Biometrika (1988), **75**, 2, *pp.* 303–10
Printed in Great Britain

Two-stage selection and testing designs for comparative clinical trials

BY PETER F. THALL
Department of Statistics/Computer & Information Systems, George Washington University, Washington, D.C. 20892, U.S.A.

RICHARD SIMON AND SUSAN S. ELLENBERG
National Cancer Institute, Bethesda, Maryland 20892, U.S.A.

SUMMARY

A two-stage design which selects the best of several experimental treatments and compares it to a standard control is proposed. The design allows early termination with acceptance of the global null hypothesis. Optimal sample size and cut-off parameters are obtained by minimizing expected total sample size for fixed significance level and power.

Some key words: Clinical trial; Expected sample size; Optimal design; Selection; Two-stage design.

1. INTRODUCTION

In clinical research where there are several experimental treatments $E_1, \ldots, E_K$ of interest, often too few patients are available to evaluate each relative to a standard 'control' therapy C. A common approach in such a circumstance is to first select the experimental treatment E_ν which appears most promising based on uncontrolled pilot studies, and then compare E_ν to C in a large randomized clinical trial. When such pilot studies are performed at different institutions, treatment effects typically are confounded with many other factors. Moreover, the usual error rate computations associated with comparative testing do not account for the preliminary selection process. Consequently, the overall procedure may be neither effective nor efficient for identifying an experimental treatment which is an improvement over C.

In this paper we propose a new approach to the problem of identifying the best of K experimental treatments and determining whether it is superior to a control. We deal with the binomial setting where patient response may be characterized as either success or failure, with θ_k the success probability for E_k; $k=0$ corresponds to C. For ease of notation assume $\theta_1 \leqslant \ldots \leqslant \theta_K$. We propose a two-stage procedure which allows early termination with acceptance of H_0: $\theta_0 = \theta_1 = \ldots = \theta_K$, with design parameters chosen to minimize expected total sample size. The design and accompanying generalized definitions of size α and power $1-\beta$ are given in § 2. The algorithm used for optimization is described in § 3. Numerical results are presented in § 4, followed by a discussion of the relative merits of the proposed design.

2. THE DESIGN

Let $a(p) = \sin^{-1}\sqrt{p}$. At stage s, denote by n_s the number of patients in each treatment arm, X_{js} the number of successes in the jth arm and define

$$Z_{js} = (4n_s)^{\frac{1}{2}} a(X_{js}/n_s) \quad (j = 0, 1, \ldots, K;\ s = 1, 2).$$

Denote $n = n_1 + n_2$, $\pi = n_1/n$ and let y_s be the test cut-off at stage s. For prespecified values of the design parameters n_1, n_2, y_1, y_2 the trial is carried out as follows.

Stage 1. Randomize $(K+1)n_1$ patients equally to $C, E_1, \ldots, E_K$. If

$$T_1 = \frac{1}{\sqrt{2}} \max_{1 \leq j \leq K} (Z_{j1} - Z_{01}) > y_1$$

then select the treatment E_ν having the highest observed success rate and proceed to a second stage. If $T_1 \leq y_1$ then stop and accept H_0.

Stage 2. Randomize $2n_2$ additional patients equally to E_ν and C. If

$$T_2 = \frac{1}{\sqrt{2}} \{\pi^{\frac{1}{2}}(Z_{\nu 1} - Z_{01}) + (1-\pi)^{\frac{1}{2}}(Z_{\nu 2} - Z_{02})\} > y_2$$

then reject H_0 and conclude $\theta_\nu > \theta_0$; if $T_2 \leq y_2$ then accept H_0.

The first stage is a combined selection procedure and test, in that the empirically best experimental treatment is compared to the control. However, the design allows a new treatment to be judged superior to the control only after a second stage, based upon data from $2n$ patients. We shall formally assume that a tie at $\max_j (X_{j1})$ is broken by randomizing fairly at stage 1, although such a tie is rather unlikely. In practice the use of other objective criteria, such as toxicity, would provide a clinically reasonable alternative to randomization.

