

Asymptotic Optimal Sample Sizes for Discovering a Promising Treatment in Phase II Group Sequential Clinical Trials

BY Yi Cheng

Indiana University South Bend, South Bend, Indiana 46634, USA

e-mail: ycheng@iusb.edu or yicheng@odin.mdacc.tmc.edu

AND Gary L. Rosner

The University of Texas, M. D. Anderson Cancer Center, Houston, Texas 77030, USA

e-mail: glr@odin.mdacc.tmc.edu

SUMMARY

We propose an asymptotic optimal procedure for determining the per-study sample size of a series of Phase II clinical studies. The sequence of studies seeks to discover an active treatment among a sequence of treatments under evaluation. Each phase II study proceeds in r stages with $r=1, 2, \text{ or } 3$.

Our result serves two practical purposes. First, it yields asymptotic optimal sample sizes. Second, it reduces the dimension of the design space for finding the exact optimal sample sizes. The criterion of optimality is to minimize the total number of observations until one promising treatment is identified, while controlling the probabilities of false positive and false negative errors to desired levels. The asymptotic optimal sample sizes are given explicitly based on analytical results and their computation is straightforward. Numerical studies show that the asymptotic optimal sample sizes and true optimal sample sizes are quite close. The Bayesian formulation of the problem is used in the procedure and choice of prior distributions is discussed.

Some key words: Asymptotic optimal sample sizes; Bayesian; False positive; False Negative; Screen clinical Trials; Group sequential design.

1 Introduction

Drug and medical device development has traditionally progressed along a path from early phase studies evaluating safety to later studies looking at efficacy (Friedman, Furberg, and Demets, 1998). In drug development, phase II studies seek to learn about the activity of new agents or therapies. A similar process is followed for medical devices. Once a therapy appears “active” in phase II, further phase II studies may be carried out or randomized phase III clinical trials may be mounted. Regulatory approval and marketing may follow if the new agent compares favorably in the randomized trials.

At large institutions dedicated to clinical research, a large number of new agents or new combinations of anticancer agents undergo evaluation for activity. The process is typically carried out through separate phase II studies with only informal learning carried out between studies—even if the studies draw patients with similar disease characteristics. Most designs for screening new agents and culling active therapies from the many agents under investigation, consider each study in isolation, even though there might be some sharing of information (borrowing strength) across studies that use rather similar agents or medical devices that differ only slightly. Yao, Begg, and Livingston (1996) proposed a formal way to carry out phase II screening activities in the context of vaccine trials. These authors view the process as one large enterprise within which multiple agents, introduced by some mechanism, undergo screening for activity and either progress to further testing or are discarded. The design decision concerns choosing the sample size for each phase II study to minimize the overall expected sample size needed (or time) until an active agent is identified. The decision problem is discussed in the frequentist paradigm, in which the type I and type II (false positive and false negative, respectively) error probabilities are prespecified and preserved over the sequence of experiments. The formal setup in Yao, Begg, and Livingston (1996) considers one treatment at a time. The underlying probability model is binomial sampling, i.e., the model assumes independent binary outcomes. For each treatment being considered, the process defines a single-arm clinical study. The design fixes the sample size n and a threshold k for the number of treatment successes beyond which a treatment is identified as

promising and recommended for further investigation.

In later work, Yao and Venkatraman (1998), Wang and Leung (1998), and Leung and Wang (2001) considered a variety of extensions leading to 2-stage designs, and fully sequential designs in the same setup. Whitehead (1985) discusses a similar problem, except that all the agents are available for testing at the start and the study proceeds by simultaneous evaluation. Strauss and Simon (1995) consider a generalization based on two-armed randomized trials for each new treatment. One arm is the new treatment, and the other arm is the best treatment found so far. At the end of this sequence of randomized studies, one chooses the “winner” that one will compare to a standard regimen in a randomized comparative trial.

None of these extensions are fully Bayesian, however, meaning that there is no learning or optimization accounting for uncertainty. Additionally, it is well known that Bayesian inferential procedures enjoy optimality properties Bergman and Gittins (1985).

In this paper, we present an easy way to compute analytically the asymptotic optimal sample size for each separate study in the setting discussed in Yao, Begg, and Livingston (1996). We also extend the results to three-stage design with a brief discussion of r -stage for general r . We describe the setting of successive non-sequential (i.e., one-stage) studies in the next section to introduce notation and provide a context for what follows. We then give the mathematical formulation behind the optimal sample size computation in the original setting and for individual one-stage, two-stage and three-stage trials. We also present numerical results, followed by a discussion mainly about reducing the dimension of the design space for finding the exact optimal sample sizes. An appendix contains mathematical details.

2 Asymptotic Optimal Procedures for One-stage Trials

2.1 The decision problem.

The design was introduced by Yao, Begg, and Livingston (1996). In their setting, the investigators have a set of agents (vaccines in their example) they wish to evaluate, and the goal is to determine which are active and worth further investigation and which are not “promising.” The studies of the new agents proceed sequentially, one agent at a time, until a

promising agent is identified. The common trial size for each agent is n . The responses from patients are independent and dichotomous. Let X_i denote the number of positive responses for the i^{th} study. The investigators have decided to identify an agent as “active” if and only if they observe $X_i > k$. Given θ_i , the probability a patient responds (i.e., a success) in the i^{th} study, X_i is binomial

$$X_i|\theta_i \sim \text{Binomial}(n, \theta_i), \quad i = 1, 2, \dots, M,$$

where

$$M = \inf\{m : \text{the } m \text{ th agent is the first agent identified as promising}\}.$$

Using a Bayesian formulation of the problem, let $F(\theta)$ be the prior distribution function, so

$$\theta_1, \theta_2, \theta_3, \dots \text{ are i.i.d. } \sim F(\theta), \quad \text{with density function } f(\theta).$$

An agent is truly promising if $\theta > \theta^*$, where θ^* is some predetermined target response probability. We adopt the notation of Yao, Begg, and Livingston (1996) and let

$$\begin{aligned} p_{--}(n, k) &= \int_0^{\theta^*} P(X_i \leq k|\theta, n)dF(\theta), & p_{+-}(n, k) &= \int_0^{\theta^*} P(X_i > k|\theta, n)dF(\theta), \\ p_{-+}(n, k) &= \int_{\theta^*}^1 P(X_i \leq k|\theta, n)dF(\theta), & p_{++}(n, k) &= \int_{\theta^*}^1 P(X_i > k|\theta, n)dF(\theta). \end{aligned} \quad (2.1)$$

Define the loss function to be the expected total sample size until a promising agent is identified. Let

$$\mathcal{N}(n, k) = E(N) = nE(M) = \frac{n}{p_{+}(n, k)}, \quad (2.2)$$

where $p_{+}(n, k) = p_{++}(n, k) + p_{+-}(n, k)$. The last equality in (2.2) follows from the fact that M has a geometric distribution with probability of success $p_{-} = 1 - p_{+}(n, k)$.

The decision problem is to choose (n, k) such that $\mathcal{N}(n, k)$ is minimized under the constraint that the probabilities of false positive (E_1) and false negative (E_2) decisions are no greater than α_1 and α_2 , respectively. Mathematically,

$$P_n(E_1) = \frac{p_{+-}(n, k)}{p_{+}(n, k)}, \quad \text{and} \quad P_n(E_2) = \frac{p_{-+}(n, k)}{p_{+}(n, k) + p_{-+}(n, k)}. \quad (2.3)$$

2.2 Asymptotic Optimal Procedures.

The search for the exact optimal solution (n^o, k^o) to minimize $\mathcal{N}(n, k)$ can be computationally intensive for multistage designs. In this section, we find the asymptotic optimal design, the computation of which is straightforward. The computational algorithm for determining the asymptotic optimal solution (n^a, k^a) rests on the following two propositions, whose proofs are in the Appendix.

