Optimal Two-Stage Designs for Phase II Clinical Trials

Richard Simon, PhD

Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

ABSTRACT: The primary objective of a phase II clinical trial of a new drug or regimen is to determine whether it has sufficient biological activity against the disease under study to warrant more extensive development. Such trials are often conducted in a multi-institution setting where designs of more than two stages are difficult to manage. This paper presents two-stage designs that are optimal in the sense that the expected sample size is minimized if the regimen has low activity subject to constraints upon the size of the type 1 and type 2 errors. Two-stage designs which minimize the maximum sample size are also determined. Optimum and "minimax" designs for a range of design parameters are tabulated. These designs can also be used for pilot studies of new regimens where toxicity is the endpoint of interest.

KEY WORDS: *clinical trials, phase II trials, optimization*

INTRODUCTION

A phase II study of a cancer treatment is an uncontrolled trial for obtaining an initial estimate of the degree of antitumor effect of the treatment. Phase I trials provide information about the maximum tolerated dose(s) of the treatment, which is important because most cancer treatments must be delivered at maximum dose for maximum effect. Phase I trials generally treat only three to six patients per dose level, however, and the patients are diverse with regard to their cancer diagnosis [1]. Consequently such trials provide little or no information about antitumor activity. The proportion of patients whose tumors shrink by at least 50% is the primary endpoint of most phase II trials although the durability of such responses is also of interest. Such trials are not controlled and do not determine the "effectiveness" of the treatment or the role of the drug in the treatment of the disease. The purpose of a phase II trial of a new anticancer drug is to determine whether the drug has sufficient activity against a specified type of tumor to warrant its further development. Further development may mean combining the drug with other drugs, evaluation in patients with less advanced disease, or initiation of phase III studies in which survival results are compared to those for a standard treatment. Phase II trials of combination regimens are also conducted to determine whether

Address reprint requests to: Richard Simon, PhD, Biometric Research Branch, National Cancer Institute, EPN 739, Bethesda, MD 20892

Received January 6, 1988; revised July 28, 1988.

655 Avenue of the Americas, New York, New York 10010 0197-2456/1989/$3.50

the treatment is sufficiently promising to warrant a major controlled clinical evaluation against the standard therapy.

The designs developed here are based on testing a null hypothesis $H_0{:}p \leq p_0$ that the true response probability is less than some uninteresting level p_0. If the null hypothesis is true, then we require that the probability should be less than α of concluding that the drug is sufficiently promising that it should be accepted for further study in other clinical trials. We also require that if a specified alternative hypothesis $H_1{:}p \geq p_1$ that the true response probability is at least some desirable target level p_1 is true, then the probability of rejecting the drug for further study should be less than β. In addition to these constraints, we wish to minimize the number of patients treated with a drug of low activity. We shall restrict our attention to two-stage designs because of practical considerations in the management of multi-institution clinical trials. The main practical consideration is that evaluation of a patient's response is not instantaneous and may require observation for weeks or months. Consequently patient accrual at the end of a stage may have to be suspended until it is determined whether the criteria for continuing are satisfied. Such suspension of accrual is awkward for physicians who are entering patients on the study. More than one such disruption during the trial would, in many cases, not be acceptable. Although more stages are desirable from the standpoint of efficiency, two-stage designs often achieve a substantial portion of the savings of fully sequential designs [2–4].

OPTIMAL TWO-STAGE DESIGNS

If the numbers of patients studied in the first and second stage are denoted by n_1 and n_2 respectively, then the expected sample size is EN = n_1 + (1 − PET)n_2, where PET represents the probability of early termination after the first stage. The decision of whether or not to terminate after the first stage will be based on the number of responses observed for those n_1 patients. The expected sample size EN and the probability of early termination depend on the true probability of response p. We will terminate the experiment at the end of the first stage and reject the drug if r_1 or fewer responses are observed. This occurs with probability PET = $B(r_1;p,n_1)$, where B denotes the cumulative binomial distribution. We will reject the drug at the end of the second stage if r or fewer responses are observed. Hence the probability of rejecting a drug with success probability p is

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

where b denotes the binomial probability mass function.

