STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:561–569 (DOI: 10.1002/sim.1600)

Admissible two-stage designs for phase II cancer clinical trials

Sin-Ho Jung[1,*,†] Taiyeong Lee[2], KyungMann Kim[3] and Stephen L. George[1]

[1]*Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, U.S.A.*
[2]*SAS Institute Inc., Cary, NC, U.S.A.*
[3]*Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, U.S.A.*

SUMMARY

In a typical two-stage design for a phase II cancer clinical trial for efficacy screening of cytotoxic agents, a fixed number of patients are initially enrolled and treated. The trial may be terminated for lack of efficacy if the observed number of tumour responses after the first stage is too small, thus avoiding treatment of patient with inefficacious regimen. Otherwise, an additional fixed number of patients are enrolled and treated to accumulate additional information on efficacy as well as safety. The minimax and the so-called 'optimal' designs by Simon have been widely used, and other designs have largely been ignored in the past for such two-stage cancer clinical trials. Recently Jung *et al.* proposed a graphical method to search for compromise designs with features more favourable than either the minimax or the optimal design. In this paper, we develop a family of two-stage designs that are admissible according to a Bayesian decision-theoretic criterion based on an ethically justifiable loss function. We show that the admissible designs include as special cases the Simon's minimax and the optimal designs as well as the compromise designs introduced by Jung *et al.* We also present a Java program to search for admissible designs that are compromises between the minimax and the optimal designs. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Bayes design; compromise design; minimax design; optimal design; sequential design

1. INTRODUCTION

In the early development of a new therapeutic agent, the dose and schedule to be used in subsequent trials is determined in a phase I trial. With traditional cytotoxic agents for cancer treatment, this dose is generally known as the maximum tolerated dose (MTD), although other choices, e.g. a dose level before the MTD, may be used. In a subsequent phase II study, patients are treated at the dose level established as safe in phase I trials to screen if the treatment has sufficiently promising clinical activity or efficacy, usually evaluated by the

*Correspondence to: Sin-Ho Jung, Department of Biostatistics and Bioinformatics, Duke University, Medical Center Box 3627, Durham, NC 27710, U.S.A.
†E-mail: jung0005@surgerytrials.duke.edu

Contract/grant sponsor: National Institutes of Health, Department of Health and Human Service, U.S. Government; contract/grant number: CA 52733

Received May 2002
Accepted May 2003

probability of tumour response, say treatment success, for further investigation in subsequent phase III studies.

In designing phase II studies, it is desirable to achieve the goals of the study with a minimal number of patients, especially if the treatment turns out to be inefficacious. To this end, multi-stage sequential designs in which a fixed number of patients are accrued in each stage have been proposed, and the study is stopped or continued depending on the observed number of treatment successes and failures [1, 2].

A first of this type of two-stage design was proposed by Gehan [3]: 14 patients are accrued during stage 1. The study stops if no responders are observed, or proceeds to stage 2 to accrue another cohort of patients if one or more responders are observed. The number of patients during stage 1, $n_1 = 14$, is designed to keep the probability of early termination very small, say < 0.05, when the treatment is active, e.g. with the probability of response $p \geqslant 0.20$, and stage 2 sample size is determined to achieve a specified level of confidence in estimating the probability of response with the maximum sample size. Later this notion of two-stage designs has been formalized as a test of statistical hypotheses regarding the probability of response, e.g. by Lee *et al.* [4] and Simon [1]. Also multi-stage designs have been formalized by Fleming [5] and Chang *et al.* [6]. These approaches were all based on the frequentist paradigm. Others have developed similar designs based on Bayesian paradigm [7, 8] or predictive probabilities [9].

In this paper, we develop a family of two-stage designs that are admissible according to a Bayesian decision-theoretic criterion based on an ethically justified loss function. We show that the admissible designs include as special cases the minimax and the Simon's optimal designs as well as the compromise designs introduced by Jung *et al.* [2]. We also present a Java program to search for the admissible designs. The rest of the paper is organized as follows: In Section 2, we review two-stage minimax and optimal designs by Simon [1] and compromise designs by Jung *et al.* [2] with illustrative examples. In Section 3, we develop admissible designs according to Bayesian decision theory with a loss function justified on ethical grounds. In Section 4, we describe two different ways to identify admissible designs and introduce a Java program which facilitates search for admissible designs given design parameters with examples. We close with a short discussion in Section 5.

