

TWO-STAGE GROUP SEQUENTIAL DESIGNS FOR COMPARATIVE CLINICAL TRIALS

Sevil Bacanlı¹, Özgür K. Peker², Yaprak Parlak Demirhan³.

^{1,3} Hacettepe University, Faculty of Science, Department of Statistics, 06800 Beytepe, Ankara, Turkey. sevil@hacettepe.edu.tr

² Anadolu University, Faculty of Science, Department of Statistics, Eskişehir, Turkey.

Abstract- Group sequential designs are widely used in Phase II clinical trials, which are usually undertaken to evaluate the response probability of specific treatment regimen. In most randomized clinical trials with sequential patient entry, fixed sample size design is unjustified on ethical grounds and sequential designs are often impractical. However group sequential designs are generally more practical and they provide much of the saving possible from sequential designs. Optimal restricted two-stage design is the simplest form of a group sequential design.

In this study, group sequential design obtained by $\alpha^*(t)$ functions characterized using the type-I error probability and optimal restricted two stage design has been compared for the cases that the group sizes are equal. Furthermore, their efficiency regarding fixed sample size design has been calculated and the results have been discussed.

Keywords- Group sequential design, Clinical trials, Two-stage design, Efficiency

1. INTRODUCTION

For ethical, scientific and economic reasons, clinical trials are often repeatedly monitored for evidence of treatment benefit or harm. To achieve this, statisticians conduct interim analyses periodically on accumulating data [1]. Repeated testing at conventional critical values can substantially inflate the overall type I error rate, and various group sequential testing procedures have been proposed to achieve the desired levels of type I error [2, 3].

Pocock [2], O'Brien&Fleming [3] were among the first workers to develop group sequential design by modifying the initial work of Armitage [4]. DeMets&Ware [5, 6] considered asymmetric group sequential boundaries adapted from Pocock and O'Brien&Fleming designs and from Wald's sequential probability ratio test. These designs are based on equally spaced analyses. Jennison [7] and Eales&Jennison [8] explored the extent of possible reductions in expected sample size by searching for optimal symmetric group sequential one-sided tests. Barber&Jennison [9] extend the optimal symmetric group sequential tests of Eales&Jennison [8] to the broader class of asymmetric designs. Jennison&Turnbull [10] describe the parametric family of tests proposed by Emerson&Fleming [11] and extended by Pampallona&Tsiatis [12] to include asymmetric tests with unequal Type I and Type II error probabilities. We refer

to the recent book by Jennison&Turnbull [10] for more details about group sequential test.

Lan&DeMets [13], Kim&DeMets [14] proposed the group sequential design obtained by $\alpha^*(t)$, use function, which characterizes spending the type I error probability (α). This design is more useful because it is not necessary that each group has equal observation. So that in this test group size is equal or unequal.

Two-stage group sequential design is the simplest form of a group sequential design. Owen [15] described two-stage tests for one-sided hypotheses about a normal mean with known variance. Hald [16] derived optimal designs for this same problem using minimax and Bayes weighted average optimality criteria. Calton&McPherson [17] considered hypothesis tests for normal and binomial responses and presented optimal two-stage designs, which did not allow acceptance of the null hypothesis at the first stage. Dewith [18] extended the work of Calton and McPherson for binomial responses by developing optimal designs that allowed acceptance or rejection at the first stage none of these designs used the fixed sample critical value at the final stage. Case et. all [19] developed optimal two-stage designs that have the restriction of using the fixed sample critical value at the final stage.

In this paper, we compared the statistical properties of the group sequential design based on $\alpha^*(t)$ function, and optimal restricted two-stage design in the setting of one-sided comparative clinical trials with normal response.

2. GROUP SEQUENTIAL DESIGN BASED ON THE TYPE-I ERROR SPENDING RATE FUNCTION

Lan&DeMets [13] and Kim&DeMets [14] have suggested a flexible group sequential design based on the use function approach. The design, needs only the specification in advance an α spending rate function, $\alpha^*(t)$, which characterizes the rate at which the type-I error probability is spent. The boundary is determined by $\alpha^*(t)$, and by discrete times t_j , $j = 1, \dots, i$; but it does not depend on times t_j , $j > i$ nor the total number of repeated tests, K , to be performed.

$\alpha^*(t)$ function can be defined as,

$$\begin{aligned}\alpha^*(t) &= P\{\tau \leq t, 0 \leq t \leq 1 | H_0\} \\ \alpha^*(t_i) &= P\{\tau \leq t_i | H_0\} \\ &= P\{\tau = t_i | H_0\} + P\{\tau = t_{i-1} | H_0\}\end{aligned}\tag{1}$$

where τ is the first time that a standard Brownian process W_t . In this paper the following $\alpha^*(t)$'s are studied;

$$1. \alpha_1^*(t) = 2 \left[1 - \Phi \left(Z_{1-\alpha/2} / \sqrt{t} \right) \right] \quad 0 \leq t \leq 1\tag{2}$$

$$2. \alpha_2^*(t) = \alpha \left[\ln[1 + (e-1)t] \right] \quad 0 \leq t \leq 1 \quad (3)$$

$$3. \alpha_3^*(t) = \alpha t \quad 0 \leq t \leq 1 \quad (4)$$

$$4. \alpha_4^*(t) = \alpha t^{3/2} \quad 0 \leq t \leq 1 \quad (5)$$

$$5. \alpha_5^*(t) = \alpha t^2 \quad 0 \leq t \leq 1 \quad (6)$$

When the group sizes are equal ($t_k = k/K$ $k = 1, \dots, K$), it generates $\alpha_1^*(t)$, discrete group sequential boundaries approximate to those of O'Brien&Fleming and it generates $\alpha_2^*(t)$, boundaries approximate to those of Pocock.