The design approach considered here is to specify the parameters p_0,p_1,α, and β and then determine the two-stage design that satisfies the error probability constraints and minimizes the expected sample size when the response probability is p_0. The optimization is taken over all values of n_1 and n_2 as well as r_1 and r. Early acceptance of the drug is not permitted here. The ethical imperative for early termination occurs when the drug has low activity. When

the drug has substantial activity ($\geq p_1$) there is often interest in studying additional patients in order to estimate the proportion, extent, and durability of response. We have optimized the designs to minimize expected sample size when the response probability is p_0. This is because of ethical considerations. It would be possible to reduce expected sample sizes for the designs considered here by termination early if the number of responses in the first stage exceeds the final decision criterion r. This would have a negligible effect on performance under H_0. If early rejection of H_0 is really of interest, however, a less conservative early rejection rule should be used.

For specified values of p_0,p_1,α, and β we have determined optimal designs by enumeration using exact binomial probabilities. For each value of total sample size n and each value of n_1 in the range $(1,n-1)$ we determined the integer values of r_1 and r, which satisfied the two constraints and minimized the expected sample size when $p = p_0$. This was found by searching over the range $r_1 \in (0,n_1)$. For each value of r_1 we determined the maximum value of r that satisfied the type 2 error constraint. We then examine whether that set of parameters (n,n_1,r_1,r) satisfied the type 1 error constraint. If it did, then we compared the expected sample size to the minimum achieved by previous feasible designs and continued the search over r_1. Keeping n fixed we searched over the range of n_1 to find the optimal two-stage design for that maximum sample size n. The search over n ranged from a lower value of about

$$\bar{p}(1-\bar{p})\left[\frac{z_{1-\alpha}+z_{1-\beta}}{p_1-p_0}\right]^2,$$

where $\bar{p} = (p_0 + p_1)/2$. We checked below this starting point to ensure that we had determined the smallest maximum sample size n for which there was a nontrivial $(n_1,n_2 > 0)$ two-stage design that satisfied the error probability constraints. The enumeration procedure searched upwards from this minimum value of n until it was clear that the optimum had been determined. The minimum expected sample size for fixed n is not a unimodal function of n because of discreteness of the underlying binomial distributions. Nevertheless, eventually as n increased the value of the local minima increased and it was clear that a global minimum had been found. Calculations were carried out in APL on a Microvax II computer. The computer program is available on request.

Tables 1 and 2 show optimal designs for a variety of design parameters. Table 1 applies to trials with $p_1 - p_0 = 0.20$ and Table 2 is for trials with $p_1 - p_0 = 0.15$. The optimal designs are shown on the left half of the tables. For each (p_0,p_1), the three rows correspond to optimal designs for $(\alpha,\beta) = (0.10,0.10)$, $(0.05,0.20)$, and $(0.05,0.10)$, respectively. The tabulated results include the optimal size of the first stage (n_1), the maximum sample size (n), the upper limits on observed response rate that result in rejection of the drug at the end of the first stage (r_1/n_1) and at the end of the trial (r/n), the expected sample size [EN(p_0)], and probability of terminating the trial at the end of the first stage [PET(p_0)] if the response probability is p_0. For example, the first line in Table 1 corresponds to a design with $p_0 = 0.05$ and $p_1 = 0.25$. The first stage consists of nine patients. If no responses are seen then the trial is terminated. Otherwise accrual continues to a total of 24 patients. The average

Table 1 Designs for $p_1 - p_0 = 0.20^a$

		Optimal Design				Minimax Design			
		Reject Drug if Response Rate				Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.25	0/9	2/24	14.5	0.63	0/13	2/20	16.4	0.51
		0/9	2/17	12.0	0.63	0/12	2/16	13.8	0.54
		0/9	3/30	16.8	0.63	0/15	3/25	20.4	0.46
0.10	0.30	1/12	5/35	19.8	0.65	1/16	4/25	20.4	0.51
		1/10	5/29	15.0	0.74	1/15	5/25	19.5	0.55
		2/18	6/35	22.5	0.71	2/22	6/33	26.2	0.62
0.20	0.40	3/17	10/37	26.0	0.55	3/19	10/36	28.3	0.46
		3/13	12/43	20.6	0.75	4/18	10/33	22.3	0.50
		4/19	15/54	30.4	0.67	5/24	13/45	31.2	0.66
0.30	0.50	7/22	17/46	29.9	0.67	7/28	15/39	35.0	0.36
		5/15	18/46	23.6	0.72	6/19	16/39	25.7	0.48
		8/24	24/63	34.7	0.73	7/24	21/53	36.6	0.56
0.40	0.60	7/18	22/46	30.2	0.56	11/28	20/41	33.8	0.55
		7/16	23/46	24.5	0.72	17/34	20/39	34.4	0.91
		11/25	32/66	36.0	0.73	12/29	27/54	38.1	0.64
0.50	0.70	11/21	26/45	29.0	0.67	11/23	23/39	31.0	0.50
		8/15	26/43	23.5	0.70	12/23	23/37	27.7	0.66
		13/24	36/61	34.0	0.73	14/27	32/53	36.1	0.65
0.60	0.80	6/11	26/38	25.4	0.47	18/27	24/35	28.5	0.82
		7/11	30/43	20.5	0.70	8/13	25/35	20.8	0.65
		12/19	37/53	29.5	0.69	15/26	32/45	35.9	0.48
0.70	0.90	6/9	22/28	17.8	0.54	11/16	20/25	20.1	0.55
		4/6	22/27	14.8	0.58	19/23	21/26	23.2	0.95
		11/15	29/36	21.2	0.70	13/18	26/32	22.7	0.67