PROPOSITION 1. *Assume the prior distribution of θ has a continuous density function $f(\theta)$ that is positive at θ^* . The optimal solution (n^o, k^o) satisfies $k^o/n^o = \theta^* + o(1)$. \square*

The result given in Proposition 1 is natural, since x/n is an unbiased and consistent estimate of θ . Based on this result, we can narrow the design space when searching for the optimal design. Let

$$k^a = \begin{cases} \lfloor n^a \theta^* \rfloor, & \text{if } \alpha_1 \leq \alpha_2 < 2\alpha_1 ; \\ \lfloor n^a \theta^* \rfloor + 1, & \text{if } 2\alpha_1 \leq \alpha_2 \leq 3\alpha_1 , \end{cases} \quad (2.4)$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x . The association between the optimal k and the values of the α_1 and α_2 is that increasing k will increase the probability of a false negative decision and decrease the probability of a false positive decision.

Computation of the optimal solution requires evaluation of the error probabilities (2.1), and the next proposition gives their asymptotic values.

PROPOSITION 2. *Assume a prior distribution with density function $f(\theta)$ that is continuous and positive at θ^* . k is defined in (2.4). Then*

$$\begin{aligned} \int_0^{\theta^*} P(X > k|\theta) f(\theta) d\theta &= g(\theta^*)/n^{1/2} + o(1/n^{1/2}), \\ \int_{\theta^*}^1 P(X \leq k|\theta) f(\theta) d\theta &= g(\theta^*)/n^{1/2} + o(1/n^{1/2}), \end{aligned} \quad (2.5)$$

where

$$g(\theta) = \frac{[\theta(1-\theta)]^{1/2} f(\theta)}{(2\pi)^{1/2}}, \quad \text{for } 0 < \theta < 1. \quad \square \quad (2.6)$$

If f has a continuous derivative in $(0,1)$, $o(1/n^{1/2})$ can be replaced by $O(1/n)$ in (2.5) .

Applying Proposition 2 to the expressions in (2.1), we obtain

$$\begin{aligned} p_{+-}(n, k) &= g(\theta^*)/n^{1/2} + o(1/n^{1/2}), \\ p_{++}(n, k) &= P(\theta > \theta^*) - g(\theta^*)/n^{1/2} + o(1/n^{1/2}), \end{aligned}$$

$$p_{-+}(n, k) = g(\theta^*)/n^{1/2} + o(1/n^{1/2}).$$

As a result,

$$\begin{aligned} P_n(E_1) &= \frac{g(\theta^*)}{\sqrt{n}P(\theta > \theta^*)} + o(1/n^{1/2}), \\ P_n(E_2) &= \frac{g(\theta^*)}{g(\theta^*) + \sqrt{n}P(\theta > \theta^*)} + o(1/n^{1/2}). \end{aligned} \quad (2.7)$$

Based on these properties, we propose the following procedure for determining the asymptotic optimal solution.

ALGORITHM

- (1) Determine the target value θ^* , the prior density $f(\theta)$, and the error rates α_1 and α_2 .
- (2) Compute $g(\theta^*) = [\theta^*(1 - \theta^*)]^{1/2}f(\theta^*)/(2\pi)^{1/2}$.
- (3) Tabulate $P_n(\widehat{E}_1)$ and $P_n(\widehat{E}_2)$, defined explicitly in (2.8) and (2.9), as functions of n .
- (4) Let $n^a = \min\{n : P_n(\widehat{E}_1) \leq \alpha_1, P_n(\widehat{E}_2) \leq \alpha_2\}$, and compute the k^a defined in (2.4).
(n^a, k^a) is an asymptotic optimal solution.

$P_n(E_1)$ and $P_n(E_2)$ are probabilities of discrete random variables. As functions of n , their graphs are “jagged”. For moderate n (40–50 patients, as is typical for phase II clinical studies), the dominant terms defined on the right-hand sides of (2.7) arise from lack of accuracy in estimating $P_n(E_1)$ and $P_n(E_2)$. We suggest using correction factors to adjust the bias and improve estimation. The correction factors should tend to 1 as n increases to preserve asymptotic optimality.

We define the *corrected* estimated probabilities of false positive and false negative errors, $P_n(\widehat{E}_1)$ and $P_n(\widehat{E}_2)$ as the following.

$$P_n(\widehat{E}_1) = \begin{cases} \frac{g(\theta^*)}{\sqrt{n}P(\theta > \theta^*)} \cdot \left[\frac{\sqrt{n} + 3.5(n\theta^* - k)}{\sqrt{n} + 2(1 - \theta^*) + 1.8E(\theta)} \right], & \text{if } k = \lfloor n\theta^* \rfloor; \\ \frac{g(\theta^*)}{\sqrt{n}P(\theta > \theta^*)} \cdot \left[\frac{\sqrt{n} + 2 + 1.6(k - n\theta^*)}{\sqrt{n} + 2 + 2(1 - \theta^*) + 1.5E(\theta) + 0.5} \right], & \text{if } k = \lfloor n\theta^* \rfloor + 1, \end{cases} \quad (2.8)$$

and

$$P_n(\widehat{E}_2) = \begin{cases} \frac{g(\theta^*)}{g(\theta^*) + \sqrt{n}P(\theta > \theta^*)} \cdot \left[\frac{\sqrt{n} + 2(1 - \theta^*) + 1.8E(\theta) - 0.4}{\sqrt{n} + 3.5(n\theta^* - k)} \right], & \text{if } k = \lfloor n\theta^* \rfloor; \\ \frac{g(\theta^*)}{g(\theta^*) + \sqrt{n}P(\theta > \theta^*)} \cdot \left[\frac{\sqrt{n} + 2 + 2(1 - \theta^*) + 1.5E(\theta) + 0.1}{\sqrt{n} + 2 - 1.6(n\theta^* - k)} \right], & \text{if } k = \lfloor n\theta^* \rfloor + 1, \end{cases} \quad (2.9)$$

The correction factors (in square brackets) arise from the following considerations. The oscillations in $P_n(E_1)$ and $P_n(E_2)$ partly come from the discreteness that is directly associated with $t = n\theta^* - k$. In general, $P_n(E_1)$ increases and $P_n(E_2)$ decreases as t increases. The discrepancies between dominant terms on the right-hand sides of (2.7) and $P_n(E_1)$ and $P_n(E_2)$ are at the rate of $O(n^{-1/2})$ and related to the prior mean $E(\theta)$ and target value θ^* . The coefficients in the correction factors in (2.8) and (2.9) are established by extensive numerical studies.

EXAMPLE 1. Assume that the target value $\theta^* = 0.4$. The Jeffreys invariant prior distribution for the Bernoulli parameter θ is $\text{beta}(0.5, 0.5)$. The prior mean and variance are .5 and .125, respectively. $g(\theta^*) = [.4(1 - .4)]^{1/2}f(.4)/(2\pi)^{1/2} = .1270$, $P(\theta > \theta^*) = 1 - F(.4) = .5641$, where f and F are the density and distribution functions of $\text{beta}(.5, .5)$, respectively. Applying the 4-step procedure proposed in this section, we compute Table 1 for the $k = \lfloor n\theta^* \rfloor$ scenario,

Table 1

Table of corrected error probabilities by sample size for computing the asymptotic optimal sample size in Example 1.

n	3	4	5	6	7	8	...	19	20	21	22	23
$P_n(\widehat{E}_1)$.083	.113	.052	.078	.098	.057052	.034	.044	.053	.037
$P_n(\widehat{E}_2)$.162	.091	.161	.091	.063	.095046	.066	.049	.039	.053

For $\alpha_1 = \alpha_2 = 0.15$, the asymptotic optimal solution is $n^a = 4$ and $k^a = 1$. For $\alpha_1 = \alpha_2 = 0.1$, $n^a = 6$ and $k^a = 2$. For $\alpha_1 = \alpha_2 = 0.05$, $n^a = 21$ and $k^a = 8$. \square

Remark. Although we require $2\alpha_1 \leq \alpha_2 \leq 3\alpha_1$ for $k^a = \lfloor n\theta^* \rfloor + 1$ scenario, the same procedure works well for $\alpha_1 = 0.05$ and $\alpha_2 = 0.2$ if the prior density is truly in u-shape. For other types of densities, it is possible that $k^a > \lfloor n\theta^* \rfloor + 1$, but in general, $k^a \leq \lfloor n\theta^* \rfloor + 2$ for typical values of α_1 and α_2 .

