(-x)$ , expression (3.2) can be obtained as well.

The computational formula for  $L_{con}(x_j)$  defined by (2.2) is derive by the following proposition.

**Proposition 3.2**—Let  $h_0(\theta) = h_1(\theta) = |\theta| + c$ . Then

$$L_{con}(\mathcal{X}_j) = 2K_2 \sum_{i=1}^{j+1} B_i + \int_{-\infty}^\infty \min\{l_A(x|\delta_j), l_R(x|\delta_j)\} \frac{1}{\sigma_{j+1}} \varphi\left(\frac{x-\delta_j}{\sigma_{j+1}}\right) dx, \quad (3.4)$$

where  $\sigma_{j+1}^2 = s_j^2 + \sigma^2 / B_{j+1}$ .

**Proof**—Let  $I_D(x|\delta_j)$  define the integral term of  $L_D(x_{j+1})$  given by the right-hand side of (2.5) and (2.6), where  $D = A$  or  $R$ . Substituting  $j$  by  $j + 1$  in equations (3.1) and (3.2), the second term on the right-hand side of each equation gives, respectively, the expression of  $I_A(x|\delta_j)$  and  $I_R(x|\delta_j)$  below:

$$\begin{aligned} l_A(x|\delta_j) &= K_1 \left\{ (\delta_{j+1} + c) \Phi \left( \frac{\delta_{j+1}}{s_{j+1}} \right) + s_{j+1} \varphi \left( \frac{\delta_{j+1}}{s_{j+1}} \right) \right\}, \\ l_R(x|\delta_j) &= K_0 \left\{ (-\delta_{j+1} + c) \Phi \left( -\frac{\delta_{j+1}}{s_{j+1}} \right) + s_{j+1} \varphi \left( \frac{\delta_{j+1}}{s_{j+1}} \right) \right\}. \end{aligned} \tag{3.5}$$

Given  $\mathcal{X}_j$  and  $X_{j+1} = x$ , the posterior mean  $\delta_{j+1}$  satisfies the following recursive equation in terms of  $\delta_j$  and  $x$ :

$$\delta_{j+1} = \frac{B_{j+1}s_{j+1}^2}{\sigma^2}x + \frac{s_{j+1}^2}{s_j^2}\delta_j.$$

It is known that the conditional distribution of  $X_{j+1}$  given  $\mathcal{X}_j$  is also normal, specifically,

$$X_{j+1}|\mathcal{X}_j \sim N(\delta_j, s_j^2 + \sigma^2/B_{j+1}).$$

Thus, we have (3.4) since  $\sigma_{j+1}^2 = s_j^2 + \sigma^2/B_{j+1}$ .

The next proposition provides a recursive expression for the estimated probability of the rejection region.

**Proposition 3.3**—Let  $h_0(\theta) = h_1(\theta) = |\theta| + c$ . Then

$$P(R_{j+1}|\theta = \delta_j, \mathcal{X}_j) = \Phi \left( \frac{\delta_j \sigma - \xi_{j+1} \sigma s_{j+1}}{s_{j+1}^2 \sqrt{B_{j+1}}} \right), \tag{3.6}$$

where  $\xi_j$  is the unique solution of the following equation in variable  $x$ :

$$(1-2r) \{s_j x \Phi(x) + s_j \varphi(x)\} + c \Phi(x) + r s_j x = cr$$

and  $r = K_0/(K_0 + K_1)$ .

**Proof**—In view of (3.3) and that  $h(\theta) = |\theta| + c$ , the critical region  $R_j$  defined by (2.7) can be expressed as

$$\begin{aligned} R_j &= \left\{ \mathcal{X}_j: \int_0^\infty h(\theta) \pi(\theta|\mathcal{X}_j) d\theta > \frac{K_0}{K_1} \left\{ E(h(\theta)|\mathcal{X}_j) - \int_0^\infty h(\theta) \pi(\theta|\mathcal{X}_j) d\theta \right\} \right\} \\ &= \left\{ \mathcal{X}_j: \int_0^\infty (\theta + c) \pi(\theta|\mathcal{X}_j) d\theta \geq r \left\{ -\delta_j + 2 \int_0^\infty \theta \pi(\theta|\mathcal{X}_j) d\theta + c \right\} \right\} \\ &= \left\{ \mathcal{X}_j: (1-2r) \left\{ \delta_j \Phi \left( \frac{\delta_j}{s_j} \right) + s_j \varphi \left( \frac{\delta_j}{s_j} \right) \right\} + c \Phi \left( \frac{\delta_j}{s_j} \right) + r \delta_j \geq cr \right\}, \end{aligned} \tag{3.7}$$

where  $r = K_0/(K_0 + K_1)$ . For given constants  $c > 0$ ,  $s > 0$ , and  $0 < r < 1$ , it can be verified that the continuous function

$$u(x, s, r) = (1-2r) \{s x \Phi(x) + s \varphi(x)\} + c \Phi(x) + r s x \tag{3.8}$$



increases in  $x$ ,  $\lim_{x \rightarrow -\infty} u(x, s, r) = -\infty$ , and  $\lim_{x \rightarrow +\infty} u(x, s, r) = +\infty$ . By the intermediate value theorem, there exists a unique real number  $\xi_j$  such that

$$R_j = \{\mathcal{X}_j: u(\delta_j/s_j, s_j, r) > rc\} = \{\mathcal{X}_j: \delta_j/s_j > \xi_j\}, \quad (3.9)$$

where  $\xi_j$  is obtained by solving equation  $u(x, s_j, r) = rc$  for  $x$ . It is easy to see that

$$X_{j+1} | \theta = \delta_j, \mathcal{X}_j \sim N\left(\delta_j, \frac{\sigma^2}{B_{j+1}}\right). \quad (3.10)$$

The posterior mean satisfies the following recursive equation:

$$\delta_{j+1} = \frac{B_{j+1}s_{j+1}^2}{\sigma^2} X_{j+1} + \frac{s_{j+1}^2}{s_j^2} \delta_j.$$

In addition to the fact that  $\sigma^2/s_j^2 + B_{j+1} = \sigma^2/s_{j+1}^2$ , we have

$$\frac{\delta_{j+1}}{s_{j+1}} = \frac{s_{j+1}\sqrt{B_{j+1}}}{\sigma} \left( \frac{X_{j+1} - \delta_j}{\sigma/\sqrt{B_{j+1}}} \right) + \frac{\delta_j}{s_{j+1}}. \quad (3.11)$$

Following (3.9) through (3.11), the predicted power,  $P(R_{j+1} | \theta = \delta_j, \mathcal{X}_j)$ , after observing  $j$  blocks of data can be expressed as

$$P\left(\frac{\delta_{j+1}}{s_{j+1}} > \xi_{j+1} | \theta = \delta_j, \mathcal{X}_j\right) = \Phi\left(\frac{\delta_j \sigma - \xi_{j+1} \sigma s_{j+1}}{s_{j+1}^2 \sqrt{B_{j+1}}}\right).$$

Hence, (3.6), the following expression for the predicted power, is derived.