[a]For each value of (p_0, p_1), designs are given for three sets of error probabilities (α, β). The first, second and third rows correspond to error probability limits (0.10, 0.10), (0.05, 0.20), and (0.05, 0.10) respectively. For each design, EN(p_0) and PET(p_0) denote the expected sample size and the probability of early termination when the true response probability is p_0.

sample size is 14.5 and the probability of early termination is 0.63 for a drug with a response probability of 0.05. All calculations are based on exact binomial probabilities.

As pointed out above, the optimal two-stage design does not necessarily minimize the maximum sample size n subject to the error probability constraints. For example, consider the case of $(p_0,p_1) = (0.30,0.50)$ and $(\alpha,\beta) = (0.10,0.10)$. The optimal design, as seen in Table 1, has a maximum sample size of 46 patients. There is a two-stage design based on a maximum of 39 patients that also satisfies the error constraints. That design has $n_1 = 28$, $r_1 = 7$, $r = 15$. The expected sample size for that design is 35.0, a 17% increase over the expected size of 29.9 for the design with the minimum expected sample size. For each set of design parameters in Tables 1 and 2, the left side shows the optimal design and the right side shows the two-stage design

Table 2 Designs for $p_1 - p_0 = 0.15$[a]

		Optimal Design				Minimax Design			
		Reject Drug if Response Rate				Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.20	0/12	3/37	23.5	0.54	0/18	3/32	26.4	0.40
		0/10	3/29	17.6	0.60	0/13	3/27	19.8	0.51
		1/21	4/41	26.7	0.72	1/29	4/38	32.9	0.57
0.10	0.25	2/21	7/50	31.2	0.65	2/27	6/40	33.7	0.48
		2/18	7/43	24.7	0.73	2/22	7/40	28.8	0.62
		2/21	10/66	36.8	0.65	3/31	9/55	40.0	0.62
0.20	0.35	5/27	16/63	43.6	0.54	6/33	15/58	45.5	0.50
		5/22	19/72	35.4	0.73	6/31	15/53	40.4	0.57
		8/37	22/83	51.4	0.69	8/42	21/77	58.4	0.53
0.30	0.45	9/30	29/82	51.4	0.59	16/50	25/69	56.0	0.68
		9/27	30/81	41.7	0.73	16/46	25/65	49.6	0.81
		13/40	40/110	60.8	0.70	27/77	33/88	78.5	0.86
0.40	0.55	16/38	40/88	54.5	0.67	18/45	34/73	57.2	0.56
		11/26	40/84	44.9	0.67	28/59	34/70	60.1	0.90
		19/45	49/104	64.0	0.68	24/62	45/94	78.9	0.47
0.50	0.65	18/35	47/84	53.0	0.63	19/40	41/72	58.0	0.44
		15/28	48/83	43.7	0.71	39/66	40/68	66.1	0.95
		22/42	60/105	62.3	0.68	28/57	54/93	75.0	0.50
0.60	0.75	21/34	47/71	47.1	0.65	25/43	43/64	54.4	0.46
		17/27	46/67	39.4	0.69	18/30	43/62	43.8	0.57
		21/34	64/95	55.6	0.65	48/72	57/84	73.2	0.90
0.70	0.85	14/20	45/59	36.2	0.58	15/22	40/52	36.8	0.51
		14/19	46/59	30.3	0.72	16/23	39/49	34.4	0.56
		18/25	61/79	43.4	0.66	33/44	53/68	48.5	0.81
0.80	0.95	5/7	27/31	20.8	0.42	5/7	27/31	20.8	0.42
		7/9	26/29	17.7	0.56	7/9	26/29	17.7	0.56
		16/19	37/42	24.4	0.76	31/35	35/40	35.3	0.94

[a]For each value of (p_0, p_1), designs are given for three sets of error probabilities (α, β). The first, second, and third rows correspond to error probability limits (0.10, 0.10), (0.05, 0.20), and (0.05, 0.10) respectively. For each design, EN(p_0) and PET(p_0) denote the expected sample size and the probability of early termination when the true response probability is p_0.

having the smallest maximum sample size n that satisfies the design constraints.