3 Asymptotic Optimal Procedures for Two-stage Trials

Yao and Venkatraman (1998) proposed a two-stage design for screening trials to identify promising new agents. Including an interim analysis within each single-agent study reduces

the expected total sample size, compared to a one-stage design with the same constraint of probabilities of false positive and false negative errors. The two-stage design shares the philosophy of Simon’s optimal design, in that the trial ends at first stage only if it does not appear that the agent will show activity (i.e., for futility). For each trial, let n_i be the number of observations and X_i be the number of positive responses from the i^{th} stage, $i = 1, 2$. For given critical values k_1 and k_2 , if $X_1 \leq k_1$, we consider stopping and abandoning the agent. Otherwise, move on to the second stage, which is also the last stage. The agent is “promising” if and only if $X_1 + X_2 > k_2$. The decision problem is to find optimal (n_1, n_2, k_1, k_2) to minimize the overall sample size until one agent is identified as promising.

Stopping early to discard a treatment (i.e., for futility) does not inflate the probability of a false positive error for the two-stage design, but such stopping possibly inflates the probability of a false negative error if k_1 is not conservative. Therefore, when we address the asymptotic optimal solution for two-stage design, it is sensible to add the following condition,

$$\limsup \frac{k_1}{n_1} = \theta_1, \quad 0 < \theta_1 < \theta^*. \quad (3.1)$$

Numerical study suggests $\theta^*/4 \leq \theta_1 \leq \theta^*/2$ and $k_1^a = \lfloor n_1^a \theta_1 \rfloor$.

Use the same notation as in Yao and Venkatraman (1998). For each trial, let

$$\begin{aligned} p_{--}(n_1, n_2, k_1, k_2) &= \int_0^{\theta^*} [P(X_1 \leq k_1 | \theta) + P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta)] dF(\theta), \\ p_{-+}(n_1, n_2, k_1, k_2) &= \int_{\theta^*}^1 [P(X_1 \leq k_1 | \theta) + P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta)] dF(\theta), \\ p_{+-}(n_1, n_2, k_1, k_2) &= \int_0^{\theta^*} P(X_1 > k_1, X_1 + X_2 > k_2 | \theta) dF(\theta), \\ p_{++}(n_1, n_2, k_1, k_2) &= \int_{\theta^*}^1 P(X_1 > k_1, X_1 + X_2 > k_2 | \theta) dF(\theta). \end{aligned} \quad (3.2)$$

PROPOSITION 3. *Assume that prior distribution has a density function $f(\theta)$ that is continuous and positive at θ^* . Under the condition (3.1), we have*

$$\begin{aligned} p_{+-}(n_1, n_2, k_1, k_2) &= g(\theta^*)/(n_1 + n_2)^{1/2} + o[1/(n_1 + n_2)^{1/2}], \\ p_{++}(n_1, n_2, k_1, k_2) &= P(\theta > \theta^*) - g(\theta^*)/(n_1 + n_2)^{1/2} + o[1/(n_1 + n_2)^{1/2}], \\ p_{-+}(n_1, n_2, k_1, k_2) &= g(\theta^*)/(n_1 + n_2)^{1/2} + o[1/(n_1 + n_2)^{1/2}]. \end{aligned} \quad (3.3)$$

Furthermore,

$$\begin{aligned} P_n(E_1) &= \frac{g(\theta^*)}{\sqrt{n_1 + n_2}P(\theta > \theta^*)} + o[1/(n_1 + n_2)^{1/2}], \\ P_n(E_2) &= \frac{g(\theta^*)}{g(\theta^*) + \sqrt{n_1 + n_2}P(\theta > \theta^*)} + o[1/(n_1 + n_2)^{1/2}]. \quad \square \end{aligned} \quad (3.4)$$

We note that the dominant terms on the right-hand sides (3.4) depend on n_1 and n_2 only through their sum $n_1 + n_2$. Observing the analogy between (2.7) and (3.4), the optimal $n_1 + n_2$ for the two-stage design is about the same to the optimal n for the one-stage design asymptotically. This observation is also consistent with numerical results. It further verifies that (3.1) is a legitimate condition. The exact same 4-step procedure given in section 2.2 will yield an asymptotic optimal solution for the total sample size $n_1 + n_2$ in a two-stage design. Since some trials may stop for futility at the first stage, two-stage designs will reduce the overall sample size, on average. The expected total sample size is (with P_{et} = the probability of early termination)

$$\mathcal{N}(n_1, n_2, k_1, k_2) = \frac{n_1 + n_2(1 - P_{\text{et}})}{p_+}, \quad (3.5)$$

where

$$\begin{aligned} P_{\text{et}} &= \int_0^1 P(X \leq k_1 | \theta, n_1) f(\theta) d\theta = P(\theta \leq \theta_1) + o(1/n_1^{1/2}), \\ P_+ &= P_{+-} + P_{++} = P(\theta > \theta^*) + o[1/(n_1 + n_2)^{1/2}]. \end{aligned} \quad (3.6)$$

We wish to find the optimal (n_1, n_2, k_1, k_2) such that $\mathcal{N}(n_1, n_2, k_1, k_2)$ is minimized under the constrains of $P_n(E_1) \leq \alpha_1$ and $P_n(E_2) \leq \alpha_2$.

The expansions in (3.3) and (3.6) only give the first dominant term. If we expand them to the second term, we can verify analytically that $n_1^a = O((n_2^a)^{1/2})$ under assumption (3.1). Proposition 4 presents a more precise statement.

PROPOSITION 4. *Assume that prior distribution has a continuously differentiable density function $f(\theta)$ with $f(\theta_1) > 0$, $0 < \theta_1 < \theta^*$, and $f(\theta^*) > 0$. Under the condition (3.1), we have*

$$n_1^a \approx \left[\frac{h(\theta_1)}{P(\theta \leq \theta_1)} \right]^{1/2} (n_1^a + n_2^a)^{1/2},$$

where $h(x) = xf(x) - x(1-x)f'(x)/2$

We propose the following procedure for finding the asymptotic optimal solution for the two-stage design.

- (1) Use the four-step procedure in Section 2.2 to find the asymptotic optimal solution n^a . Let $n_1^a + n_2^a = n^a$ and $k_2^a = \lfloor n^a \theta^* \rfloor$ if $\alpha_1 \leq \alpha_2 < 2\alpha_1$, or $k_2^a = \lfloor n^a \theta^* \rfloor + 1$ if $2\alpha_1 \leq \alpha_2 \leq 3\alpha_1$.
- (2) Set $n_1^a = \lceil [h(\theta_1)/P(\theta \leq \theta_1)]^{1/2}(n_1^a + n_2^a)^{1/2} \rceil + 1$, $k_1^a = \lfloor n_1^a \theta_1 \rfloor$, and $n_2^a = n^a - n_1^a$, where $0 < \theta_1 < \theta^*$, $\lceil x \rceil$ is the smallest integer greater than or equal to x , and

$$h(\theta_1) = \theta_1 f(\theta_1) - \theta_1(1 - \theta_1) f'(\theta_1)/2.$$

$(n_1^a, n_2^a, k_1^a, k_2^a)$ is the asymptotic optimal solution. We recommend $\theta_1 = \theta^*/3$.

Remark. If the prior is beta(a,b), denoting its density as $f(x)$, then

$$h(x) = f(x) [1 + bx - a(1 - x)]/2.$$

Following the next section, we examine the performance of the above procedure by numerical comparison between the optimal solutions and asymptotic solutions.

4 Asymptotic Optimal Procedures for Three-stage Trials

In this section, we extend the analytical results in Propositions 3 and 4 to the three-stage case. One can extend the results to the r -stage case for general r .