The expressions (3.1), (3.4), and (3.6) are computationally straightforward and hence the decision procedures are user-friendly to carry out.

### 3.2. Connection to type I error rate

Our goal in this section is to find a connection between the type I error rate and  $K_0/K_1$  in the loss function, so that the operating characteristic can be directly assessed and ensured.

Recall that  $K_0/(K_0 + K_1) = r$ , equivalently,

$$K_0/K_1 = r/(1-r). \quad (3.12)$$

The following two terms are defined for the preparation of Proposition 3.4 and its proof.

Let

$$\xi = \begin{cases} \sqrt{Z_{\alpha/2} + n_0 \delta^2 / \sigma^2}, & \text{if } n_1 \leq \sigma^2 Z_{\alpha/2} / \delta^2 + n_0; \\ Z_{\alpha/2} \sqrt{1 - n_0 / n_1} + \frac{n_0 \delta}{\sigma \sqrt{n_1}}, & \text{if } n_1 > \sigma^2 Z_{\alpha/2}^2 / \delta^2 + n_0 \end{cases} \quad (3.13)$$

and

$$r = \begin{cases} \Phi(\xi), & \text{if } \Phi(\xi) > v(\xi); \\ \frac{s_1 \{ \xi \Phi(\xi) + \varphi(\xi) \} + c \Phi(\xi)}{s_1 \{ 2\xi \Phi(\xi) + 2\varphi(\xi) - \xi \} + c}, & \text{if } \Phi(\xi) \leq v(\xi), \end{cases} \quad (3.14)$$

where

$$v(\xi) = \frac{\xi \Phi(\xi) + \varphi(\xi)}{2\xi \Phi(\xi) + 2\varphi(\xi) - \xi}. \quad (3.15)$$

**Proposition 3.4**—Let  $h_0(\theta) = h_1(\theta) = |\theta| + c$ . For  $K_0 = K_1$  defined by (3.13) through (3.12), the probability of having a false positive conclusion is controlled under the nominal significance level  $\alpha$ . Moreover, the probability to conclude futility at  $\theta = 0$  in Theorem 2.2 with  $h_0(0) = h_1(0)$  is  $K_0 = (K_0 + K_1)$ , which is greater than or equal to  $1 - \alpha$ .

**Proof**—In light of (3.7), the probability to reject the null hypothesis can be expressed as

$$P(R_j | \theta = 0) = P\{u(\delta_j / s_j, s_j, r) > rc | \theta = 0\}.$$

Since  $u(x, s, r)$  defined by (3.8) is increasing in  $x$  and linear in  $s$ , noting that  $0 < s_1, u(\delta_j / s_j, s_j, r) > rc$  leads to

$$\max\{u(\delta_j / s_j, s_1, r), u(\delta_j / s_j, 0, r)\} > rc, \quad (3.16)$$

we have

$$P(R_j | \theta = 0) \leq P\left(\frac{\delta_j}{s_j} > \xi | \theta = 0\right) = \Phi\left\{\frac{n_0 \delta - \xi \sigma \sqrt{n_j}}{\sigma \sqrt{(n_j - n_0)}}\right\}, \quad (3.17)$$

where  $\xi = \min\{u^{-1}(rc, s_j, r), u^{-1}(rc, 0, r)\}$ . The last equation holds due to the fact that

$$\frac{\delta_j}{s_j} | \theta = 0 \sim N\left(\frac{n_0 \delta}{\sigma \sqrt{n_j}}, \frac{n_j - n_0}{n_j}\right).$$

For a given upper boundary (e.g.  $\alpha$ ) to the probability in the right-hand side of (3.17), the corresponding quantile  $\xi$  can be derived. Once  $\xi$  is obtained, we can calculate  $r$  and then the ratio of  $K_0$  and  $K_1$  subsequently.

Because efficacy is assessed at each interim analysis, we shall use  $\alpha(j)$  at the  $j$ th interim in order to control the overall type I error rate at  $\alpha$  (Slud and Wei, 1982). Let

$$\alpha(0)=0, \quad \alpha(j)=\sum_{i=1}^j \frac{1}{2^i} \alpha, \quad \text{for } j=1, 2, 3, \dots,$$

where  $\alpha(j)$  monotonically increases in  $j$  and  $\lim_{j \rightarrow \infty} \alpha(j) = \alpha$  (i.e.  $\alpha(M) = \alpha$ ). Use the conservative upper boundary  $\alpha(1)$  for all  $j \geq 1$  in the above inequality,

$$\sup_{n_j: j \geq 1} \Phi \left\{ \frac{n_0 \delta - \xi \sigma \sqrt{n_j}}{\sigma \sqrt{(n_j - n_0)}} \right\} \leq \alpha(1), \quad (3.18)$$

which leads to  $P(R_j | \theta = 0) = \alpha(1) < \alpha$  for all  $j$ .

We need to solve the inequality (3.18) for  $\xi$ . Following the argument of the first derivative, it can be shown that the component inside of  $\Phi$ , which is  $(n_0 \delta - \xi \sigma \sqrt{n_j}) / \{\sigma \sqrt{(n_j - n_0)}\}$ , as a function of  $n_j$ , increases when  $0 < n_j < (\xi \sigma / \delta)^2$  and decreases when  $(\xi \sigma / \delta)^2 < n_j < \infty$ . Therefore,

$$\sup_{j \geq 1} \Phi \left\{ \frac{n_0 \delta - \xi \sigma \sqrt{n_j}}{\sigma \sqrt{(n_j - n_0)}} \right\} \leq \begin{cases} \Phi \left\{ -\frac{\sqrt{(\xi^2 \sigma^2 - n_0 \delta^2)}}{\sigma} \right\}, & \text{if } n_1 \leq (\xi \sigma / \delta)^2; \\ \Phi \left\{ \frac{n_0 \delta - \xi \sigma \sqrt{n_1}}{\sigma \sqrt{(n_1 - n_0)}} \right\}, & \text{if } n_1 > (\xi \sigma / \delta)^2. \end{cases}$$