2. TWO-STAGE DESIGNS

For ethical reasons, most clinical trials are required to have sequential designs. Yet, for practical reasons, they are usually conducted as multi-stage experiments, instead of being fully sequential. Two-stage designs are commonly used for phase II cancer clinical trials because of simplicity and diminishing returns beyond two stages.

A typical two-stage trial is conducted as follows. During stage 1, n_1 patients are enrolled and treated. If the number of responders is less than or equal to r_1, the trial is terminated for lack of efficacy and it is concluded that the treatment does not warrant further investigation. Otherwise, the study is continued to stage 2 during which an additional n_2 patients are enrolled and treated. If the cumulative number of responders after stage 2 does not exceed r, it is concluded that the treatment lacks sufficient efficacy. Otherwise, it is concluded that the treatment has sufficient activity, and the treatment will be considered for further investigation in subsequent trials. One may also employ an upper boundary to stop the trial early when a

significantly high efficacy is observed from stage 1 [6, 7]. However, there being no compelling ethical argument and thus rarely used, we consider early stopping only for lack of efficacy in this paper.

A two-stage design is defined by the number of patients to be accrued during stages 1 and 2, n_1 and n_2, and the boundary values r_1 and r ($r_1 < r$), so we denote any two-stage design by $(r_1/n_1, r/n)$ where $n = n_1 + n_2$, the maximum sample size. The values of $(r_1/n_1, r/n)$ are determined based on some pre-specified design parameters as described below. To be more specific, let p_0 denote the maximum unacceptable probability of response, and p_1 the minimum acceptable probability of response with $p_0 < p_1$. We want to test $H_0 : p \leqslant p_0$ against $H_1 : p \geqslant p_1$ with type I error probability α and power $1 - \beta$. These parameters $(p_0, p_1, \alpha, \beta)$ are termed design parameters.

Let $b(\cdot; p, m)$ and $B(\cdot; p, m)$ be the probability mass and distribution function, respectively, for the binomial distribution with probability of success p and number of trials m. For a two-stage design, the probability of rejecting the treatment, or equivalently accepting H_0, is expressed as

$$R(p) = B(r_1; p, n_1) + \sum_{x=r_1+1}^{\min(n_1, r)} b(x; p, n_1) B(r - x; p, n_2)$$

when the probability of response is p. Note that $B(r_1; p, n_1) = \mathrm{PET}(p)$ is the probability of early termination after stage 1. The constraints on type I error probability and power are expressed as $R(p_0) \geqslant 1 - \alpha$ and $R(p_1) \leqslant \beta$. Given $(p_0, p_1, \alpha, \beta)$, there are many two-stage designs $(r_1/n_1, r/n)$ satisfying the constraints.

Simon [1] proposed two criteria to select a good two-stage design among these designs. The minimax design minimizes the maximum sample size, n, among the designs satisfying the (α, β)-constraint. On the other hand, the so-called 'optimal' design minimizes the expected sample size EN under the null hypothesis determined by

$$\mathrm{EN} = \mathrm{PET}(p_0) \times n_1 + \{1 - \mathrm{PET}(p_0)\} \times n$$

The minimax and the optimal designs have both been widely used, and other designs have largely been ignored in the past for such two-stage cancer clinical trials. However, the two Simon's designs may result in highly divergent sample size requirements as shown in the example below. For example, the minimax design may have an excessively large EN as compared to the optimal design or the optimal design may have an excessively large maximum sample size n as compared to the minimax design. This results from the discrete nature of the binomial problem.

Example 1

For the design parameters $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$, the minimax design is given by $(r_1/n_1, r/n) = (2/18, 5/27)$ and the optimal design by $(1/11, 6/35)$. The maximum sample size n for the minimax design is eight less than that for the optimal design. However, the expected sample size EN under H_0 for the optimal design is 18.3 which is only slightly smaller than $\mathrm{EN} = 20.4$ for the minimax design.