In some cases, the "minimax" design may be more attractive than that with the minimum expected sample size. This will be the case when the difference in expected sample sizes is small and the patient accrual rate is low. Consider, for example, the case of distinguishing $p_0 = 0.10$ from $p_1 = 0.30$ with $\alpha = \beta = 0.10$. The optimal design in Table 1 has an expected sample size under H_0 of 19.8 and a maximum sample size of 35. The minimax two-stage design has an expected sample size of 20.4 and a maximum sample size of 25. If the accrual rate is only ten patients per year, it could take 1 year longer to complete

the optimal design than the minimax design. This may be more important than the slight reduction in expected sample size. Also, the optimal designs achieve reductions in EN(p_0) by having smaller first stages than the minimax designs. The small first stage exposes few patients to an inactive treatment. In cases where the patient population is very heterogeneous, however, a very small first stage may not be desirable because patients entered early in the study may be unrepresentative of the eligible population. Hence there are circumstances where the minimax designs are preferable.

Usually the minimax two-stage design has the same maximum sample size n as the smallest single-stage design that satisfies the error probabilities. The minimax two-stage design has a smaller expected sample size under H_0, however. In determining the minimax designs, we have limited attention to non-trivial two-stage designs; those with n_1 and $n_2 > 0$. In some cases, the minimax two-stage design has a smaller maximum sample size than the optimal single stage design. For example, the optimal single-stage design for distinguishing $p_0 = 0.60$ from $p_1 = 0.80$ with $\alpha = \beta = 0.10$ has a sample size of 36 and rejects H_1 if 25 or fewer responses are observed. As shown in Table 1, the minimax two-stage design for this case has a maximum sample size of 35. This is due to the discreteness of the binomial distribution and the fact that although the one- and two-stage designs both satisfy the error constraints, they do not have the same error probabilities.

DISCUSSION

A number of statistical designs have been previously proposed for phase II clinical trials. The first, and most commonly used design was developed by Gehan [5]. It is a two-stage design for estimating the response rate. It is most commonly employed with a first stage of 14 patients. If no responses are observed in the first stage, then the trial is terminated because this event has probability less than 0.05 if the true response probability is greater than or equal to 0.20. If at least one response is observed in the first 14 patients, then a second stage of accrual is carried out in order to obtain an estimate of the response probability having a specified standard error. The number of patients accrued in the second stage depends on the number of responses observed in the first stage and the desired standard error. Gehan's design is often used with a second stage of 11 patients. This provides for estimation with approximately a 10% standard error, although this may provide very broad confidence limits. Requiring that the standard error be 5% instead of 10% provides estimates with more satisfactory precision, but requires much larger sample sizes. If the second stage is to be much larger, then it is not clear that the first stage should include only 14 patients. This is because even for a poor drug with a true response probability of 5%, there is a 51% chance of obtaining at least one response in the first 14 patients.

Although one can make a cogent case that the main objective of phase III clinical trials should be estimation rather than hypothesis testing, in planning phase II trials it often seems more useful to identify levels of activity to distinguish than to select a precision for estimation. The two design principles are mathematically similar, but the hypothesis testing formulation encourages

phase II trial designers to think carefully about the objectives of the experiment and to define how decisions will be influenced by results. It is for this reason that we have adopted the hypothesis testing framework employed by most authors. Jennison and Turnbull [6] and Chang and O'Brien [7] have described methods for calculating confidence intervals following sequential sampling procedures of the type proposed here.