Letting each expression on the right-hand side of the last inequality be less than or equal to  $\alpha(j)$ , it yields

$$\xi \geq \begin{cases} \sqrt{Z_{\alpha(j)}^2 + n_0 \delta^2 / \sigma^2}, & \text{if } n_1 \leq \sigma^2 Z_{\alpha(j)}^2 / \delta^2 + n_0; \\ Z_{\alpha(j)} \sqrt{1 - n_0 / n_1} + \frac{n_0 \delta}{\sigma \sqrt{n_1}}, & \text{if } n_1 > \sigma^2 Z_{\alpha(j)}^2 / \delta^2 + n_0. \end{cases}$$

By using the conservative upper boundary  $\alpha(1) = \alpha/2$  instead of  $\alpha(j)$  in the above derivation, (3.13) is verified. Next, we find the algebraic expression for  $r$ . In light of (3.16) and (3.17),  $\xi$  and  $r$  satisfy the equation:

$$\max\{u(\xi, s_1, r), u(\xi, 0, r)\} = rc,$$

where  $u$  is defined by (3.8). For the scenario of  $\max\{u(\xi, s_1, r), u(\xi, 0, r)\} = u(\xi, 0, r) = c\Phi(\xi)$ ,  $c\Phi(\xi) = rc$  implies  $r = \Phi(\xi)$ . For the complementary scenario of  $\max\{u(\xi, s_1, r); u(\xi, 0, r)\} = u(\xi, s_1, r)$ , it occurs if and only if  $r = v(\xi)$ , where  $v(\xi)$  is defined by (3.15). Hence (3.14) can be verified after working out the algebra in solving  $u(\xi, s_1, r) = rc$  for  $r$ .

In general, the choice of constants  $K_j$  depends on the objectives of a trial. In a special case: when  $h_i(\theta) = 1$  for  $i=0, 1$ ,  $K_0$  and  $K_1$  are the costs of making type I and type II errors, respectively, relative to the cost of enrolling a patient,  $K_2$ .

#### 4. Estimation following the proposed design

After the inference, the estimation of  $\theta$  is often of interest. It is well known that the naive estimate of  $\theta$  (sample mean) following a classical group sequential design is biased. This

phenomenon also occurs with the posterior mean of  $\theta$  following the proposed design. Typically,  $\theta$  is overestimated for trials concluding efficacy and underestimated for trials concluding futility.

We take a different perspective and approach to all existing methods. Since  $\theta$  is the treatment difference, it is reasonable to assume that the domain of  $\theta$ , denoted as  $\mathcal{D}$ , is bounded and can be predetermined. Let  $\delta_M$  and  $s_M$  be the posterior mean and posterior standard deviation of  $\theta$  from the original data at the end of the trial that follows the proposed design. We perform a grid search to find a near unbiased estimator of  $\theta$  in  $\mathcal{D}$  based on Monte Carlo simulations.

Because the proposed efficient design (ED) is uniquely determined by a set of design parameters, we denote such a design as  $ED(\delta, K_0, K_1, K_2, B_0, B_1, B, \beta)$ . Let  $x_M$  be the original data collected under the specified design  $ED(\delta, K_0, K_1, K_2, B_0, B_1, B, \beta)$ . To perform a grid search, we divide  $\mathcal{D}$  into  $N$  subintervals with an equal length. The partition is fine enough, so that the midpoint  $\theta_n$  in the  $n$ th subinterval is representative for the  $n$ th interval and  $\theta_n, n = 1, \dots, N$ , should cover the whole  $\mathcal{D}$ .

**Definition 4.1**

The simulation based estimation procedure follows:

1. For each  $\theta_n$ , we simulate one data set from  $N(\theta_n, \sigma^2/B_i)$  by applying the same monitoring rules of  $ED(\delta, K_0, K_1, K_2, B_0, B_1, B, \beta)$  as in the original trial. At the end of the simulated trial, compute the posterior mean of  $\theta_n$ . To stabilize the estimation, we repeat the procedure for  $L$  times and denote the posterior mean from the  $l$ th simulation as  $\widehat{\delta}_{\theta_n}^{(l)}$  for each  $\theta_n$ . We denote the average of the posterior means from the  $L$  simulations as

$$\zeta_{MC}(\theta_n|L) = \frac{1}{L} \sum_{l=1}^L \widehat{\delta}_{\theta_n}^{(l)}.$$

2. When  $\zeta_{MC}(\theta_n|L), n = 1, \dots, N$ , are computed, the estimator  $\widehat{\theta}$  will be searched from  $\{\theta_1, \dots, \theta_N\}$  with the following criterion: among all candidates in  $\{\theta_1, \dots, \theta_N\}$ ,  $\zeta_{MC}(\widehat{\theta}|L)$  is the closest to  $\delta_M$ , where  $\delta_M$  is the posterior mean obtained from the observed data, i.e.,

$$\widehat{\theta} = \arg \min_{\theta_n \in \mathcal{D}} |\zeta_{MC}(\theta_n|L) - \delta_M| = \arg \min_{\theta_n \in \mathcal{D}} \{\zeta_{MC}(\theta_n|L) - \delta_M\}^2. \quad (4.1)$$

We reason that the  $\widehat{\theta}$  proposed in (4.1) is a reduced-bias estimator. Let  $g(\theta) = E_{\theta}(\delta_M)$ . For any given  $\theta_n$ , we have

$$E_{\theta_n} \{\zeta_{MC}(\theta_n|L)\} = \frac{1}{L} \sum_{l=1}^L E_{\theta_n}(\widehat{\delta}_{\theta_n}^{(l)}) = g(\theta_n).$$

It is realistic to assume that the  $Var_{\theta_n}(\widehat{\delta}_{\theta_n}^{(1)})$  is uniformly bounded when  $\theta_n$  takes a value in a predetermined bounded domain  $\mathcal{D}$ . Based on the above estimation procedure (1), we have

$$\text{Var}_{\theta_n} \{ \zeta_{MC}(\theta_n|L) \} = \frac{1}{L^2} \sum_{l=1}^L \text{Var}_{\theta_n} \{ \widehat{\delta}_{\theta_n}^{(l)} \} = O(1/L)$$

uniformly for  $\theta_n$  in  $\mathcal{D}$ . Moreover,

$$\begin{aligned} E_{\theta, \theta_n} \{ \zeta_{MC}(\theta_n|L) - \delta_M \}^2 &= E_{\theta, \theta_n} \{ \zeta_{MC}(\theta_n|L) - g(\theta_n) + g(\theta_n) - \delta_M \}^2 \\ &= \text{Var}_{\theta_n} \{ \zeta_{MC}(\theta_n|L) \} + \text{Var}_{\theta} \{ \delta_M \} + \{ g(\theta) - g(\theta_n) \}^2. \end{aligned} \tag{4.2}$$