In the three-stage design, we denote the exact optimal design and asymptotic optimal designs as

$$(n_1^o, n_2^o, n_3^o, k_1^o, k_2^o, k_3^o) \quad \text{and} \quad (n_1^a, n_2^a, n_3^a, k_1^a, k_2^a, k_3^a),$$

respectively. Given a design $(n_1, n_2, n_3, k_1, k_2, k_3)$, the first stage sample size is n_1 . If $X_1 \leq k_1$, we consider the treatment not worth further study; otherwise, we continue to the second stage with additional sample size n_2 . If $X_1 + X_2 \leq k_2$, we abandon further study of the treatment; otherwise, we continue to the third stage with additional sample size n_3 . If $X_1 + X_2 + X_3 \leq k_3$, the treatment is identified as not promising. If $X_1 + X_2 + X_3 > k_3$, the treatment is identified as promising.

As in (3.2), we define

$$\begin{aligned}
p_{-+}(n_1, n_2, n_3, k_1, k_2, k_3) &= \int_{\theta^*}^1 \left[P(X_1 \leq k_1 | \theta) + P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta) \right. \\
&\quad \left. + P(X_1 > k_1, X_1 + X_2 > k_2, X_1 + X_2 + X_3 \leq k_3 | \theta) \right] dF(\theta), \\
p_{+-}(n_1, n_2, n_3, k_1, k_2, k_3) &= \int_0^{\theta^*} P(X_1 > k_1, X_1 + X_2 > k_2, X_1 + X_2 + X_3 > k_3 | \theta) dF(\theta), \\
p_{++}(n_1, n_2, n_3, k_1, k_2, k_3) &= \int_{\theta^*}^1 P(X_1 > k_1, X_1 + X_2 > k_2, X_1 + X_2 + X_3 > k_3 | \theta) dF(\theta).
\end{aligned}$$

Extending assumption (3.1) to the three-stage case, we assume

$$\limsup \frac{k_1^a}{n_1^a} = \theta_1 \quad \text{and} \quad \limsup \frac{k_2^a}{n_1^a + n_2^a} = \theta_2, \quad 0 < \theta_1 < \theta_2 < \theta^*. \quad (4.1)$$

Propositions 5 and 6 are three-stage analogues of Propositions 3 and 4, respectively.

PROPOSITION 5. *Assume a prior distribution with density function $f(\theta)$ that is continuous and positive at θ^* . Under condition (4.1), with $n = n_1 + n_2 + n_3$, we have*

$$\begin{aligned}
p_{+-}(n_1, n_2, n_3, k_1, k_2, k_3) &= g(\theta^*)/n^{1/2} + o(n^{-1/2}), \\
p_{++}(n_1, n_2, n_3, k_1, k_2, k_3) &= P(\theta > \theta^*) - g(\theta^*)/n^{1/2} + o(n^{-1/2}), \\
p_{-+}(n_1, n_2, n_3, k_1, k_2, k_3) &= g(\theta^*)/n^{1/2} + o(n^{-1/2}).
\end{aligned}$$

Moreover,

$$\begin{aligned}
P_n(E_1) &= \frac{g(\theta^*)}{n^{1/2}P(\theta > \theta^*)} + o(n^{-1/2}), \\
P_n(E_2) &= \frac{g(\theta^*)}{g(\theta^*) + n^{1/2}P(\theta > \theta^*)} + o(n^{-1/2}). \quad \square
\end{aligned}$$

The expected total sample size is

$$\mathcal{N}(n_1, n_2, n_3, k_1, k_2, k_3) = \frac{n_1 + n_2(1 - P_{\text{et}1}) + n_3(1 - P_{\text{et}1} - P_{\text{et}2})}{p_{+}},$$

where

$$\begin{aligned}
P_{\text{et}1} &= \int_0^1 P(X_1 \leq k_1 | \theta, n_1) f(\theta) d\theta = P(\theta < \theta_1) + o(1/n_1^{1/2}), \\
P_{\text{et}2} &= \int_0^1 P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta, n_1, n_2) f(\theta) d\theta = P(\theta_1 < \theta \leq \theta_2) + o(1/n_1^{1/2}), \\
P_{+} &= P_{+-} + P_{++} = P(\theta > \theta^*) + o[1/n^{1/2}].
\end{aligned}$$

The goal is to choose the optimal $(n_1, n_2, n_3, k_1, k_2, k_3)$ such that $\mathcal{N}(n_1, n_2, n_3, k_1, k_2, k_3)$ is minimized under the constraints of $P_n(E_1) \leq \alpha_1$ and $P_n(E_2) \leq \alpha_2$.

PROPOSITION 6. Assume a prior distribution with continuously differentiable density function $f(\theta)$ such that $f(\theta_i) > 0$, $0 < \theta_1 < \theta_2 < \theta^*$, and $f(\theta^*) > 0$. Under condition (4.1), we have

$$\begin{aligned} n_1^a + n_2^a &\approx \left[\frac{h(\theta_2)}{P(\theta_1 < \theta \leq \theta_2)} \right]^{1/2} (n_1^a + n_2^a + n_3^a)^{1/2}, \\ n_1^a &\approx \left[\frac{h(\theta_1)}{P(\theta \leq \theta_1)} \right]^{1/2} (n_1^a + n_2^a)^{1/2}, \end{aligned}$$

where $h(x) = xf(x) - x(1-x)f'(x)/2$. \square

The corresponding procedure for finding the asymptotic optimal solution for a three-stage design is as follows.

- (1) Use the four-step procedure in Section 2.2 to find the asymptotic optimal solution n^a .
Let $n_1^a + n_2^a + n_3^a = n^a$ and $k_3^a = \lfloor n^a \theta^* \rfloor$ if $\alpha_1 \leq \alpha_2 < 2\alpha_1$, or $k_3^a = \lfloor n^a \theta^* \rfloor + 1$, if $2\alpha_1 \leq \alpha_2 \leq 3\alpha_1$.
- (2) Let $n_1^a + n_2^a = \lceil [n^a h(\theta_2)/P(\theta_1 < \theta \leq \theta_2)]^{1/2} \rceil + 1$, $n_1^a = \lceil [(n_1^a + n_2^a)h(\theta_1)/P(\theta \leq \theta_1)]^{1/2} \rceil + 1$, $k_2^a = \lfloor (n_1^a + n_2^a)\theta_1 \rfloor$, and $k_1^a = \lfloor n_1^a \theta_1 \rfloor$, $(n_1^a, n_2^a, n_3^a, k_1^a, k_2^a, k_3^a)$ is the asymptotic optimal solution. We recommend $\theta_1 = \theta^*/3$ and $\theta_2 = 2\theta^*/3$.

5 Numerical Results

We investigate the performance of the asymptotic optimal solutions numerically by comparing them with the values of true optimal solutions for one-stage, two-stage, and three-stage designs in Tables 2, 3, and 4, respectively. For the target values of θ^* equal to 0.3 and 0.6, we computed by extensive grid search the asymptotic optimal sample sizes over a wide range of beta distributions having prior mean at 0.2, 0.3, 0.4, 0.5 and prior variance at 0.08, 0.1, 0.12, respectively. For each case, we also computed the exact false positive and false negative error probabilities, $P_n(E_1)$ and $P_n(E_2)$ associated with the asymptotic optimal solutions. We also examine the properties of truncated trials, as in Section 2.3 of Yao, Begg, & Livingston (1996). Truncation-type early stopping occurs when there will not be enough responders (“successes”), even if all remaining patients at this stage respond. The tables also list the expected total sample size until a promising agent is identified for each optimal design with and without truncation (\mathcal{N}_T^o and \mathcal{N}^o , respectively).

Table 2 gives the numerical results for the one-stage case for two different combinations of (α_1, α_2) , $(0, 1, 0.1)$ and $(0.05, 0.15)$. The trial sizes n and critical values k for both exact and asymptotic optimal solutions are listed. We see that the asymptotic optimal solutions behave very well in the sense that (n^a, k^a) and (n^o, k^o) are either very close or the same across all scenarios. Also, the false positive and false negative error probabilities are generally at the nominal level.