The stage 1 and 2 summary statistics are weighted in T_2 to minimize the asymptotic variance. Both T_1 and T_2 are defined on the standard normal scale to facilitate interpretation of y_1 and y_2.

We shall broaden the usual definitions of size and power to account for the hybrid nature of the proposed procedure. An experimental treatment E_ν is said to be 'chosen' if E_ν is selected at stage 1 and H_0 is rejected in favour of the alternative $\theta_\nu > \theta_0$ at stage 2. The 'size' of the procedure is the probability that any E_ν is chosen when H_0 holds. To develop a definition of power, let δ_1 and δ_2 be constants $(0 < \delta_1 < \delta_2 < 1 - \theta_0)$ such that, from a medical viewpoint, the success rate $\theta_0 + \delta_1$ is only a marginal improvement over θ_0 while $\theta_0 + \delta_2$ is a clinically significant improvement. Any E_j for which $\theta_j \geq \theta_0 + \delta_2$ is considered 'acceptable,' and we shall assume that (i) at least one E_j is acceptable, and (ii) no θ_j lies in the interval $(\theta_0 + \delta_1, \theta_0 + \delta_2)$. The second assumption is made because no statistical procedure can effectively discriminate between groups whose means are arbitrarily close. The 'power' function is now defined as $1 - \beta(\theta)$ equal to the probability of an acceptable choice given $\theta = (\theta_0, \theta_1, \ldots, \theta_K)$.

The following result will be utilized to compute power; the proof is given in the Appendix.

THEOREM 1. *Under assumptions* (i) *and* (ii), $1 - \beta(\theta)$ *is minimized for given* θ_0, δ_1 *and* δ_2 *when* $\theta_1 = \ldots = \theta_{K-1} = \theta_0 + \delta_1$ *and* $\theta_K = \theta_0 + \delta_2$.

The configuration θ^* minimizing $1-\beta(\theta)$ will be called least favourable, and hereafter power will be computed exclusively under θ^*, where only E_K is acceptable. Writing $\beta^*=\beta(\theta^*)$, it follows that $1-\beta^*$, the probability that E_K is chosen given θ^*, is

$$\sum_{x_0=0}^{n_1}\sum_{x_K=0}^{n_1}\text{pr}\{X_{01}=x_0, X_{K1}=x_K\mid\theta^*\}I\{a(x_K/n_1)-a(x_0/n_1)>y_1/\sqrt{(2n_1)}\}$$
$$\times\,\text{pr}\{M\mid X_{K1}=x_K;\,\theta^*\}\,\text{pr}\{T_2>y_2\mid X_{01}=x_0, X_{K1}=x_K;\,\theta^*\},\quad(1)$$

where I is the indicator function, and M denotes the event that $X_{K1}\geqslant\max_j(X_{j1})$ and E_K wins the randomization in case of a tie. Denote the binomial mass and cumulative distribution functions by $b(x;\,n,p)$ and $B(x;\,n,p)$, respectively. The first factor in (1) is simply $b(x_0;\,n_1,\theta_0)b(x_K;\,n_1,\theta_0+\delta_2)$, while

$$\text{pr}\{M\mid X_{K1}=x_K;\,\theta^*\}$$
$$=\begin{cases}K^{-1}\{b(0;\,n_1,\theta_0+\delta_1)\}^{K-1} & (x_K=0),\\ \sum_{j=0}^{K-1}\dfrac{1}{j+1}\binom{K-1}{j}\{b(x_K;\,n_1,\theta_0+\delta_1)\}^j\{B(x_K-1;\,n_1,\theta_0+\delta_1)\}^{K-1-j} & (x_K>0).\end{cases}$$

Denote $\Delta_s=a(\theta_0+\delta_s)-a(\theta_0)$ $(s=1,2)$. Since $(Z_{\nu 2}-Z_{02})/\sqrt{2}$ is asymptotically normal with mean $(2n_2)^{\frac{1}{2}}\Delta_2$ and variance 1, the last factor in (1) may be approximated by

$$1-\Phi\left\{\frac{y_2-(2/n)^{\frac{1}{2}}[n_1\{a(x_K/n_1)-a(x_0/n_1)\}+n_2\Delta_2]}{(1-\pi)^{\frac{1}{2}}}\right\},\quad(2)$$

where Φ is the standard normal distribution function. The size is computed by evaluating (1) at $\delta_1=\delta_2=0$ and multiplying the result by K.