The last equality follows from the facts that

$$\begin{aligned} E_{\theta, \theta_n} \{ \delta_M - g(\theta_n) \}^2 &= \text{Var}(\delta_M) + \{ g(\theta) - g(\theta_n) \}^2, \\ E_{\theta} [ E_{\theta_n} \{ \zeta_{MC}(\theta_n|L) - g(\theta_n) \} \{ g(\theta_n) - \delta_M \} | \delta_M ] &= 0. \end{aligned}$$

Since  $\text{Var}_{\theta}(\delta_M)$  is free of  $\theta_n$  and  $\text{Var}_{\theta_n} \{ \zeta_{MC}(\theta_n|L) \} = O(1/L)$ , for sufficiently large  $L$ , choosing  $\theta_n$  to minimize  $E_{\theta, \theta_n} \{ \zeta_{MC}(\theta_n|L) - \delta_M \}^2$ , the left-hand side of (4.2), is asymptotically equivalent to choosing  $\theta_n$  to minimize  $\{ g(\theta_n) - g(\theta) \}^2$ .

The above proposed estimation procedure has the flavor of a bootstrap procedure in the resampling part (Wang & Leung, 1997). The main difference between our procedure and their procedure is that we use a searching step to locate the estimator of  $\theta$  that is closest to the posterior estimator; whereas Wang and Leung (1997) used the maximum likelihood estimator.

### 5. Numerical studies

Using Monte Carlo simulations, we compare the performance of the proposed efficient design (ED) with existing group sequential designs in a similar setting. The OBF design (O'Brien & Fleming, 1979) is a classical group sequential design with a fixed maximum sample size. The self-designing trial of Shen & Fisher (1999) is a frequentist adaptive design, and the Bayesian adaptive design (Cheng & Shen, 2005) is based on the decision-theoretic approach. The ED, Bayes adaptive design and self-designing trials do not enforce a maximum sample size, in contrast to the O'Brien-Fleming (OBF) design, in which the maximum sample size is estimated given the design parameters:  $\delta$ ,  $\alpha$  and  $\beta$ .

It is worth noting that the original OBF design is two-sided. DeMets and Ware (1982) extended the results of the OBF design to the one-sided version of the symmetric OBF design. Our numerical study is based on the one-sided boundaries proposed by DeMets and Ware (1982).

For each scenario, the same type I and II error rates,  $\delta$ , and block sizes are used among the four designs under comparison. The block sizes are prespecified,  $B_0 = 1$ ,  $B_i = B = 6$ , for  $i = 2$ . For the OBF design, let  $B_1 = B$  for simplicity. For the remaining three designs,  $B_1 = 15$  in order to have a more stable estimator of  $\theta$  at the first step. The Bayesian adaptive design uses the loss structure with  $h_1(\theta) = h_0(\theta) = 1$ , while  $K_j$  is chosen to achieve the corresponding  $\alpha$  and  $\beta$ , respectively. The number of interim analysis of the OBF design is backward derived given the block size  $B$  and the fixed maximum sample size. The asymptotic boundary for the one-sided OBF test is obtained from DeMets & Ware (1982) and Shen & Fisher (1999) with the given design parameters.

The values of  $\theta$  are mostly limited to the set  $\{\theta: |\theta - \delta| < 0.5\delta\}$ . For a fair comparison among the methods, the true values of  $\theta$  are given as fixed when the  $X_j$ 's are generated from a normal distribution with  $\sigma^2 = 1$ . We used mean  $\theta = 0$  under the null hypothesis, and mean  $\theta = 0.5$  under the alternative hypothesis. The design parameters are specified as follows:  $\alpha = 0.025$ ,  $\beta = 0.1$ , and  $\delta = 0.4$  or  $0.7$ . Since the type I error rate in the proposed designs depends on  $K_0$  and  $K_1$  only through the ratio  $K_0/K_1$ , we let  $K_1 = 1$  and  $K_0 = r/(1 - r)$ , where  $r$  is computed from (3.14). Let  $h_0(\theta) = h_1(\theta) = |\theta| + c$  and  $c = BK_2$ , which satisfy the conditions of Theorems 2.1 and 2.2.

The empirical type I error rate, power and average sample number (ASN) of each design are summarized in Table 1 with 10,000 repetitions. The proposed efficient design maintains the nominal level at  $\alpha = 0.025$ , and achieves the specified power. Among the designs under comparison, the ED is the most robust one to different values of  $\delta$ . It can achieve the specified power even if  $\delta$  is considerably optimistically planned, compared with the true  $\theta$ . When  $\delta (= 0.7)$  overestimates  $\theta$ , the ED and Bayesian designs are the two designs that can still maintain the power at 0.9, but the ASN of ED is 17% lower than the ASN of the Bayesian design. When  $\delta$  is underestimated relative to the true  $\theta$  (at  $\delta = 0.4$ ), ASNs for ED are at least 25% lower than those for the other designs under the alternative.

While all sequential methods typically lead to savings in sample size compared with the standard fixed sample design, ED performs the best among the designs we investigated. As noted in Table 1, type I error rates of all designs under comparison are under control. The proposed design with assessment of predicted power at each interim analysis results in some further saving in the ASN while achieving the specified power.

The cost of each patient,  $K_2$ , relative to the cost of making a wrong conclusion is chosen to match the specified power. In general, the larger  $K_2$  is, the less power the trial has. We found that  $K_2$  should be in the magnitude of  $0.1^5$  of  $K_1$ , the cost of mistakenly concluding a negative trial. The results show that the proposed design is quite robust when  $K_2$  varies in the range of  $0.1^5$  to  $5(0.1)^5$  over different values of  $\theta$  and  $\delta$ . We performed following sensitivity analysis based on simulations (not shown in tables due to space limitation). The role of  $c$  is relatively minor other than that  $c$  should satisfy the inequality  $0 < cK_1 < 2K_2 \min\{B_j, j - 1\}$ , which is one of the conditions for Theorem 2.1. The performance of the proposed design is quite robust when  $c$  varies within its domain. When the block size is increased from  $B = 6$  to 16, we found that the performances were also similar for the two settings. The only difference is that average sample size increased by 10 under  $H_0$  and by 6 under  $H_1$  with slightly higher power. Figure 1 shows two histograms of the total number of blocks for the proposed design without a constraint on the maximum number of blocks. Our numerical studies also shows little difference in the error rates or the average sample size if the number of blocks is truncated at 15 for the scenarios shown in the histograms.