Table 3 lists the numerical results for the two-stage case, where $n^o = n_1^o + n_2^o$ and $n^a = n_1^a + n_2^a$. We let $\theta_1 = \theta^*/3$, and $k_1^a = \lfloor n_1^a \theta^*/3 \rfloor$. Again, the asymptotic optimal procedure performs very well. $(n_1^a, n_2^a, k_1^a, k_2^a)$ is very close to $(n_1^o, n_2^o, k_1^o, k_2^o)$ for all prior distributions considered and the false positive and false negative error probabilities are generally controlled at or below the nominal levels α_1 and α_2 .

Numerical results for the three-stage case are in Table 4, where $n^o = n_1^o + n_2^o + n_3^o$ and $n^a = n_1^a + n_2^a + n_3^a$. Again, the asymptotic optimal procedure performs well even for moderate n . The probability of a false positive error is well controlled, while the probability of a false negative decision exceeds the nominal level in some cases by a slight amount. For a phase II study, such a possible minor increase in the false negative error rate seems worthwhile in exchange for a design procedure that reduces the costs with simplicity in the implementation. Also, Table 4 only shows values of \mathcal{N}_T^o , because our numerical results indicate that, unlike the one-stage and two-stage designs, \mathcal{N}^o and \mathcal{N}_T^o are very close for all three-stage scenarios listed in the table. Thus, the three-stage design is very efficient regarding sample size.

6 Prior Distributions

The optimal sample size is sensitive to the prior distribution and the target value of θ^* . In general, larger θ^* values lead to larger sample sizes and larger prior variances lead to smaller sample sizes. Additionally, the shape of the prior density function has significant impact. A U-shape density leads to a smaller sample size, compared to L-shaped and bell-shaped (i.e., unimodal) densities. For instance, assuming a beta prior with mean 0.3, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, and target value $\theta^* = 0.6$, the optimal sample size $n^o = 23$ if $var(\theta) = 0.12$, $n^o = 46$ if $var(\theta) = 0.1$, and $n^o = 82$ if $var(\theta) = 0.08$. The density with $var(\theta) = 0.08$ has

an L-shape, while the other two are U-shaped.

6.1 Caution with the prior distributions with a U-shaped density.

Consider two prior distributions with common prior mean. A naive way to think about one prior variance being larger than the other is that the prior distribution with the smaller variance is more informative. Such a reading can be misleading, however, when the support of the prior distribution is bounded, as is the case studied in this paper (i.e., $0 < \theta < 1$). Comparing a bell-shaped density to a flatter density, the one with smaller variance is more informative. However, comparing a flat density to a U-shaped density, the one with larger variance, the U-shaped density, is more informative, as in the following example.

EXAMPLE 2. Let $f_1(\theta)$ and $f_2(\theta)$ be the beta(1, 1) and beta(0.05, 0.05) density functions, respectively. The common prior mean is 0.5 but the prior variances are 0.083 and 0.227, respectively. After one observation, we have the following posterior means.

$$\begin{aligned} E(\theta|X = 0, f_1) &= 0.3333 & E(\theta|X = 1, f_1) &= 0.6667 \\ E(\theta|X = 0, f_2) &= 0.0455 & E(\theta|X = 1, f_2) &= 0.9545 \end{aligned}$$

We see a dramatic difference in the posterior means after one observation when f_2 is the prior distribution, while the difference under f_1 is much less dramatic. With U-shaped prior f_2 , $P(\theta \leq 0.1|f_2) = 0.45$ and $P(\theta \geq 0.9|f_2) = 0.45$. The f_2 prior proposes that θ is either close to zero or one. Therefore, the outcome of one single observation determines whether it is close to zero or one. In such a case, the optimal solution is $n^o = 1$ for $0.2 \leq \theta^* \leq 0.6$, $\alpha_1 = \alpha_2 = 0.1$. Thus, f_2 is more informative, despite having larger variance than f_1 , and can lead to smaller sample sizes.

Although we value prior information, we consider the prior distributions f_2 in Example 2 overly informative and to be avoided. For the same reason, we suggest caution when using prior distributions having U-shaped densities, especially if the prior variance is above 0.12, and recommend one compensate by reducing the error rate from 0.1 to 0.05, or to some value between 0.05 and 0.1 if the optimal sample size is too small. (We note that the prior variance for the often-used Jeffrey's prior is 0.125.)

6.2 Non-beta prior distributions

Other than being computationally straightforward, the asymptotic procedure has the advantage that it can deal with non-beta prior distributions as easily as beta distributions. For this procedure, $P(\theta > \theta^*)$ and $E(\theta)$ in $P_n(\widehat{E}_1)$ and $P_n(\widehat{E}_1)$ may need integration. All other terms in these formulas are explicit and algebraic.

7 Discussion

We have proposed asymptotic optimal designs for deciding the sample size for a sequence phase II clinical studies that evaluate treatments to discover one that shows promise. We discussed trials with fixed sample sizes and also trials having r stages, where $r = 2$, or 3 , at which the trial could stop to abandon evaluating a treatment not showing promise. Although our asymptotic optimal procedure can be extended in a straightforward manner to r -stage designs for general r , the further reduction in overall expected sample size is practically negligible beyond three stages.

Our results can serve two purposes in practice. First, it provides an explicit procedure for finding asymptotic optimal sample sizes for r -stage design with $r = 1, 2, 3$. Also, we showed analytically that the maximum optimal sample size for each study is asymptotically the same, regardless of the number of stages. In other words, $n^o = n_1^o + \dots + n_r^o$ is approximately the same as r varies. This result helps to narrow the design space when searching for the optimal design. When considering a multi-stage design, therefore, a practical way to find the optimal design is as follows. First, start by finding the optimal sample size for a 1-stage design, $n^o(1\text{-stage})$. Then, search for the optimal sample size n^o for an r -stage design in the range of $n^o(1\text{-stage}) - 5$ to $n^o(1\text{-stage}) + 5$. If $n^o(1\text{-stage})$ is less than ten, a three-stage design is not necessary, and a two-stage design might not make much difference, either.

We also found that $k_r^a/n^a = \theta^* + o(1)$. This result reduces the dimension of the design space if one wishes to find the exact optimal sample size. In practice, we can add the constraint

$$\lfloor n\theta^* \rfloor - 2 \leq k_r \leq \lfloor n\theta^* \rfloor + 2$$

to the design space when searching for the exact optimal sample size, reducing the compu-

tation time substantially for multi-stage design.

Previous research in this area has mainly focused on one-stage and two-stage designs for the reason stated by McPherson (1982), Mariani and Marubini (1996), and Yao and Venkatraman (1998), namely, that most of the reduction in the expected total sample size in a multi-stage design is achieved with two stages. Our numerical studies indicate that the optimal sample size is usually small when the prior density is U-shaped with variance greater than 0.1 (e.g., Jeffrey's prior) and $\theta^* \geq E(\theta)$. Therefore, further reduction in expected sample size from a two-stage design to one with three or more stages is likely to be minor. We discussed the impact of the shape of the prior on the optimal sample size in Section 6. For many other situations, however, a reduction in the expected sample size can be substantial by increasing the number of stages and should still be considered in practice.

We used a Bayesian formulation when setting up the design problem by averaging over the parameter space for θ using prior information. We have not discussed designs that update the prior to the posterior distribution when entering the next stage as would other Bayesian group sequential designs (e.g., Berry, 1987; Freedman & Spiegelhalter, 1989). Further study into reformulating the design in terms of the probabilities of the two types of error and the overall sample size to be minimized (loss function) is ongoing for the design in which the posterior will replace the prior for the next stage.

ACKNOWLEDGMENT

This research was done when the first author is visiting the Department of Biostatistics, M. D. Anderson Cancer Center, The University of Texas .

APPENDIX

Since the proof of Proposition 1 uses a result from Proposition 2, we prove Proposition 2 first.