It follows from the approximate normality of each Z_{js} that (1) is asymptotically equivalent to

$$\int_{-\infty}^{\infty}\phi(w_0)\int_{y_1^*}^{\infty}\phi(w_K)[\Phi\{w_K+2(\Delta_2-\Delta_1)\sqrt{n_1}\}]^{K-1}\{1-\Phi(y_2^*)\}\,dw_K\,dw_0,\quad(3)$$

where

$$y_1^*=y_1\sqrt{2}+w_0-2\Delta_2\sqrt{n_1},\quad y_2^*=\frac{y_2-(\pi/2)^{\frac{1}{2}}(w_K-w_0)-(2n_2)^{\frac{1}{2}}\Delta_2}{(1-\pi)^{\frac{1}{2}}},$$

and ϕ is the standard normal density. Thus asymptotically the size is independent of θ_0 and the power depends on θ only through Δ_1 and Δ_2.

The formulation (1) is studied, rather than an approximation of (3) via numerical integration, because the former is more accurate. The only numerical approximation used is that required for evaluation of (2), and this is done to avoid a fourfold sum. In addition, the optimal stage 2 sample size n_2 is sufficiently large, for all designs considered, to ensure high accuracy of the normal approximations for $(Z_{K2}-Z_{02})/\sqrt{2}$ in evaluating (2).

The approach here is to choose the design parameters n_1, n_2, y_1, y_2 to minimize expected total sample size $E(N)=(K+1)n_1+2n_2\,\text{pr}(T_1>y_1)$ subject to the constraints imposed by specifying α and $1-\beta^*$. Since $\text{pr}(T_1>y_1)$ depends on θ, minimization may be carried out under H_0, θ^* or with $E(N)$ averaged over a specified prior. See, for example, Hald (1975), Colton & McPherson (1976), Jennison (1987), or Case, Morgan & Davis (1987). In the present context, the goal is to obtain reasonable designs which perform well under both H_0 and θ^*. We thus derive designs to minimize the simple average $\frac{1}{2}E(N\mid H_0)+\frac{1}{2}E(N\mid\theta^*)$.

3. DERIVATION OF OPTIMAL DESIGNS

All designs given here correspond to $\alpha = 0{\cdot}05$, $\delta_1 = 0{\cdot}05$, $\delta_2 = 0{\cdot}20$, with $K = 2, 3, 4$; $\theta_0 = 0{\cdot}2, 0{\cdot}4, 0{\cdot}6$, and $1-\beta^* = 0{\cdot}70, 0{\cdot}75, 0{\cdot}80$. The range of values for $1-\beta^*$ is somewhat lower than conventional power figures since it accounts for both the selection and test in choosing an acceptable treatment.

The search algorithm proceeded with n_1 fixed in an outer loop, then y_1 fixed at a value in a discrete domain having grid width 0·025 to locate local minima. For each n_1 and y_1, the two equations in size and power were solved for y_2 and π to an accuracy of $\pm 10^{-4}$ for both α and $1-\beta^*$, using a least-squares algorithm due to Shrager (1970). Then $E(N)$ was computed based upon the smallest integer $n_2 \geq n_1(\pi^{-1}-1)$. Regarded as a function of y_1 for n_1 fixed, $E(N)$ has two distinct local minima E_1 and E_2, say, in all cases. An exhaustive search for y_1 along the T_1 domain

$$(2n_1)^{\frac{1}{2}}\{a(x_K/n_1) - a(x_0/n_1)\} \quad (0 \leq x_0, x_K \leq n_1)$$

in the neighbourhoods of E_1 and E_2 was carried out to obtain the minimum for given n_1. As a function of n_1 this minimum is unimodal, thus yielding the global optimum.