We also assess the bias of the proposed estimator and compare it with the posterior mean. Table 2 presents the proposed estimate and the naive posterior mean estimate for  $\theta$ , along with their empirical estimates of the standard error with  $\beta = 0.1$ ,  $\delta = 0.6$ , and  $\alpha = 0.025$  or  $.05$ . The true underlying  $\theta$  varies from 0 to 0.6. The grid search was conducted over the predetermined interval  $[-0.5, 1]$  with an equal subinterval length of 0.005. To obtain the reduced-bias estimator, we used 10,000 repetitions for each grid point. The obtained estimator,  $\hat{\theta}$ , has a substantially reduced bias compared with that of the posterior mean, denoted by  $\delta_M$  in Table 2. The standard deviation of the proposed estimators are not significantly different from the standard deviation of the posterior mean.

## 6. Example

We apply the proposed design to a randomized, placebo-controlled, double-blind trial involving patients with acne papulopustulosa of Plewig's grade 11–111. Investigators examined the effect of treatment under a combination of 1% chloramphenicol and 0.5% pale sulfonated shale oil versus an alcoholic vehicle (placebo) (Lehmacher & Wassmer, 1999; Fluhr et al., 1998). After 6 weeks of treatment, the level of bacteria was compared to the baseline level for patients in the active treatment group versus those in the placebo group. With  $\alpha = 0.01$  and  $1 - \beta = 0.95$ ,  $h_0(\theta) = h_1(\theta) = |\theta| + c$ , we used  $K_2 = 3 \times 10^5$ ,  $c = BK_2 = 0.00018$ ,  $K_1 = 1$  and  $K_0 = r/(1 - r) = 1933.9$ , where  $r = 0.9995$  is calculated from expression (3.14). Assume that  $\delta = 1$  and  $\hat{\sigma}_0 = 2$  based on the prior information. The quantities needed for conducting the proposed design are listed in Table 3, where  $\hat{\sigma}_j$  is the sample standard deviation from the cumulative data and Pred. Power is the notation for the predicted power  $P(R_{j+1} | \theta = \delta_j, x_j)$ . The block means, not cumulative means, are denoted by  $x_j = \bar{x}_{T_j} - \bar{x}_{C_j}$ .

In the first interim analysis,  $L_A(x_1) = 1.507 > L_{cont}(x_1) = 0.210$ , therefore, the next step is to check the predicted power. Since the predicted power, 0.946, is less than  $1 - \beta = 0.95$ , the trial continues to the next block. At the second interim analysis,  $L_A(x_2) > L_{cont}(x_2)$ , and the predicted power is 0.997, which is great than 0.95. Therefore, we stop the trial. Because  $L_R(x_2) = 0.061 < L_A(x_2) = 1.530$ , the conclusion is to reject  $H_0: \theta = 0$ . The procedure is robust in terms of the choices of  $\delta$  and  $\hat{\sigma}_0$ . When  $\delta$  varies from 0.5 to 2.5 and  $\hat{\sigma}_0$  varies from 1 to 2.5, the conclusion remains the same and the trial is always terminated after no more than two interim analyses.

## 7. Discussion

We have proposed an efficient design with the new aspects of allowing the investigators to evaluate both efficacy and futility at each interim analysis, which leads to a saving in sample size. The one-stage-ahead expected loss and predicted power assessment allow the trial to be terminated once evidence favoring a treatment, or not, is sufficient. The proposed design has the following desirable merits. First, the method can terminate a trial as soon as the evidence is sufficient to show the efficacy of a treatment. Second, the method ensures the control of the type I error rate, which is a practical requirement for trial designs under the regulatory setting, whether Bayesian or frequentist. Third, the proposed method avoids premature decisions with a flexible maximum sample size by relying on the mechanism of assessing the conditional probability. Compared with the other two adaptive designs in simulations, the proposed design is more robust to a range of assumed prior distributions, but more sensitive to the current data.

We have provided a new estimation procedure to obtain a reduced-bias estimator for the parameter of interest. One major advantage of this method is that the estimation procedure takes into account the whole process of the complicated stopping rule through Monte Carlo simulations. The empirical studies have shown that the estimator behaves well under various scenarios. Although this estimation procedure is developed for the proposed design, the methodology of searching for an unbiased parameter estimator may be applied to other Bayesian group sequential designs, as well as classical group sequential designs with the posterior mean replaced by its naive estimate, as in (4.1).

Although we assumed normal observations in illustrating the design, the large sample properties in Section 3 only require independent increments from each block of data to be approximately normally distributed, which will allow extension of the method to outcomes with other distributions and time-to-event outcomes with staggered entry using log-rank type test statistics (Tsiatis et al., 1985 and Shen & Cai, 2003).

The maximum number of blocks is sequentially determined using interim data. When the maximum number of interim analyses is fixed, optimizing the block size  $B$  is possible theoretically, but could be difficult computationally. At the first interim analysis, the conditional power can be variable if the size of the first block of data is small. That is why we recommend choosing a relatively larger block size initially, but a constant, smaller block size later. Beyond the initial block, the smaller the block size, the more efficient is the procedure. The most efficient block size would be one, if we do not take the cost of the interim analysis into consideration. But such a fully sequential design after the first block may not be practical in clinical trial conduct. Therefore, we recommend a small block size, such as 5 to 10, after the first block.

We recognize the tradeoff between efficiency and flexibility for clinical trial designs. While being more flexible, the proposed group sequential design may not be the most efficient design as defined in the mathematical sense by Tsiatis and Mehta (2003). As shown in several recent papers (e.g., Bauer, Brannath and Posch, 2001; Brannath, Bauer and Posch, 2006), the most efficient design (in a mathematical sense), however, relies on many underlying assumptions such as the right spending functions and that the sample sizes should be specified a priori. The designs could be poorly underpowered if any of these design parameters were misspecified in the planning phase, which is very likely in practice.

There are different approaches to specifying  $\delta$ , at which the power is set for group sequential designs, such as using a minimal clinical significant treatment difference (Barber and Jennison, 2002) or an anticipated effective size. Achieving the desired power at the true effective size in a neighborhood of  $\delta$  has gained more attention for its practicality and flexibility in clinical trial designs in recent years. That is the rationale we use as we set our objectives for the proposed design.