PROOF OF PROPOSITION 2. We only prove the first expression of (2.5). The second expression can be proved by a similar argument. Let $I_n = \int_0^{\theta^*} P(X > k|\theta, n)f(\theta)d\theta$ and

$k = \lfloor n\theta^* \rfloor$. For any given $\delta > 0$,

$$\int_0^{\theta^* - \delta} P(X > k|\theta, n) f(\theta) d\theta = o(n^{-r})$$

for any $r > 0$. Since $(X - n\theta)/[\theta(1 - \theta)]^{1/2}$ converges weakly to the standard normal distribution, we have

$$\begin{aligned} I_n &= \int_{\theta^* - \delta}^{\theta^*} P(X > k|\theta, n) f(\theta) d\theta + o(n^{-r}) \\ &= f(\xi)[1 + o(1)] \int_{\theta^* - \delta}^{\theta^*} \left[1 - \Phi \left(\frac{k - n\theta}{[n\xi(1 - \xi)]^{1/2}} \right) \right] f(\theta) d\theta + o(n^{-r}), \end{aligned}$$

where ξ is between $\theta^* - \delta$ and θ^* . Let $t = \sqrt{n}(\theta^* - \theta)/[\xi(1 - \xi)]^{1/2}$.

$$I_n = \frac{[\xi(1 - \xi)]^{1/2} f(\xi)}{n^{1/2}} (1 + o(1)) \int_0^1 [1 - \Phi(t)] dt = \frac{[\theta^*(1 - \theta^*)]^{1/2} f(\theta^*)}{(2\pi n)^{1/2}} + o(1/n^{1/2}).$$

The last equality follows, since $\int_0^\infty [1 - \Phi(t)] dt = 1/(2\pi)^{1/2}$ and ξ converges to θ^* as δ goes to 0. \square

PROOF OF PROPOSITION 1. For a given target value θ^* and a prior density function $f(\theta)$, n goes to infinity only if α_1 and α_2 go to zero. Let

$$\liminf_{\alpha_1, \alpha_2 \rightarrow 0} \frac{k^o}{n^o} = \eta.$$

There exists a convergent subsequence of $\{k^o/n^o\}$ with limit equal to η . Assume that $\eta < \theta^*$. According to Proposition 2, for such a subsequence,

$$P(E_1) = \frac{\int_0^{\theta^*} P(X > k|\theta, n) f(\theta) d\theta}{\int_0^1 P(X > k|\theta, n) f(\theta) d\theta} = \frac{P(\eta < \theta < \theta^*)}{P(\theta > \eta)} + o(1/n^{1/2}),$$

which contradicts to the fact that α_1 goes to zero. Therefore,

$$\liminf_{\alpha_1, \alpha_2 \rightarrow 0} \frac{k^o}{n^o} \geq \theta^*.$$

By an analogous argument, we also have

$$\limsup_{\alpha_1, \alpha_2 \rightarrow 0} \frac{k^o}{n^o} \leq \theta^*. \quad \square$$

PROOF OF PROPOSITION 3. We only prove the second expression in (3.3). The other two expressions can be proved by similar arguments.

$$P_{++} = \int_{\theta^*}^1 P(X_1 > k_1 | \theta, n_1) f(\theta) d\theta - \int_{\theta^*}^1 P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta, n_1, n_2) f(\theta) d\theta,$$

where

$$\int_{\theta^*}^1 P(X_1 > k_1 | \theta, n_1) f(\theta) d\theta = \int_{\theta^*}^1 f(\theta) d\theta - \int_{\theta^*}^1 P(X_1 \leq k_1 | \theta, n_1) f(\theta) d\theta = P(\theta > \theta^*) + o(n_1^{-r})$$

for any $r > 0$ since $k_1/n_1 = \lambda\theta^* + o(n_1^{-1/2})$ and

$$\begin{aligned} & \int_{\theta^*}^1 P(X_1 > k_1, X_1 + X_2 \leq k_2 | \theta, n_1, n_2) f(\theta) d\theta \\ &= \int_{\theta^*}^1 P(X_1 + X_2 \leq k_2 | \theta, n_1, n_2) f(\theta) d\theta - \int_{\theta^*}^1 P(X_1 \leq k_1, X_1 + X_2 \leq k_2 | \theta, n_1, n_2) f(\theta) d\theta \\ &= \frac{g(\theta^*)}{(n_1 + n_2)^{1/2}} + o((n_1 + n_2)^{-1/2}) + o(n_1^{-r}) \end{aligned}$$

for any $r > 0$. Thus (3.3) is proved. (3.4) is a direct result from (2.3) and (3.3). \square

The proof of Proposition 2 is established for an expansion to one dominant term. The following lemma extends it further to a two-term Edgeworth expansion.

LEMMA FOR PROPOSITION 4. *Assume that the prior distribution has a continuously differentiable density function $f(\theta)$ with $f(\theta_0) > 0$, $0 < \theta_0 < 1$. Let $X \sim B(n, \theta)$ and $k = \lfloor n\theta_0 \rfloor$. Then*

$$J_n \triangleq \int_0^1 P(X > k | \theta, n) f(\theta) d\theta \approx P(\theta > \theta_0) + \frac{h(\theta_0)}{n} + o(n^{-1}),$$

where $h(\theta_0) = \theta_0 f(\theta_0) - f'(\theta_0) \sigma^2(\theta_0) / 2$.

PROOF. Let $\sigma^2(\theta) = \theta(1 - \theta)$, we have

$$J_n = P(\theta > \theta_0) + \int_{\theta_0 - \delta}^{\theta_0} P(X > k | \theta, n) f(\theta) d\theta - \int_{\theta_0}^{\theta_0 + \delta} P(X \leq k | \theta, n) f(\theta) d\theta + o(n^{-r})$$

for any $r > 0$. By Edgeworth expansion (Bhattacharya & Rao, 1976),

$$P(X \leq k | \theta, n) = \Phi(z_{n,\theta}) + \left[w(\theta, z_{n,\theta}) + \frac{(1 - 2\theta)(1 - z_{n,\theta}^2)}{6} \right] \frac{\phi(z_{n,\theta})}{n^{1/2} \sigma(\theta)} [1 + o(1)],$$

where $z_{n,\theta} = n^{1/2}(k/n - \theta)/\sigma(\theta)$ and $-0.5 \leq w(\theta, z_n, \theta) \leq 0.5$. Here, $w(\theta, z_n, \theta)$ represents the rounding error because of the discreteness of the problem. Using an argument similar to the proof of Proposition 2,

$$\int_{\theta_0}^{\theta_0+\delta} \frac{(1-2\theta)(1-z_{n,\theta}^2)}{6} \frac{\phi(z_{n,\theta})}{n^{1/2}\sigma(\theta)} f(\theta) d\theta = \frac{(1-2\theta_0)f(\theta_0)}{6n} \int_0^\infty (1-t^2)\phi(t) dt + o(n^{-1}) = o(n^{-1}),$$

since $\int_0^\infty (1-t^2)\phi(t) dt = 0$. Furthermore, let $t = n^{1/2}(\theta - \theta_0)/\sigma(\theta_0)$ and $\Delta_n = k - n\theta_0$. Note that $\Phi(-t) = 1 - \Phi(t)$ and $\phi(-t) = \phi(t)$. Applying a Taylor expansion,

$$\begin{aligned} \Phi(z_{n,\theta}) &\approx \Phi(-t) + \phi(-t)(z_{n,\theta} + t) = 1 - \Phi(t) + \phi(t) \left[\frac{\Delta_n}{n^{1/2}\sigma(\theta)} - n^{1/2}(\theta - \theta_0) \left(\frac{1}{\sigma(\theta)} - \frac{1}{\sigma(\theta_0)} \right) \right] \\ &= 1 - \Phi(t) + \phi(t) \left[\frac{\Delta_n}{n^{1/2}\sigma(\theta_0)} + \frac{\sigma'(\theta_0)t^2}{n^{1/2}} + O(|\theta - \theta_0|/n^{1/2}) + O(|\theta - \theta_0|^3/n^{1/2}) \right], \end{aligned}$$