For n_1 outside an integer domain of width up to about 20 in the cases considered, both E_1 and E_2 rapidly increase with y_1, until no solution for π and y_2 exists. The value of y_1 yielding $\min(E_1, E_2)$ is quite distinct for $\theta_0 = 0{\cdot}2$ or 0·6, but often is unique only up to one or two decimal places for $\theta_0 = 0{\cdot}4$. For example, $0{\cdot}485 \leq y_1 \leq 0{\cdot}606$ minimizes $E(N)$ for the design with $K = 4$, $\theta_0 = 0{\cdot}4$ and $1-\beta^* = 0{\cdot}75$ in Table 1. In such cases, a value of y_1 near the centre of the optimizing interval is given. The values α and $1-\beta^*$ are rather more sensitive to changes in y_2, so that the tabled cut-off y_2 is fully accurate to all digits reported, given the method of computation.

4. NUMERICAL RESULTS

Optimum design parameters are presented in Table 1, along with other numerical values which describe each design's behaviour.

The values of n_1 are close to the sample sizes conventionally used for pilot or 'phase II' studies in clinical cancer research. This is desirable since it would be difficult to persuade clinical investigators to use a design having a substantially larger selection stage. The values of n_1 and n_2 in Table 1 all are based upon the prespecified increments $\delta_1 = 0{\cdot}05$ and $\delta_2 = 0{\cdot}20$, the latter chosen to correspond to the target improvement in response rate used in many clinical trials of advanced cancer. The optimum n_1 increases with K for fixed θ_0 and $1-\beta^*$, essentially because E_K must overcome more competing treatments in order to be selected at stage 1. This was also found to be the case by Simon, Wittes & Ellenberg (1985) for randomized phase II trials without a control group. For K and $1-\beta^*$ fixed, the values of n_1 and n_2 are larger for $\theta_0 = 0{\cdot}40$ than for $\theta_0 = 0{\cdot}20$ or 0·60, probably due to the fact that Δ_2, as computed on the $\sin^{-1}$ square root scale, is somewhat smaller at $\theta_0 = 0{\cdot}40$.

The inclusion of a control arm at stage 1 allows use of the pooled statistic T_2 for comparison of E_ν to C at stage 2. Thus, the sample size $2n$ available for that comparison is only moderately larger than that required for a single-stage two-sample test. The stage 2 cut-off values of y_2 are more extreme than the conventional one-sided test cut-off 1·645. This may be regarded as an adjustment for the stage 1 selection process, and it is consistent with the fact that y_2 increases with K. The increase in y_2 over 1·645 is rather moderate in all cases, with conditional stage 2 test sizes ranging from 0·022 to 0·037.

Table 1. *Designs having minimal* $E(N)$ *for given* K, θ_0, $1-\beta^*$, $\alpha=0{\cdot}05$, $\delta_1=0{\cdot}05$ *and* $\delta_2=0{\cdot}20$