## References

1. Barber, Jennison. Optimal asymmetric one-sided group sequential tests. *Biometrika*. 2002; 89:49–60.
2. Berry DA. Discussion of a paper by Spiegelhalter, D.J., Freedman, L.S. & Parmar, K.B. Bayesian approaches to randomized trials. *J R Statist Soc A*. 1994; 157:399.
3. Berry DA, Ho C. One-sided sequential stopping boundaries for clinical trials: a decision-theoretical approach. *Biometrics*. 1988; 44:219–227. [PubMed: 3358990]
4. Bickel, P.J.; Doksum, K.A. *Mathematical Statistics*. 2. Vol. I. Upper Saddle River, New Jersey: Prentice Hall; 2001.
5. Cheng Y, Shen Y. Bayesian adaptive designs for clinical trials. *Biometrika*. 2005; 92:633–646.
6. DeMets DL, Ware JH. Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika*. 1982; 69:661–3.
7. Fisher L. Self-designing clinical trials. *Statist Med*. 1998; 17:1551–1562.
8. Fluhr JW, Gloor M, Merkel W, Warnecke J, Hoffler U, Lehmacher W, Glutsch J. Antibacterial and sebosuppressive efficacy of a combination of chloramphenicol and pale sulfonated shale oil. *Arzneimittel-Forschung/Drug Research*. 1998; 48(I):188–196. [PubMed: 9541732]
9. Gittins J, Pezeshk H. How large should a clinical trial be? *Statistician*. 2000; 49:177–187.
10. Jennison, C.; Turnbull, B. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC; 2000.
11. Jennison C, Turnbull B. Adaptive and nonadaptive group sequential tests. *Biometrika*. 2006; 93:1–21.
12. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999; 55:1286–1290. [PubMed: 11315085]
13. Lewis RJ, Berry DA. Group sequential clinical trials: a classical evaluation of Bayesian decision-theoretic designs. *J Am Statist Ass*. 1994; 89:1528–1534.



14. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35:549–556. [PubMed: 497341]
15. Schervish, MJ. *Theory of Statistics*. New York: Springer-Verlag; 1995.
16. Shen Y, Cai J. Sample size reestimation for clinical trials with censored survival data. *J Amer Statist Assoc*. 2003; 98:418–26.
17. Shen Y, Fisher L. Statistical inference for self-designing clinical trials with a one-sided hypothesis. *Biometrics*. 1999; 55:190–197. [PubMed: 11318154]
18. Slud EV, Wei LJ. Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J Am Statist Assoc*. 1982; 77:862–868.
19. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *J R Statist Soc A*. 1994; 157:357–416.
20. Tsiatis AA, Mehta C. On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*. 2003; 90:367–378.
21. Tsiatis AA, Rosner GL, Trichler DL. Group sequential tests with censored survival data adjusting for covariates. *Biometrika*. 1985; 72:365–373.
22. Wang, Leung. Bias reduction via resampling for estimation following sequential tests. *Sequential Analysis: Design Methods and Applications*. 1997; 16:249–267.

## Appendix A. Proofs of Theorems 2.1 and 2.2

For the natural flow of the proofs, we prove Theorem 2.2 first and Theorem 2.1 next.

### Proof of Theorem 2.2

We prove (2.4) first. Assume that  $\theta^* > 0$ . For any given positive  $\varepsilon$ , there exists an  $N > \varepsilon$  such that  $h_1(\theta) = (1 + |\theta|^w) < \varepsilon$  whenever  $\theta > N$ . Let

$$U_{h_1}(N) = \max\{h_1(\theta) : |\theta - \theta^*| \leq N\}, \quad U = \limsup_{j \rightarrow \infty} \left\{ E \left( |\theta|^w | \mathcal{X}_j \right) \right\}.$$

Under the regularity conditions (Schervish, 1995, p. 429, 450),

$$\liminf_{j \rightarrow \infty} P(|\theta - \theta^*| < \varepsilon | \mathcal{X}_j) = 1$$

almost surely and  $E(|\theta|^w | \mathcal{X}_j)$  differs from  $|\hat{\theta}|^w$  by  $O(m_j^{-1})$ , where  $\hat{\theta}$  is MLE of  $\theta$  and  $m_j = \sum_{i=1}^j B_i$ . Hence,  $U$  is finite almost surely and we have

$$\begin{aligned} & \limsup_{j \rightarrow \infty} \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_j) d\theta \\ & \leq \limsup_{j \rightarrow \infty} \left[ \{h_1(\theta^*) + O(\varepsilon)\} P(|\theta - \theta^*| < \varepsilon | \mathcal{X}_j) \right. \\ & \left. + U_{h_1}(N) P(|\theta - \theta^*| \geq \varepsilon | \mathcal{X}_j) + \int_{|\theta - \theta^*| \geq N} \frac{h_1(\theta)(1 + |\theta|^w)}{(1 + |\theta|^w)} \pi(\theta | \mathcal{X}_j) d\theta \right] \quad (\text{A.1}) \\ & \leq [\{h_1(\theta^*) + O(\varepsilon)\} + \varepsilon(1 + U)] \end{aligned}$$

almost surely. A similar argument holds for  $\liminf$ . Letting  $\varepsilon$  approach to zero, we obtain the following inequalities:

$$h_1(\theta^*) \leq \liminf_{j \rightarrow \infty} \int_0^\infty h_1(\theta)\pi(\theta|\mathcal{X}_j)d\theta \leq \limsup_{j \rightarrow \infty} \int_0^\infty h_1(\theta)\pi(\theta|\mathcal{X}_j)d\theta \leq h_1(\theta^*).$$

Hence,

$$\lim_{j \rightarrow \infty} \int_0^\infty h_1(\theta)\pi(\theta|\mathcal{X}_j)d\theta = h_1(\theta^*) \quad (A.2)$$

almost surely when  $\theta^* > 0$ . Similarly, we can show the follows:

$$\begin{aligned} \lim_{j \rightarrow \infty} \int_{-\infty}^0 h_0(\theta)\pi(\theta|\mathcal{X}_j)d\theta &= h_0(\theta^*) \quad \text{a.s., if } \theta^* < 0; \\ \lim_{j \rightarrow \infty} \int_0^\infty h_1(\theta)\pi(\theta|\mathcal{X}_j)d\theta &= 0 \quad \text{a.s., if } \theta^* < 0; \\ \lim_{j \rightarrow \infty} \int_{-\infty}^0 h_0(\theta)\pi(\theta|\mathcal{X}_j)d\theta &= 0 \quad \text{a.s., if } \theta^* > 0. \end{aligned} \quad (A.3)$$

Since

$$\left[ E\{L(\theta, A, j)|\mathcal{X}_j\} - E\{L(\theta, R, j)|\mathcal{X}_j\} \right] = K_1 \int_0^\infty h_1(\theta)\pi(\theta|\mathcal{X}_j)d\theta - K_0 \int_{-\infty}^0 h_0(\theta)\pi(\theta|\mathcal{X}_j)d\theta,$$

(2.4) is a direct result of (A.2) and (A.3).