Simon's designs are to some extent mathematical niceties. Also, as indicated above, the minimax and optimal designs can be quite different. Often, practical compromises are possible without changing the statistical operating characteristics appreciably. To avoid these

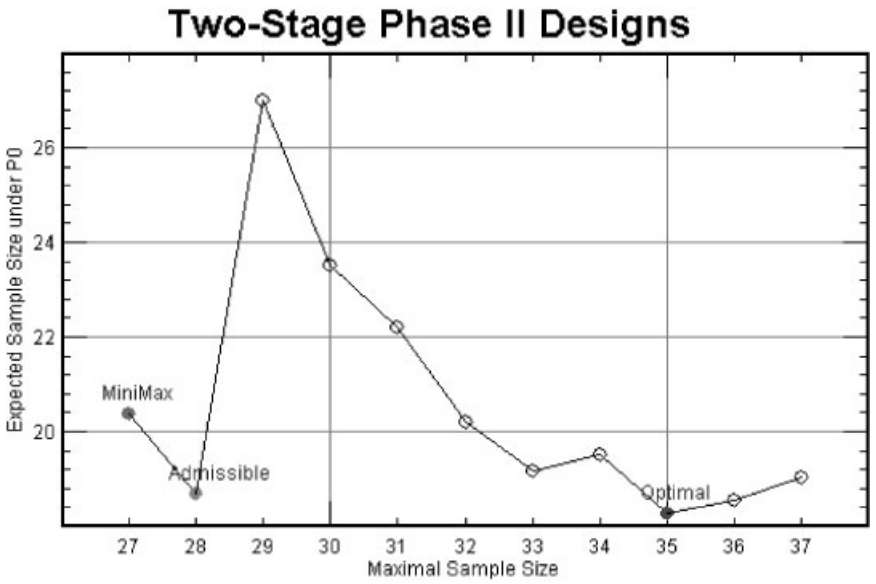

Figure 1. Two-stage designs for $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$ with $N = 37$.

discrepancies in the maximum sample size and the expected sample size under the null hypothesis between the minimax and the optimal designs, Jung *et al.* [2] propose a heuristic graphical method to search for compromise designs, neither minimax nor optimal, but with more desirable and practically appealing features.

Example 1 (*revisited*)
For the same design parameters $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$, the design given by $(r_1/n_1, r/n) = (1/13, 5/28)$ requires only one more patient in the maximum sample size n than the minimax design but its expected sample size EN under H_0 is very comparable to that of the optimal design (18.7 vs 18.3). Jung *et al.* [2] recommend this design as a good compromise between the minimax design and the optimal design.

For a given maximum sample size n, the designs satisfying the (α, β)-constraint can be determined by an exhaustive enumeration. This can be achieved readily by changing n_1 $(= 1, \ldots, n-1)$, r_1 and r $(0 \leqslant r_1 \leqslant r \wedge n_1;\ r_1 \leqslant r \leqslant r_1 + n - n_1)$. From these designs, the one that minimizes EN is determined. This design dominates all other designs for the given n. So, our search procedure for a good design will go through only these dominating designs within a range of n values. We will call them candidate designs. If n is too small, there may exist no designs satisfying the (α, β)-constraint, and as a result no candidate design either. This process is repeated by increasing n by 1 each time until an arbitrary upper limit, say N. Typically, N may represent the number of available subjects that can be accrued in a reasonable time period.

We have developed a program that plots EN of the candidate designs against n given the design parameters $(p_0, p_1, \alpha, \beta)$ and N. The plot starts with Simon's minimax design and ends with $n = N$. From the plot, the design minimizing EN within the range can be easily identified

Table I. Two-stage admissible designs for $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$.

$(r_1/n_1, r/n)$	EN	α	$1-\beta$	PET(p_0)	PET(p_1)	q
$(2/18, 5/27)$	20.4	0.0444	0.851	0.734	0.060	[0.632,1]
$(1/13, 5/28)$	18.7	0.0498	0.858	0.621	0.063	[0.057,0.632]
$(1/11, 6/35)$	18.3	0.0422	0.851	0.697	0.113	[0,0.057]

The first and the third designs are Simon's minimal and optimal designs, respectively.

and is marked as 'optimal'. When N is large enough, this local optimal design is the Simon's optimal design. This program is written in Java language and is thus platform independent.