K	θ_0	$1-\beta^*$	n_1	n_2	y_1	y_2	$E(N)$	$N_{\max}$	τ_0	γ^*
2	0·2	0·70	30	44	0·787	1·787	141·71	178	0·667	0·030
		0·75	36	44	0·730	1·818	163·71	196	0·640	0·026
		0·80	40	52	0·689	1·812	187·64	224	0·626	0·025
2	0·4	0·70	36	50	0·684	1·811	172·99	208	0·588	0·027
		0·75	40	58	0·590	1·808	197·36	236	0·575	0·026
		0·80	47	62	0·580	1·822	226·53	265	0·557	0·023
2	0·6	0·70	27	45	0·578	1·794	139·62	181	0·589	0·028
		0·75	31	50	0·543	1·798	159·54	193	0·584	0·026
		0·80	36	55	0·500	1·803	183·68	218	0·563	0·023
3	0·2	0·70	33	55	0·762	1·902	205·09	242	0·571	0·046
		0·75	38	59	0·709	1·916	233·33	270	0·548	0·042
		0·80	48	57	0·835	1·926	266·97	306	0·619	0·035
3	0·4	0·70	39	64	0·591	1·917	247·09	284	0·492	0·042
		0·75	47	63	0·550	1·944	280·89	314	0·469	0·036
		0·80	52	75	0·500	1·936	320·37	358	0·457	0·034
3	0·6	0·70	32	51	0·530	1·928	201·04	230	0·496	0·039
		0·75	37	55	0·509	1·935	227·94	258	0·490	0·036
		0·80	42	62	0·472	1·938	260·28	292	0·471	0·032
4	0·2	0·70	36	61	0·721	1·984	267·26	302	0·496	0·055
		0·75	44	62	0·868	1·983	303·14	344	0·583	0·049
		0·80	51	65	0·800	2·004	345·64	385	0·555	0·043
4	0·4	0·70	45	69	0·675	1·987	321·58	363	0·512	0·049
		0·75	49	77	0·550	2·004	364·32	399	0·404	0·046
		0·80	58	84	0·700	2·001	414·47	458	0·471	0·041
4	0·6	0·70	35	58	0·529	2·002	262·05	291	0·440	0·047
		0·75	42	61	0·717	2·000	296·28	332	0·520	0·042
		0·80	48	66	0·655	2·014	337·10	372	0·486	0·037

$E(N)=\frac{1}{2}\{E(N\mid H_0)+E(N\mid\theta^*)\}$; $N_{\max}=(K+1)n_1+2n_2$; $\tau_0=\text{pr}\,(T_1\leqslant y_1\mid H_0)$
$\gamma^*=\text{pr}$ (choose suboptimal $E_\nu\mid\theta^*$)

The discreteness of the design parameter space in general and the shallowness of $E(N)$ as a function of y_1 for $\theta_0=0{\cdot}40$ tend to obscure consistent patterns in the design with regard to power and the number K of experimental agents. However, the optimal designs all terminate early with probability about $\frac{1}{2}$ when H_0 is true. This is a highly desirable property, since acceptance at stage 1 obviates randomization of an additional $2n_2$ patients for a second stage. The values of $\tau_0=\text{pr}\,(T_1\leqslant y_1\mid H_0)$ are given in Table 1.

The error of choosing one of the $K-1$ suboptimal treatments under θ^* is highly undesirable. While this decision is not a type II error, since H_0 is not accepted, it does imply a failure of the trial to correctly detect an important treatment advance. The probability γ^* of this third type of error is small for all designs given here, as shown in the last column of Table 1.

Table 2 presents comparisons between the optimal two-stage designs presented here and the balanced case of single-stage designs of Dunnett (1984), for the case $\theta_0=0{\cdot}20$. Although Dunnett's designs were developed for a slightly different formulation of the problem, the least favourable configuration of his formulation, with $\delta_0^*=0$, is identical to the least favourable configuration of our formulation. Dunnett's designs provide more

Table 2. *Comparison of sample sizes for Dunnett balanced one-stage and proposed optimal two-stage designs at* $\alpha = 0{\cdot}05$, $\theta_0 = 0{\cdot}20$, $\delta_1 = 0{\cdot}05$, $\delta_2 = 0{\cdot}20$

		Dunnett	Optimal two-stage			
K	$1-\beta^*$	N	$(K+1)n_1$	$E(N \mid H_0)$	$E(N \mid \theta^*)$	$N_{\max}$
2	0·70	189	90	119·3	166·1	178
	0·75	213	108	139·7	187·7	196
	0·80	240	120	158·9	216·4	224
3	0·70	284	132	179·2	231·0	242
	0·75	316	152	205·3	261·4	270
	0·80	356	192	235·4	298·5	306
4	0·70	385	180	241·5	293·0	302
	0·75	425	220	271·7	334·6	344
	0·80	475	255	312·9	378·4	385