and $f(\theta) \approx f(\theta_0) + f'(\theta_0)(\theta - \theta_0) + O(|\theta - \theta_0|)$. Since $\sigma'(\theta_0) = (1 - 2\theta_0)/(2\sigma(\theta_0))$ and $d\theta = (\sigma(\theta_0)/n^{1/2})dt$, by approximating the bounded quantity $w(\theta, z_n, \theta)$ by its mid-range 0, we have

$$\begin{aligned} \int_{\theta_0}^{\theta_0+\delta} P(X \leq k|\theta, n) f(\theta) d\theta &\approx \int_{\theta_0}^{\theta_0+\delta} \Phi(z_{n,\theta}) f(\theta) d\theta \\ &= \frac{\sigma(\theta_0)}{n^{1/2}} \int_0^\infty \left\{ 1 - \Phi(t) + \phi(t) \left[\frac{\Delta_n}{n^{1/2}\sigma(\theta_0)} + \frac{\sigma'(\theta_0)t^2}{n^{1/2}} \right] \right\} \left(f(\theta_0) + \frac{f'(\theta_0)\sigma(\theta_0)t}{n^{1/2}} \right) dt + o(n^{-1}) \\ &= \frac{g(\theta_0)}{n^{1/2}} + \frac{1}{n} \left[\frac{\Delta_n f(\theta_0)}{2} + \frac{(1-2\theta_0)f(\theta_0)}{4} + \frac{f'(\theta_0)\sigma^2(\theta_0)}{4} \right] + o(n^{-1}) \end{aligned}$$

where $g(\cdot)$ is defined in (2.6). By a similar argument,

$$\int_{\theta_0-\delta}^{\theta_0} P(X > k|\theta, n) f(\theta) d\theta \approx \frac{g(\theta_0)}{n^{1/2}} - \frac{1}{n} \left[\frac{\Delta_n f(\theta_0)}{2} + \frac{(1-2\theta_0)f(\theta_0)}{4} + \frac{f'(\theta_0)\sigma^2(\theta_0)}{4} \right] + o(n^{-1}).$$

Approximating $\Delta_n = \lfloor n\theta_0 \rfloor - n\theta_0$ by its mid-range 0, we obtain

$$J_n \approx P(\theta > \theta_0) + \frac{1}{n} [\theta_0 f(\theta_0) - f'(\theta_0)\sigma^2(\theta_0)/2]. \quad \square$$

PROOF OF PROPOSITION 4. In light of (3.5), (3.6) and the above lemma,

$$\mathcal{N}(n_1, n_2, k_1, k_2) = \frac{n_1 + n_2(1 - P_{et})}{p_+} \approx \frac{n_1 + n_2[P(\theta > \theta_1) + h(\theta_1)/n_1]}{P(\theta > \theta^*)}.$$

For each fixed value of $n_1 + n_2$, a straightforward calculus manipulation shows that the numerator of the right-hand side of the above equation is minimized at

$$n_1^a \approx \left[\frac{h(\theta_1)}{P(\theta < \theta_1)} \right]^{1/2} (n_1^a + n_2^a)^{1/2}. \quad \square$$

The proofs of Propositions 5 and 6 are similar to the proofs of Propositions 3 and 4.

REFERENCES

- Bergman, S. W. & Gittins, J. C. (1985). *Statistical Methods for Planning Pharmaceutical Research*. New York: Marcel Dekker.
- Berry, D. A. (1987). Interim analysis in clinical trials: The role of the likelihood principle. *The Amer. Statistician* **41**, 117-122.
- Freedman, L. S. and Spiegelhalter, D. J. (1989). Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Contr. Clin. Trials* **10** 357-367.
- Friedman, L., Furberg, C., and Demets, D. (1998). *Fundamentals of Clinical Trials, third edition*, New York, N.Y.: Springer-Verlag New York., Inc.
- Mariani, M. and Marubini, E. (1996) Design and analysis of phase II cancer trials: A review of statistical methods and guidelines for medical researchers. *International Statistical Review* **64**, 61-88.
- McPherson, K. (1982). On choosing the number of interim analyses in clinical trials. *Statistics in Medicine* **1**, 25-36.
- Strauss, N. & Simon, R. (1995). Investigating a sequence of randomized phase-II trials to discover promising treatments. *Statistics in Medicine* **14**, 1479-1489.
- Wang, Y. G. and Leung, D. H. Y. (1998). An optimal design for screening trials. *Biometrics* **54**, 243-250.
- Whitehead J. (1985). Designing phase II studies in the context of a programme of clinical research. *Biometrics* **41**, 373-383.
- Wang, Y. G. and Leung, D. H. Y. (2001). Optimal design for evaluating a series of treatments. *Biometrics* **57**, 168-171.
- Yao, T., Begg, C. B., & Liveingston P. O. (1996). Optimal sample size for a series of pilot trials of new agents. *Biometrics* **52**, 992-1001.
- Yao, T. & VENKATRAMAN E. S. (1998). Optimal two-stage design for a series of pilot trials of new agents. *Biometrics* **54**, 1183-1189.

Table 2

Comparison between true and asymptotic optimal solutions with their realized probabilities of false positive and false negative for one-stage design.

α_1/α_2	θ^*	$E(\theta)$	$\text{var}(\theta)$	True optimal design			Asymptotic optimal design			
				k^o/n^o	\mathcal{N}^o	\mathcal{N}_T^o	k^a/n^a	$P(E_1)$	$P(E_2)$	
0.1/0.1	0.3	0.2	0.08	4/15	58.2	48.3	4/15	0.097	0.083	
			0.10	2/8	33.0	27.4	2/8	0.084	0.089	
		0.3	0.08	5/18	44.6	39.2	5/18	0.089	0.093	
	0.10		3/12	30.2	26.6	3/12	0.098	0.078		
	0.12		2/8	21.7	18.8	2/8	0.080	0.090		
	0.4	0.4	0.08	3/12	21.2	19.7	4/15	0.083	0.087	
			0.10	3/12	22.2	20.4	3/11	0.071	0.100	
		0.12	2/8	16.0	14.4	2/8	0.075	0.092		
	0.6	0.2	0.08	13/22	166.7	90.7	13/22	0.099	0.099	
			0.10	7/12	77.6	42.4	7/12	0.080	0.095	
		0.3	0.08	24/40	219.5	133.7	22/37	0.100	0.095	
			0.10	13/22	103.2	63.3	13/22	0.090	0.095	
			0.12	7/12	51.4	31.5	7/12	0.084	0.099	
		0.4	0.4	0.08	22/37	136.5	94.2	22/37	0.096	0.094
				0.10	13/22	74.1	50.8	13/22	0.090	0.097
0.12			8/14	43.2	29.9	8/14	0.093	0.085		
0.5		0.5	0.08	16/27	69.4	53.0	14/24	0.101 [†]	0.091	
	0.10		10/17	41.9	31.6	8/14	0.100	0.094		
.05/0.15	0.3	0.2	0.08	6/18	78.8	61.3	6/18	0.044	0.143	
			0.10	3/9	41.3	31.5	3/9	0.080	0.089	
		0.3	0.08	8/24	66.3	55.9	8/24	0.040	0.145	
	0.10		5/15	43.1	35.6	5/15	0.041	0.142		
	0.12		3/9	27.1	22.0	3/9	0.040	0.145		
	0.4	0.4	0.08	6/18	36.1	32.2	6/18	0.038	0.149	
			0.10	5/15	31.3	27.5	4/13	0.050	0.120	
		0.12	3/9	19.9	17.1	3/9	0.038	0.147		
	0.6	0.2	0.08	19/30	251.4	126.1	22/35	0.048	0.129	
			0.10	9/14	99.5	48.7	9/14	0.043	0.145	
		0.3	0.10	19/30	154.5	88.4	19/30	0.045	0.137	
			0.12	11/17	80.3	44.6	9/14	0.045	0.151 [†]	
		0.4	0.4	0.10	19/30	111.0	71.7	19/30	0.045	0.140
				0.12	12/19	64.6	41.5	12/19	0.093	0.085
		0.5	0.5	0.08	22/35	98.2	71.8	22/35	0.048	0.138
0.10				14/22	59.4	42.4	12/19	0.050	0.144	
0.12			9/14	36.6	25.6	9/14	0.040	0.146		

†: Where $P(E_i)$ exceeds α_i , $i = 1, 2$.