Figure 1 shows the plot of EN against n for $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$ and $N = 37$ discussed in Example 1. Simon's minimax design is given by $(r_1/n_1, r/n) = (2/18, 5/27)$ and optimal design by $(1/11, 6/35)$, which is also the Simon's optimal design. The program provides specification of a design $(r_1/n_1, r/n)$ along with EN, PET(p_0), PET(p_1), and exact type I error probability and power when the circle representing the candidate design, actually (n, EN), is clicked with a pointer. Table I summarizes the operating characteristics of various designs.

3. ADMISSIBLE DESIGNS

In this section, we will use formal statistical criteria to define a class of admissible designs according to which compromise designs between the minimax and the optimal designs as proposed by Jung *et al.* [2] can be justified. In order to identify appropriate designs, we need to specify $(p_0, p_1, \alpha, \beta)$ and N.

Given $(p_0, p_1, \alpha, \beta)$ and N, let $\mathscr{D}$ denote the space of all candidate designs with $n \leqslant N$ satisfying the (α, β)-constraint. We consider two outcomes $\omega_1 = n$ and $\omega_2 = \mathrm{EN}$ from each design $d \in \mathscr{D}$. We use notations $n(d)$ and $\mathrm{EN}(d)$, in place of n and EN, to relate each design to its outcomes. Define a loss function

$$L(\omega, d) = n(d) I(\omega = \omega_1) + \mathrm{EN}(d) I(\omega = \omega_2)$$

in $\Omega \times \mathscr{D}$, where $\Omega = \{\omega_1, \omega_2\}$ and $I(\cdot)$ is an indicator function. This loss function is justified on the ethical grounds that it is desirable to minimize both the maximum number and the expected number of patients under the null hypothesis in two-stage phase II cancer clinical trials.

Let Q be a probability distribution defined over Ω as $Q(\omega = \omega_1) = q$ and $Q(\omega = \omega_2) = 1 - q$ for $q \in [0, 1]$. For any design $d \in \mathscr{D}$, the expected loss, or risk, is defined by

$$\rho(Q, d) = \int_{\Omega} L(\omega, d)\, \mathrm{d}Q(\omega) = q \times n(d) + (1 - q) \times \mathrm{EN}(d)$$

By considering only the designs with $n \leqslant N$ with N prespecified, $\rho(Q, d)$ is finite for every $d \in \mathscr{D}$. For a probability distribution Q, the Bayes risk is defined by

$$\rho^*(Q) = \inf_{d \in \mathscr{D}} \rho(Q, d) \tag{1}$$

Any design $d^* \in \mathscr{D}$ whose risk equals the Bayes risk is called a Bayes design against the distribution Q under the specified loss function. Note that the minimax design is a Bayes design against Q with $q=1$ and the Simon's optimal design is a Bayes design against Q with $q=0$. Since Q is uniquely defined by a constant $q \in [0,1]$, we may use q and Q interchangeably.

A design d^* is admissible if it is a Bayes design against a distribution Q. Equivalently, a design $d_0 \in \mathscr{D}$ is inadmissible if it is not a Bayes design for any choice of $q \in [0,1]$, i.e. there exists $d_q \in \mathscr{D}$ such that, for some $q \in [0,1]$,

$$\rho(q, d_0) > \rho(q, d_q)$$

For $d_1, d_2 \in \mathscr{D}$, it is said that d_1 dominates d_2 if $n(d_1) \leqslant n(d_2)$ and $\mathrm{EN}(d_1) < \mathrm{EN}(d_2)$ or $n(d_1) < n(d_2)$ and $\mathrm{EN}(d_1) \leqslant \mathrm{EN}(d_2)$. In this case, d_2 cannot be an admissible design.

This approach can be easily modified to handle any number of stages and different loss functions. Unlike Bayesian multi-stage designs or designs based on predictive probabilities which have to assign a prior probability to the probability of success, our method assumes that the probability of success is fixed. Instead, we combine some existing optimality criteria, i.e. assign prior probabilities to the criteria, and identify admissible designs under the criteria defined by various combinations of the existing criteria.

4. SEARCH FOR ADMISSIBLE DESIGNS

Now suppose that we want to find a Bayes design against the distribution with a specified q in $[0,1]$ according to the derivation given in the previous section. There are two ways to identify admissible designs.