information about those experimental treatments not chosen as best essentially because none is discarded early. This is an advantage of single-stage designs. For identifying one 'acceptable' or best treatment, however, the balanced single-stage designs are quite inferior to the two-stage designs given here. Note that even the maximum possible sample sizes of the corresponding two-stage designs are smaller than the fixed sample sizes of Dunnett's balanced designs in each case. Analogous comparisons for $\theta_0 = 0{\cdot}40$ and $0{\cdot}60$ yielded similar results. Although unbalanced single-stage designs are not strictly comparable to the two-stage designs presented here, the total sample size of the single-stage designs could be reduced by optimizing the ratio of the number of patients placed on the control treatment to the number on each of the experimental treatments. Dunnett provides examples where savings of up to 10% can be achieved using a ratio of about 2 for parameter values of interest here.

5. DISCUSSION

The numerical results for the proposed two-stage designs indicate desirable operating characteristics under both H_0 and θ^*. The alternative considered here specifies an interval $(\theta_0+\delta_1,\ \theta_0+\delta_2)$ between the success rates of one acceptable and $K-1$ unacceptable experimental treatments. The designs given in Table 1 are thus most appropriate for situations where at most one experimental treatment is likely to be a substantial improvement over C. While the probability of an acceptable choice would decrease with $\delta_2-\delta_1$, the loss due to choosing a suboptimal treatment would also decrease. Naturally, the optimal designs given here will perform well under admissible configurations more favourable than θ^*, that is when any suboptimal treatment has rate $\theta_k < \theta_0+\delta_1$ or when one or more $\theta_k > \theta_0+\delta_2$.

Perhaps the most important practical feature of this design is the use of a randomized first stage to select an experimental treatment for further study. Although small pilot studies still might be necessary to determine tolerable dosages, the randomized selection approach avoids institutional effects which confound and obscure treatment differences.

The designs proposed here obviously are not applicable to all clinical situations. First, we have assumed the outcome is binary and quickly observed. The design could easily be generalized to accommodate survival data, but its efficiency would be limited severely

by long survival times. Secondly, for some diseases sufficiently many patients are available to allow comparison of each experimental treatment to the control. The design proposed here does not provide such information. Designs such as those proposed by DeMets & Ware (1982), Lan, Simon & Halperin (1982) or Ellenberg & Eisenberger (1985) permit early termination when interim results are not favourable for a single experimental treatment. These could be generalized to accommodate several experimental treatments, dropping treatment arms as the trial progresses.

Sylvester & Staquet (1980) used a decision-theoretic framework for determining the sample size of a phase II clinical trial for a single experimental treatment. The loss function for each patient is proportional to the difference between the response probability for the treatment assigned and that of a standard control. Losses are considered both for patients in the phase II screening study and for those in a possible subsequent phase III study, and over the time horizon until a superior treatment is found. The sample size of the phase III study, time horizon, and prior distribution for the response probability to the experimental treatment all are assumed known.

Although there is a large literature on ranking and selection (Gibbons, Olkin & Sobel, 1977; Gupta & Panchapakesan, 1979; Santner & Tamhane, 1984), such methods have seen little use in clinical trials. Most clinical trials require a quantification of evidence concerning the relative merits of two or more treatments, rather than only a selection of one. Dunnett's (1984) procedure formally incorporates selection into the decision making process, but without the possibility of early termination.

Whitehead (1986) has studied a problem similar to that considered here, but where the number of experimental treatments is large relative to the number of available patients. His approach requires a prior distribution for $\theta_1, \ldots, \theta_K$ and does not permit early termination. Whitehead's approach and ours may be regarded as complementary in that they are directed toward somewhat different clinical situations. Both are attempts to improve the efficiency of a program of treatment development when more than one new treatment is available.