Table 3

Comparison between true and asymptotic optimal solutions with their realized probabilities of false positive and false negative with $n = n_1 + n_2$ for two-stage design.

α_1/α_2	θ^*	$E(\theta)$	$v(\theta)$	True optimal design				Asymptotic optimal design					
				k_1^o/n_1^o	k_2^o/n^o	\mathcal{N}^o	\mathcal{N}_T^o	k_1^a/n_1^a	k_2^a/n^a	$P(E_1)$	$P(E_2)$		
.1/.1	.3	0.2	.08	0/5	4/15	35.0	34.1	0/5	4/15	.092	.095		
			.10	0/3	2/9	19.8	19.7	0/3	2/8	.075	.116 [†]		
		0.3	.08	0/5	5/19	32.8	31.8	0/6	5/18	.087	.099		
			.10	0/4	3/12	21.1	20.8	0/5	3/12	.095	.087		
			.12	0/4	2/8	16.2	16.0	0/4	2/8	.077	.099		
			.08	0/5	3/12	18.3	18.0	0/6	4/15	.081	.092		
	0.4	.10	0/4	2/9	13.7	13.6	0/5	3/11	.070	.106 [†]			
		.12	0/4	2/9	13.9	13.8	0/4	2/8	.072	.100			
		.6	0.2	.08	1/4	18/31	77.1	65.8	0/4	13/22	.098	.100	
				.10	0/2	8/14	33.1	30.1	0/3	7/12	.080	.097	
			0.3	.08	4/10	23/39	100.6	84.2	1/7	22/37	.100	.096	
				.10	1/4	15/26	54.3	47.7	1/5	13/22	.089	.100	
.12	0/2			8/14	28.6	25.7	0/4	7/12	.083	.099			
.08	3/8			24/41	81.5	72.2	1/8	22/37	.096	.095			
0.4	.10	2/6	14/24	47.8	41.8	1/6	13/22	.090	.098				
	.12	1/4	8/14	27.4	24.3	0/4	8/14	.093	.085				
	0.5	.08	2/6	15/26	44.5	40.7	1/6	14/24	.100	.093			
		.10	1/5	8/14	26.2	23.7	1/5	8/14	.099	.096			
.05/.1	.3	0.2	.08	1/8	9/30	62.4	58.9	0/6	9/29	.046	.099		
			.10	0/4	5/17	35.0	34.1	0/4	5/16	.038	.107 [†]		
		0.3	.08	1/10	11/36	63.3	60.6	0/9	11/35	.044	.100		
			.10	0/6	7/23	42.3	40.5	0/6	7/23	.047	.094		
			.12	0/4	5/17	28.7	27.9	0/5	5/16	.038	.099		
			.08	1/12	7/23	37.9	36.6	0/7	8/26	.045	.100		
		0.4	.10	1/7	6/21	30.3	29.5	0/7	7/22	.036	.102 [†]		
			.12	0/4	4/14	21.1	20.7	0/6	5/16	.037	.096		
			.6	0.2	.08	1/5	30/49	130.3	108.2	1/6	30/49	.051 [†]	.093
					.10	0/2	15/25	53.1	46.7	0/4	13/21	.045	.099
		0.3		.10	3/8	28/46	96.7	82.8	1/6	24/39	.050	.100	
				.12	2/6	14/23	50.5	43.1	1/5	14/23	.050	.091	
	0.4	.10		3/8	28/46	85.0	75.3	1/8	27/44	.049	.093		
		.12		2/6	17/28	50.8	45.1	1/6	16/26	.047	.096		
	0.5	0.2	.08	5/12	33/54	93.0	84.4	1/8	30/49	.051	.096		
			.10	2/6	20/33	53.9	49.0	1/7	19/31	.049	.095		
		0.3	.12	1/5	11/18	32.2	28.8	1/5	13/21	.042	.101 [†]		

†: Where $P(E_i)$ exceeds α_i , $i = 1, 2$.

Table 4

Comparison between true and asymptotic optimal solutions with their realized probabilities of false positive and false negative with $m = n_1 + n_2$, $n = n_1 + n_2 + n_3$ for three-stage design.

θ^*	$E(\theta)$	$v(\theta)$	True optimal design				Asymptotic optimal design				
			k_1^o/n_1^o	k_2^o/m^o	k_3^o/n^o	N_T^o	k_1^a/n_1^a	k_2^a/m^a	k_3^a/n^a	$P(E_1)/P(E_2)$	
$\alpha_1 = 0.1, \alpha_2 = 0.1$											
.3	.2	.08	0/4	1/9	4/16	31.2	0/4	1/9	4/15	.088/.110	
		.10	0/3	1/8	2/9	19.7	0/4	1/7	2/8	.080/.097	
	.3	.08	0/6	1/10	4/15	27.4	0/5	1/9	5/18	.083/.110	
		.10	0/4	1/9	3/12	20.6	0/5	1/8	3/12	.093/.090	
	.4	.08	0/5	1/9	3/12	17.8	0/5	1/8	4/15	.077/.106	
		.10	0/4	1/7	2/9	13.5	0/5	1/7	3/11	.067/.112	
.6	.2	.08	0/3	4/10	14/24	56.2	0/4	3/9	13/22	.098/.102	
		.10	0/2	1/4	6/11	24.6	0/3	2/7	7/12	.080/.097	
	.3	.08	1/5	7/16	23/39	77.7	1/5	4/11	22/37	.096/.108	
		.10	0/3	6/12	12/21	40.8	0/4	3/9	13/22	.089/.097	
		.4	.12	0/2	3/7	7/13	22.6	0/3	2/7	7/12	.083/.101
			.08	1/5	7/15	24/41	68.7	1/5	4/11	22/37	.092/.107
		.5	.10	1/5	6/12	12/21	36.8	1/5	3/9	13/22	.088/.103
			.12	1/4	6/11	7/13	22.3	0/4	2/7	8/14	.092/.086
			.08	1/5	6/14	14/24	39.2	0/4	3/8	14/24	.097/.103
			.10	0/3	3/8	8/14	22.9	0/4	2/7	8/14	.100/.095
	$\alpha_1 = .05, \alpha_2 = .1$										
	.3	.2	.08	0/5	2/13	9/30	54.1	0/6	2/12	9/29	.043/.112
.10			0/5	2/9	4/14	32.6	0/4	1/9	5/16	.038/.108	
	.3	.08	1/9	3/15	11/37	56.5	0/6	3/13	11/35	.041/.114	
		.10	0/5	2/11	7/24	37.1	0/5	2/11	7/23	.044/.108	
	.4	.12	0/4	2/10	4/14	23.6	0/4	1/9	5/16	.037/.109	
		.08	0/9	2/15	7/23	35.9	0/5	2/10	8/26	.039/.130	
		.10	1/9	4/15	5/17	26.8	0/6	2/10	7/22	.033/.118	
		.12	0/4	1/8	4/14	20.3	0/5	1/9	5/16	.036/.101	
.6	.2	.08	0/3	7/17	30/49	99.3	0/4	4/12	30/49	.051/.094	
		.10	0/2	3/8	15/25	43.8	0/3	3/8	13/21	.044/.104	
	.3	.10	0/3	7/15	28/46	77.2	0/4	4/11	24/39	.050/.103	
		.12	0/3	2/6	14/23	40.4	0/4	3/9	14/23	.051/.090	
		.4	.10	1/5	8/17	28/46	72.3	1/5	4/11	27/44	.047/.103
			.12	0/3	5/11	17/28	43.5	0/4	3/9	16/26	.047/.098
		.5	.08	2/9	10/21	30/49	78.9	1/5	4/11	30/49	.046/.127
			.10	1/5	5/11	20/33	47.8	1/5	3/9	19/31	.048/.102
			.12	0/3	3/7	12/21	28.7	0/4	3/8	13/21	.041/.102