For the second part of Theorem 2.2, assuming  $\theta^* > 0$ , since

$$\lim_{j \rightarrow \infty} \left[ E\{L(\theta, A, j+1)|\mathcal{X}_{j+1}\} - E\{L(\theta, R, j+1)|\mathcal{X}_{j+1}\} \right] = K_1 h_1(\theta^*) > 0$$

almost surely, we have

$$\lim_{j \rightarrow \infty} I_{R_{j+1}}(\mathcal{X}_{j+1}) = 1 \text{ a.s.,}$$

where

$$I_{R_{j+1}}(\mathcal{X}_{j+1}) = \begin{cases} 1, & \text{if } E\{L(\theta, A, j+1)|\mathcal{X}_{j+1}\} - E\{L(\theta, R, j+1)|\mathcal{X}_{j+1}\} > 0; \\ 0, & \text{elsewhere.} \end{cases}$$

Hence,

$$\lim_{j \rightarrow \infty} P(R_{j+1}|\theta = \delta_j, \mathcal{X}_j) = \lim_{j \rightarrow \infty} E(I_{R_{j+1}}(\mathcal{X}_{j+1})|\theta = \delta_j, \mathcal{X}_j) = 1$$

almost surely.

For the last part of Theorem 2.2, assuming  $\theta^* = 0$ , by a similar argument of (A.1), we have

$$E\{L(\theta, A, j)|\mathcal{X}_j\} - E\{L(\theta, R, j)|\mathcal{X}_j\} = K_1 h_1(0)P(0 \leq \theta \leq \varepsilon|\mathcal{X}_j) - K_0 h_0(0)P(-\varepsilon \leq \theta < 0|\mathcal{X}_j) + O(\varepsilon) + o_\varepsilon(1) \quad (\text{A.4})$$

$$= \{K_1 h_1(0) + K_0 h_0(0)\}P(0 \leq \theta \leq \varepsilon|\mathcal{X}_j) - K_0 h_0(0) + O(\varepsilon) + o_\varepsilon(1),$$

where  $o_\varepsilon(1)$  converges to 0 almost surely as  $j$  approaches to infinity for any given  $\varepsilon > 0$ . The last equality of (A.4) follows from the facts that

$$P(-\varepsilon \leq \theta < 0|\mathcal{X}_j) = 1 - P(0 \leq \theta \leq \varepsilon|\mathcal{X}_j) - P(|\theta| > \varepsilon|\mathcal{X}_j)$$

and

$$\lim_{j \rightarrow \infty} P(|\theta| > \varepsilon|\mathcal{X}_j) = 0 \text{ a.s.}$$

Letting  $\hat{\theta}_{m_j}$  be the MLE and  $I(\theta)$  be the Fisher information number, we have

$$\limsup_{j \rightarrow \infty} \left| P \left( \frac{\sqrt{m_j}(\theta - \hat{\theta}_{m_j})}{I^{-1/2}(0)} \leq x | \mathcal{X}_j \right) - \Phi(x) \right| = 0 \text{ a.s.} \quad (\text{A.5})$$

(Bickel, 2001, p. 339). In light of (A.5),

$$P(0 \leq \theta \leq \varepsilon|\mathcal{X}_j) = 1 - \Phi \left( -\frac{\sqrt{m_j} \hat{\theta}_{m_j}}{I^{-1/2}(0)} \right) + o_\varepsilon(1). \quad (\text{A.6})$$

Moreover,  $\Phi(-\sqrt{m_j} \hat{\theta}_{m_j} / I^{-1/2}(0))$  converges to  $\Phi(Z)$  in distribution, where  $Z$  follows a standard normal distribution. As a result of (A.4) and (A.6),

$$\limsup_{j \rightarrow \infty} P \left[ E\{L(\theta, A, j)|\mathcal{X}_j\} - E\{L(\theta, R, j)|\mathcal{X}_j\} < 0 \right] = \limsup_{j \rightarrow \infty} P \left[ \{K_1 h_1(0) + K_0 h_0(0)\} \{1 - \Phi(Z)\} - K_0 h_0(0) + O(\varepsilon) < 0 \right].$$

Let  $\varepsilon$  go to zero. Noting that  $1 - \Phi(Z)$  follows a uniform distribution on  $(0, 1)$  and that the above argument also holds for  $\liminf$ , we have

$$\lim_{j \rightarrow \infty} P \left[ E\{L(\theta, A, j)|X_j\} - E\{L(\theta, R, j)|X_j\} < 0 \right] = P \left\{ 1 - \Phi(Z) < \frac{K_0 h_0(0)}{K_1 h_1(0) + K_0 h_0(0)} \right\} = \frac{K_0 h_0(0)}{K_1 h_1(0) + K_0 h_0(0)}. \quad (\text{A.7})$$

Hence, the last part of Theorem 2.2 is proved.

### Proof of Theorem 2.1

When  $\theta^* < 0$ , in view of the second equation in expression (A.3),

$$\lim_{j \rightarrow \infty} \{L_A(\mathcal{X}_j) - L_{cont}(\mathcal{X}_j)\} \leq \lim_{j \rightarrow \infty} \left[ \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_j) d\theta - 2B_{j+1}K_2 \right] \leq 0 - 2BK_2 < 0$$

almost surely. Consequently,  $P_{\theta^*}(M < \infty) = 1$ .

Next, assuming  $\theta^* > 0$ , the trial stops at  $M = j$  if

$$P(R_{j+1} | \theta = \delta_j, \mathcal{X}_j) \geq 1 - \beta.$$

In the second part of Theorem 2.2, we proved that

$$\lim_{j \rightarrow \infty} P(R_{j+1} | \theta = \delta_j, \mathcal{X}_j) = 1$$

almost surely. Subsequently,  $P_{\theta^*}(M < \infty) = 1$ .