The first approach is to consider a straight line $q \times n + (1-q) \times \mathrm{EN} = \rho$ determined by ρ, on the (n, EN)-plane, i.e. a line with slope $-q/(1-q)$ and intercept $\rho/(1-q)$. Starting from a small ρ, we move the straight line upward until it touches a design. The first design touched by the line is a Bayes design with Bayes risk ρ^*, where $\rho^*/(1-q)$ is the intercept of the straight line when it touches the Bayes design.

Suppose that we choose $q = \frac{1}{2}$ for Example 1. Then, from Figure 1, design $(r_1/n_1, r/n) = (1/13, 5/28)$ is a unique Bayes design. Noting that $n=28$ and $\mathrm{EN}=18.7$ for this design, we obtain Bayes risk $\rho^* = q \times n(d) + (1-q) \times \mathrm{EN}(d) = 23.3$. Again Table I shows the two-stage designs discussed above, i.e. Simon's minimax and optimal designs and a compromise design by Jung *et al.* [2] for the design parameters $(p_0, p_1, \alpha, \beta) = (0.1, 0.3, 0.05, 0.15)$ in Example 1. It also summarizes with their operating characteristics such as the expected sample size EN under the null hypothesis, exact type I error probability and power, the probability of early termination both under the null and alternative hypotheses, and the distribution Q specified by q against which the design is admissible.

In designing a phase II study, $q \in [0,1]$ may be chosen depending on the relative importance between n and EN. For example, if the study is on a rare disease so that the accrual is very low, then we may choose a larger q to give more favour towards the minimax design. On the other hand, if the accrual is not a problem, but we want to stop the study as early as possible when the treatment is inactive, then we may choose a small q to give more favour towards the optimal design. Whichever is the case, given $(p_0, p_1, \alpha, \beta)$ and N, a design may be regarded as a good one if it is a Bayes design over a wide range of q in $[0,1]$. In Figure 1, the compromise design $(r_1/n_1, r/n) = (1/13, 5/28)$, which is admissible, can be identified by

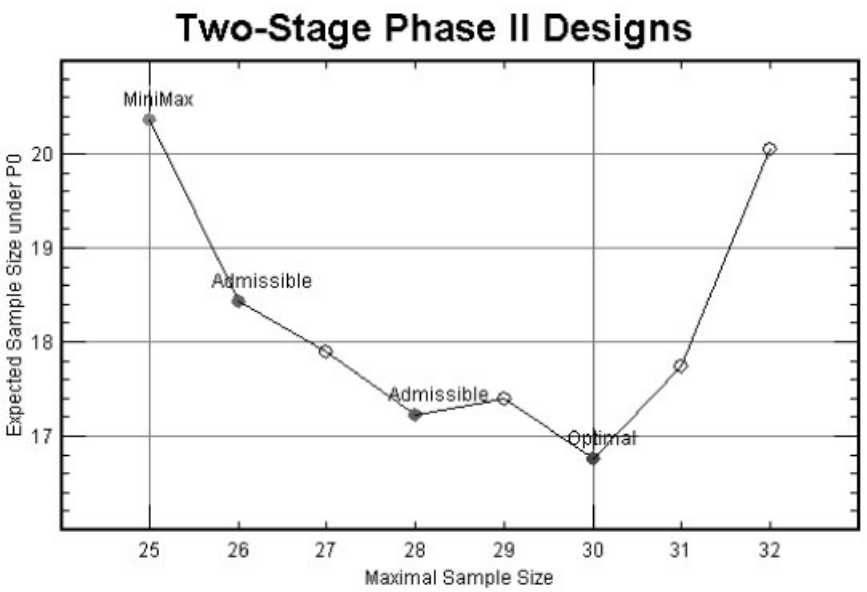

Figure 2. Two-stage designs for $(p_0, p_1, \alpha, \beta) = (0.05, 0.25, 0.05, 0.9)$ with $N = 32$.

any straight line with slope, $-q/(1-q)$, between -1.72 and -0.06, i.e. for $q \in [0.057, 0.632]$. Similarly we can show that Simon's minimax and optimal designs are Bayes designs for $q \in [0.632, 1]$ and $q \in [0, 0.057]$, respectively.

The second approach is to consider a convex hull formed by connecting candidate designs between the Simon's minimax design and the optimal design. According to DeGroot [10, pp. 125–127], any design on the convex hull is admissible. We have implemented this procedure in our Java program so that other admissible designs, besides Simon's minimax and optimal designs, can be automatically identified.