Schultz et al. [8] developed recursive formulae for calculating the operating characteristics of general k-stage designs with the possibility of acceptance and rejection at each stage. Early acceptance is appropriate for situations where patients are very limited or the drug is very expensive. Fleming [9] also studied k-stage designs with acceptance or rejection possible at each stage. Fleming's design is based on an approach developed for phase III trials [10] in which early rejection of a hypothesis occurs only when interim results are quite extreme. This conservatism permits final analysis to be unaffected by interim monitoring if early termination does not occur but is not always desirable for phase II trials of agents that are likely to be inactive. Lee et al. [11] considered two-stage designs that permit the possibility of recommending additional phase II trials. Herson [12] described a multistage Bayesian approach in which the trial is terminated early if the predictive probability of rejecting the null hypothesis at the maximum sample size falls below a specified level. The predictive probability is calculated with regard to the posterior distribution of p given the prior distribution and the data. None of the above authors attempted to optimize their designs.

Sylvester and Staquet [13] have provided an interesting decision theory approach to this problem, although the complexities of real-world decision-making are difficult to capture with simple models. Colton and McPherson [2] considered the design of two-stage clinical trials with binary outcome. They restricted attention to the case where the sample sizes of the two stages are equal and where the null hypothesis is $p = 0.05$. They determined a total sample size and rejection regions r_1 and r to minimize the expected sample size under the alternative hypothesis.

Chang et al. [14] have recently also considered the problem of optimizing the design of phase II trials. They described an algorithm that, although not guaranteed to find the optimum design, seemed to work well. For their designs early acceptance of the drug is permitted and the expected sample size, averaged over the null and alternative hypotheses, is minimized. In their published tables they have not optimized with regard to the maximum sample size.

Comparison of the designs developed here to those published by others for distinguishing a null hypothesis $p \leqslant p_0$ from an alternative $p \geqslant p_1$ is made difficult by the fact that two designs may not be equivalent with regard to the error probabilities α and β. Table 3 compares the optimum designs developed here with two-stage designs tabulated by Fleming [9] and by Chang et al. [14] for cases where the error probabilities are not too dissimilar. For most cases shown in Table 3, the new optimized designs offer a meaningful reduction in expected sample size when the null hypothesis is true. It must be recognized, however, that Fleming did not optimize with regard to the sample size within stages subject to constraints on the error probabilities.

Table 3 Comparison of Two-Stage Designs[a]

p_0	p_1	Type	r_1/n_1	a_1/n_1	r/n	α	β	EN(p_0)
0.05	0.20	Fleming	0/20	4/20	4/40	0.052	0.922	32.5
		Chang	0/20	5/20	4/40	0.047	0.920	32.8
		Optimal	1/21		4/41	0.046	0.902	26.7
0.10	0.30	Fleming	1/15	5/15	5/25	0.036	0.807	19.4
		Optimal	1/10		5/29	0.047	0.805	15.0
0.20	0.40	Fleming	4/20	9/20	11/35	0.037	0.801	25.4
		Chang	7/25	9/25	16/50	0.050	0.814	26.6
		Optimal	3/13		12/43	0.049	0.800	20.6
0.20	0.40	Fleming	4/25	11/25	15/50	0.032	0.904	39.3
		Chang	5/25	10/25	15/50	0.039	0.901	34.2
		Optimal	4/19		15/54	0.048	0.904	30.4
0.30	0.50	Fleming	8/25	14/25	19/45	0.029	0.807	31.3
		Chang	9/25	13/25	21/50	0.032	0.801	29.3
		Optimal	5/15		18/46	0.049	0.803	23.6
0.30	0.50	Fleming	7/25	14/25	20/50	0.048	0.894	37.1
		Chang	6/25	14/25	20/50	0.049	0.899	41.3
		Optimal	8/24		24/63	0.049	0.903	34.7

[a]The hypothesis $p \geq p_1$ is rejected if the number of responses is $\leq r_1$ after n_1 patients or $\leq r$ after n patients. The hypothesis $p \leq p_0$ is rejected if the number of responses is $\geq a_1$ after n_1 patients or $> r$ after n patients.

Chang et al. [14] presented optimized results only for three-stage designs. Also, the latter two designs provide for early rejection of the null hypothesis, a feature that may be useful in some clinical trials.

We have tabulated optimal phase II designs for (α,β) = (0.10,0.10), (0.05,0.20), and (0.05,0.10). In phase II trials, both kinds of error are important. β represents the probability of rejecting a treatment with response rate $\geq p_1$. α represents the probability of failing to reject a treatment with response probability $\leq p_0$. This is a less serious error from a drug discovery viewpoint, but it is serious from a cost perspective since it leads to unnecessary follow-up trials. The tabulated designs should be appropriate for most situations. It is unusual to have $\beta < \alpha$ and designs based on $\alpha = \beta = 0.05$ require too large a sample size for practical use in most phase II trials. Tabulation of the minimax designs should also be useful for those who prefer such designs or who wish to know the smallest value of maximum sample size to use for selecting designs of other types.