Appendix

Proof of Theorem 1

Let $A = \{K-m+1, \ldots, K\}$ denote the indices of acceptable treatments, where $|A| = m \geqslant 1$ by assumption. Thus θ satisfies

$$\theta_1 \leqslant \ldots \leqslant \theta_{K-m} \leqslant \theta_0 + \delta_1, \quad \theta_0 + \delta_2 \leqslant \theta_{K-m+1} \leqslant \ldots \leqslant \theta_K. \tag{A1}$$

For fixed $\zeta \in (0, 1)$ let $U_1, \ldots, U_K$ be independent random variables, each uniformly distributed on $[0, \zeta]$, and denote by $Z_{j1\zeta}$ the stage 1 statistic based upon $X_{j1} + U_j$ rather than X_{j1} $(1 \leqslant j \leqslant K)$. Since $a(p) = \sin^{-1}\sqrt{p}$ is strictly monotone and $\zeta < 1$, no strict inequalities among $\{Z_{j1}, 1 \leqslant j \leqslant K\}$ are altered by replacing each Z_{j1} with $Z_{j1\zeta}$, but this acts as a fair randomization device in the event of a tie. Thus the probability of an acceptable choice is

$$\text{pr}\{\max_{j \in A}(Z_{j1\zeta}) > \max_{r \in A^c}(Z_{r1\zeta}), Z_{\nu 1} = \max_{j \in A}(Z_{j1}) > Z_{01} + y_1\sqrt{2}, T_2 > y_2\}. \tag{A2}$$

Since $Z_{j1\zeta}$ and Z_{js} are stochastically increasing in θ_j for each j and s, it follows that (A2) is minimized subject to (A1) when $\theta_j = \theta_0 + \delta_1$, for $1 \leqslant j \leqslant K-m$, and $\theta_r = \theta_0 + \delta_2$, for $K-m+1 \leqslant r \leqslant K$. If $m > 1$, then (A2) is strictly decreased by switching one element from A to A^c, hence the minimum is achieved when $m = 1$.

REFERENCES

CASE, L. D., MORGAN, T. M. & DAVIS, C. E. (1987). Optimal restricted two-staged designs. *Controlled Clin. Trials* **8**, 146–56.

COLTON, T. & MCPHERSON, K. (1976). Two-stage plans compared with fixed-sample-size and Wald SPRT plans. *J. Am. Statist. Assoc.* **71**, 80–6.

DEMETS, D. L. & WARE, J. H. (1982). Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* **69**, 661–3.

DUNNETT, C. W. (1984). Selection of the best treatment in comparison to a control with application to a medical trial. In *Design of Experiments: Ranking and Selection*, Ed. T. J. Santner and A. C. Tamhane, pp. 47–66. New York: Marcel Dekker.

ELLENBERG, S. S. & EISENBERGER, M. A. (1985). An efficient design for phase III studies of combination chemotherapies (with discussion). *Cancer Treat. Rep.* **69**, 1147–54.

GIBBONS, J. D., OLKIN, I. & SOBEL, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* New York: Wiley.

GUPTA, S. S. & PANCHAPAKESAN, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* New York: Wiley.

HALD, A. (1975). Optimum double sampling tests of given strength. I. The normal distribution. *J. Am. Statist. Assoc.* **70**, 451–6.

JENNISON, C. (1987). Efficient group sequential tests with unpredictable group sizes. *Biometrika* **74**, 155–65.

LAN, K. K. G., SIMON, R. & HALPERIN, M. (1982). Stochastically curtailed tests in long term clinical trials. *Comm. Statist.-Sequen. Anal.* **1**, 207–14.

SANTNER, T. J. & TAMHANE, A. C. (1984). *Design of Experiments: Ranking and Selection.* New York: Marcel Dekker.

SHRAGER, R. I. (1970). Nonlinear regression with linear constraints: an extension of the magnified diagonal method. *J. Assoc. Comp. Mach.* **17**, 446–52.

SIMON, R., WITTES, R. E. & ELLENBERG, S. S. (1985). Randomized phase II clinical trials. *Cancer Treat. Rep.* **69**, 1375–81.

SYLVESTER, R. J. & STAQUET, M. M. (1980). Design of phase II clinical trials in cancer using decision theory. *Cancer Treat. Rep.* **64**, 519–24.

WHITEHEAD, J. (1986). Sample sizes for phase II and phase III clinical trials: An integrated approach. *Statist. Med.* **5**, 459–64.

[*Received August* 1987. *Revised October* 1987]