In Figure 1, the design given by $(r_1/n_1, r/n) = (1/13, 5/28)$ is the only admissible design except the minimax and the optimal designs. Figure 2 shows candidate two-stage designs for $(p_0, p_1, \alpha, \beta) = (0.05, 0.25, 0.05, 0.1)$ with $N = 32$. Admissible designs are highlighted at $n = 25$ (Simon's minimax design), 26, 28 and 30 (Simon's optimal design). The designs with $n = 26$ and 28 are admissible for $q \in [0.377, 0.660]$ and $q \in [0.187, 0.377]$, respectively. Table II summarizes admissible two-stage designs identified in Figure 2 along with their operating characteristics.

5. DISCUSSION

Phase II clinical trials form a critical step in developing treatment for cancer by screening new treatment for efficacy. These clinical trials are conducted typically as a two-stage design in order to avoid giving patients with advanced stage disease possibly inactive treatment. Statistical tests of the null and alternative hypotheses regarding the efficacy of treatment have

Table II. Two-stage admissible designs for $(p_0, p_1, \alpha, \beta) = (0.05, 0.25, 0.05, 0.1)$.

$(r_1/n_1, r/n)$	EN	α	$1-\beta$	PET(p_0)	PET(p_1)	q
$(0/15, 3/25)$	20.4	0.0336	0.901	0.463	0.013	[0.660,1]
$(0/12, 3/26)$	18.4	0.0365	0.905	0.540	0.032	[0.377,0.660]
$(0/10, 3/28)$	17.2	0.0426	0.906	0.599	0.056	[0.187,0.372]
$(0/9, 3/30)$	16.8	0.0489	0.902	0.630	0.075	[0,0.187]

The first and the last designs are Simon's minimal and optimal designs, respectively.

proven to be useful in determining efficient designs given the desirable statistical operating characteristics in terms of the type I and II error probabilities.

In this paper, we proposed a family of two-stage designs that are admissible according to a Bayesian decision-theoretic criterion. It was based on a loss function which is a weighted average of the maximum sample size and the expected sample size under the null hypothesis, the two criteria used by Simon [1]. We showed that both the minimax and the optimal designs by Simon [1] are admissible against the distribution which gives probability mass 1 to n ($q = 0$) and to EN ($q = 1$), respectively. Compromise designs proposed by Jung *et al.* [2] are also admissible against a distribution Q and possess a practical appeal as their maximum sample size is close to that of the minimax design and their expected sample size under the null hypothesis is close to that of the optimal design.

We also showed how to find other admissible designs which are compromises between the minimax and the optimal designs for a suitable distribution on the convex hull formed by the minimax and the optimal designs. We illustrated a graphical search for admissible compromise designs with examples.

The application program has been developed and implemented in Java language and it is available from the first author upon request. Our method can be readily adapted for any other form of loss functions. More importantly, it can be easily generalized to any number of stages.

ACKNOWLEDGEMENTS

Dr Kim's research was supported by grant CA 52733 from the National Institutes of Health, Department of Health and Human Service, U.S. government.

REFERENCES

1. Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989; **10**:1–10.
2. Jung SH, Carey M, Kim KM. Graphical search for two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 2001; **22**:367–372.
3. Gehan EA. The determination of the number of patients required in a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases* 1961; **13**:346–353.
4. Lee YJ, Staquet M, Simon R, Cantane R, Muggia F. Two stage plans for patient accrual in phase II cancer clinical trials. *Cancer Treatment Reports* 1979; **63**:1721–1726.
5. Fleming TR. One sample multiple testing procedures for phase II clinical trials. *Biometrics* 1982; **38**:143–151.
6. Chang MN, Therneau TM, Wieand HS, Cha SS. Designs for group sequential phase II clinical trials. *Biometrics* 1987; **43**:865–874.

7. Spiegelhalter DJ, Freedman LS, Blackburn PR. Monitoring clinical trials: conditional or predictive power? *Controlled Clinical Trials* 1986; **7**:8–17.
8. Thall PF, Simon R. A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. *Controlled Clinical Trials* 1994; **15**:463–481.
9. Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics* 1979; **35**: 775–783.
10. DeGroot MH. *Optimal Statistical Decisions.* McGraw-Hill: New York, 1970.