For phase II trials of new drugs against solid tumors, designs with (p_0,p_1) equal to (0.05,0.20), (0.05,0.25), or (0.10,0.25) will often be appropriate. This is because many new drugs are almost totally inactive against the common solid tumors. Also, new drugs that provide relatively modest (20%–25%) response rates against these refractory diseases are of interest for further development. In some cases effective treatments can be obtained by combining such drugs with other drugs, by optimizing their schedules or routes of administration, or by using them for patients with less advanced forms of the same histopathological type of cancer. Other tabulated designs will be appropriate for phase II trials of combination regimens. It is for this reason that the full range of (p_0,p_1) was produced. For pilot studies of combinations,

the level $p_1 - p_0 = 0.20$ is commonly the degree of difference targeted. Designing a trial to distinguish only larger differences is often unrealistic and uninformative. A $p_1 - p_0$ of 0.15 is probably the smallest difference that one would consider for a phase II study because the sample sizes become prohibitively large for smaller differences and because the lack of controls limits the interpretability of trials based on distinguishing smaller differences. The designs presented here could also be utilized for pilot studies of intensive regimens with toxicity as the endpoint. In this case the hypotheses should be specified in terms of the probability of no toxic event.

The optimization criterion chosen here is not unique. One could minimize the expected sample size averaged with regard to a prior distribution for the true response probability p. Historically, however, most new regimens are not successful and, more importantly, optimizing the design for performance under the null hypothesis seems ethically appropriate.

As pointed out by DeMets [15], the decision to terminate a controlled clinical trial early is often complex and sequential boundaries are generally to be regarded as guidelines rather than rigid decision rules. This also applies to phase II clinical trials because there are secondary endpoints and sometimes patient subsets of interest. A decision to terminate a phase II trial for a treatment having poor activity for a well defined and fairly homogeneous set of patients, however, is generally less complex than a decision for early termination of a large controlled study.

REFERENCES

1. Simon R: Design, Analysis and Reporting of Cancer Clinical Trials. In: Biopharmaceutical Statistics for Drug Development (Peace, KE, Ed.) New York: Marcel Dekker, 1987
2. Colton T, McPherson K: Two-stage plans compared with fixed-sample-size and Wald SPRT plans. J Am Stat Assoc 71:80–86, 1976
3. McPherson K: On choosing the number of interim analyses in clinical trials. Stat Med 1:25–36, 1982
4. Jennison C: Efficient group sequential tests with unpredictable group sizes. Biometrika 74:155–166, 1987
5. Gehan EA: The determination of the number of patients required in a follow-up trial of a new chemotherapeutic agent. J Chron Dis 13:346–353, 1961
6. Jennison C, Turnbull BW: Confidence intervals for a binomial parameter following a multistage test with application to MIL-STD 105D and medical trials. Technometrics 25:49–58, 1983
7. Chang MN, O'Brien PC: Confidence intervals following group sequential tests. Controlled Clin Trials 7:18–26, 1986
8. Schultz JR, Nichol FR, Elfring GL, Weed SD: Multi-stage procedures for drug screening. Biometrics 29:293–300, 1973
9. Fleming TR: One sample multiple testing procedure for phase II clinical trials. Biometrics 38:143–151, 1982
10. O'Brien PC, Fleming TR: A multiple testing procedure for clinical trials. Biometrics 35:549–555, 1979
11. Lee YJ, Staquet M, Simon R, Cantane R, Muggia F: Two stage plans for patient accrual in phase II cancer clinical trials. Cancer Treat Rep 63:1721–1726, 1979

12. Herson J: Predictive probability early termination plans for phase II clinical trials. Biometrics 35:775–783, 1979
13. Sylvester RJ, Staquet MJ: Design of phase II trials in clinical trials in cancer using decision theory. Cancer Treat Rep 64:519–524, 1977
14. Chang MN, Therneau TM, Wieand HS, Cha SS: Designs for group sequential phase II clinical trials. Biometrics 43:865–874, 1987
15. DeMets DL: Stopping guidelines versus stopping rules: A practitioner's point of view. Communications Stat Theory Meth 13:2395–2417, 1984