Finally, if  $\theta^* = 0$  and  $K_1 h_1(0) < 2B_{j+1}K_2$  for  $j \geq 1$ ,

$$\liminf_{j \rightarrow \infty} P_0 \{L_A(\mathcal{X}_j) - L_{cont}(\mathcal{X}_j) < 0\} \geq \liminf_{j \rightarrow \infty} P_0 [E\{L(\theta, A, j) | \mathcal{X}_j\} < 2B_{j+1}K_2] = \liminf_{j \rightarrow \infty} P_0 \left[ \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_j) d\theta < 2B_{j+1}K_2 \right] \quad (A.8)$$

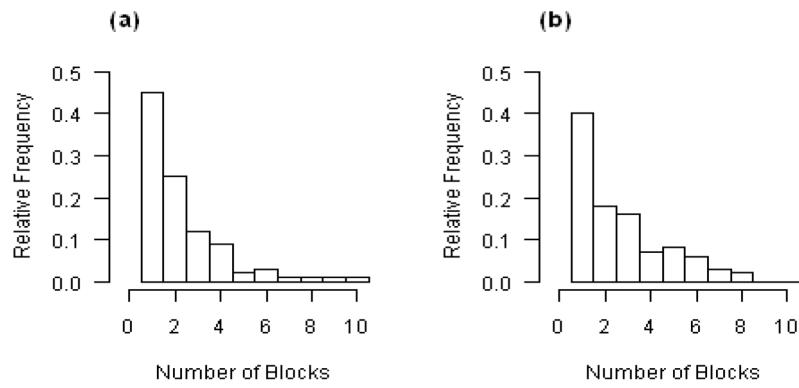
By a similar argument that yields (A.4) through (A.7), we have

$$\begin{aligned} \lim_{j \rightarrow \infty} P_0 \left[ \int_0^\infty h_1(\theta) \pi(\theta | \mathcal{X}_j) d\theta < 2B_{j+1}K_2 \right] &= P \left[ K_1 h_1(0) \{1 - \Phi(Z)\} < 2B_{j+1}K_2 \right] \\ &= P_0 \left[ \{1 - \Phi(Z)\} < \frac{2B_{j+1}K_2}{K_1 h_1(0)} \right] = 1. \end{aligned} \quad (A.9)$$

The last equality follows from the facts that  $1 - \Phi(Z)$  follows a uniform distribution on (0, 1) and that  $2B_{j+1}K_2 = K_1 h_1(0) > 1$ . Combining (A.8) and (A.9), we have

$$\lim_{j \rightarrow \infty} P_0 \{L_A(\mathcal{X}_j) - L_{cont}(\mathcal{X}_j) < 0\} = 1,$$

which implies  $P(M < \infty) = 1$ .



**Figure 1.**

The histogram of the number of blocks as relative frequencies for (a) under  $H_0, \theta = 0$  and (b) under  $H_1, \theta = 0.5$ . For both (a) and (b),  $\alpha = 0.025, \beta = 0.1, \delta = 0.5, B_1 = 15, B = 8, \sigma = 1, K_2 = 3 * 10^5, c = BK_2, K_1 = 1$  and  $K_0 = r/(1 - r)$ , where  $r$  is defined by (3.14).

**Table 1**

The comparison of power and average sample number (ASN) between the proposed design and other group sequential designs with one-sided  $\alpha = 0.025$ ,  $B = 6$ , and true  $\theta = 0$  under null and  $\theta = 0.5$  under the alternative.

Design	$\delta$	$\hat{\alpha}$	ASN $_{\alpha}$	$1-\hat{\beta}$	ASN $_{\beta}$
OBF	0.4	0.023	68.7	0.983	84.2
Self-designing	0.4	0.010	85.4	0.923	87.2
Bayes Adaptive	0.4	0.014	55.3	0.949	75.0
ED	0.4	0.025	42.9	0.908	55.0
OBF	0.7	0.024	26.9	0.644	38.2
Self-designing	0.7	0.017	49.0	0.798	62.6
Bayes Adaptive	0.7	0.014	45.4	0.912	66.4
ED	0.7	0.019	42.7	0.901	54.8

Monte Carlo simulation based on 10,000 repetitions.  $B_0 = 1$ ,  $B_1 = 15$ ,  $\sigma = 1$ ,  $K_1 = 1$ ,  $K_2 = 3 * 10^5$ ,  $c = BK_2$ , and  $K_0 = r/(1 - r)$ , where  $r$  is defined by (3.14). ASN $_{\alpha}$  and ASN $_{\beta}$  are average sample sizes under  $\theta = 0$  and  $\theta = 0.5$ , respectively.

\$watermark-text

\$watermark-text

\$watermark-text

**Table 2**

The comparison of proposed estimate  $\hat{\theta}$  and posterior mean  $\delta_M$  following an ED with  $\beta = 0.1$ ,  $B = 6$  and the assumed treatment difference  $\delta = 0.6$ .

$\alpha$	$\theta$	$\hat{\theta}$	$\delta_M$
0.025	0.0	-0.018	0.213
	0.3	0.295	0.235
	0.4	0.416	0.230
	0.5	0.519	0.224
0.05	0.6	0.625	0.214
	0.0	-0.015	0.212
	0.3	0.304	0.235
	0.4	0.425	0.230
0.1	0.5	0.523	0.219
	0.6	0.616	0.218
	0.0	-0.026	0.186
	0.3	0.295	0.300
0.2	0.4	0.416	0.436
	0.5	0.519	0.549
	0.6	0.625	0.658
	0.0	-0.015	-0.025
0.3	0.3	0.304	0.320
	0.4	0.425	0.455
	0.5	0.523	0.558
	0.6	0.616	0.645

Monte Carlo simulation based on 10,000 repetitions.  $B_0 = 1$ ,  $B_1 = 15$ ,  $\sigma = 1$ ,  $K_1 = 1$ ,  $c = BK_2$ ,  $K_2 = 3 * 10^5$  and  $K_0 = r(1 - r)$ , where  $r$  is defined by (3.14).

\$watermark-text

\$watermark-text

\$watermark-text

**Table 3**

Illustration of the steps of the ED procedures.

$j$	$B_j$	$x_j$	$\hat{\sigma}_j$	$L_A$	$L_{cont}$	Pred. Power	$L_R$	Decision
1	12	1.549	1.861	1.507	0.210	0.946		Continue
2	6	1.580	1.932	1.530	0.051	0.997	0.061	Stop; D=R