The clinical trial we shall consider involves the comparison of two normally distributed with means μ_e and μ_c for the experimental and the control treatments respectively, and common variance σ^2 . Such that $X_e \sim N(\mu_e, \sigma^2)$, $X_c \sim N(\mu_c, \sigma^2)$. It is planned as a test of the null hypothesis $H_0 : \mu_e = \mu_c$ against the one-sided alternative $H_0 : \mu_e > \mu_c$. A group sequential test is specified by the maximum number of analysis, K , the total number of observations at each analysis $n_1, \dots, n_k$. Then at the k th analysis, the standardized test statistic after each group of observations is defined as,

$$Z_k = \frac{\sum_{i=1}^{n_k} X_{ie} - \sum_{i=1}^{n_k} X_{ic}}{\sqrt{2\sigma^2 n_k}} \quad k=1, \dots, K \quad (7)$$

$Z_k \sim N(\Delta, 1)$ under H_1 and $\Delta = (\mu_e - \mu_c) \sqrt{n_k} / (\sigma \sqrt{2})$ is the noncentrality parameter with $\delta = \mu_e - \mu_c$, representing the clinically meaningful mean treatment difference to be detected at given α , $1-\beta$ and K . The test statistics Z_k , is compared with group sequential boundaries (C_k) as below following;

1. We stop to reject H_0 if $Z_k \geq C_k$ $k=1, \dots, K-1$, otherwise, we continue to the next stage.
2. We stop to reject H_0 if $Z_k \geq C_k$ $k=K$, Otherwise we stop to accept H_0 .

Therefore, the sample size for each treatment in a clinical trial is determined by $n_k = 2\sigma^2 \Delta^2 / \delta^2$ and maximum sample size is given by $n = 2n_k K$. When $K = 1$, that is for a fixed sample size design, this formula becomes the familiar sample size formula for normal data [10, 13]. But expected sample size, ESS is given by, $ESS = n \bar{\tau}^*$, where $\bar{\tau}^*$ is the expected boundary crossing time. It can be calculated ESS, either under H_0 or H_1 . $\bar{\tau}^*$ goes to 1 under the null hypothesis. So $ESS(H_0) \cong n$.

Constants (C_k), Δ , n , $\bar{\tau}^*$ and ESS under the alternative hypothesis for $K=1, 2$, $\alpha=0.05$, $1-\beta=0.90$ are given in Table 1 for one-sided tests of hypothesis using the the $\alpha^*(t)$ use function, with equal increments of information time. For more complete

tabulations of various constants can be found Reboussin et al. [20] and Jennison&Turnbull [10].

Table 1. Group sequential one-sided designs at given $\alpha = 0.05$, $1 - \beta = 0.90$ and $K=1, 2$ ($t_i = 0.5; 1$)

K	$\alpha^*(t)$	C_k	Δ	n	$\bar{\tau}^*$	ESS
1	$\alpha_1^*(t)$	1.645	2.926	34.24	1	34.24
2	$\alpha_1^*(t)$	2.538	2.079	34.57	0.839	29.01
		1.662				
	$\alpha_2^*(t)$	1.866	2.184	38.17	0.688	26.26
		1.885				
	$\alpha_3^*(t)$	1.960	2.152	37.06	0.712	26.39
		1.807				
	$\alpha_4^*(t)$	2.104	2.119	35.93	0.747	26.84
		1.737				
	$\alpha_5^*(t)$	2.241	2.102	35.35	0.778	27.51
		1.699				

^a Multiply each value by $(\sigma/\delta)^2$

3. OPTIMAL RESTRICTED TWO-STAGE DESIGNS

The two-stage design is the simplest form of a group sequential design. In this design, the null hypothesis $H_0: \theta = \theta_0$ is tested versus the alternative $H_1: \theta > \theta_0$ (or $\theta \neq \theta_0$).

The two-stage design for one-sided test as follows:

Stage 1: Accrue n_1 patients and calculate

$$Z_1 = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \quad (8)$$

where $\hat{\theta}$ is computed from data on the first n_1 patients.

- i) Accept H_0 if $Z_1 < C_1$;
- ii) Reject H_0 if $Z_1 > C_2$;
- iii) Otherwise; continue the second stage.

Stage 2: Accrue an additional n_2 patients. Let $n = n_1 + n_2$ and calculate,

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \quad (9)$$

where $\hat{\theta}$ is computed from data on all n patients.

- i) Accept H_0 if $Z < C_3$;
- ii) Otherwise, reject H_0 .

One may want to test the mean of normal distribution; in this case, Z_1 and Z statistics are defined as following,

$$Z_1 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2\sigma^2/n_1}}$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2\sigma^2/n}}$$

Z_1 and Z are distributed standard normal distribution and their joint distribution is bivariate normal with zero means, unit variances, and correlation $(n_1/n)^{1/2}$. The maximum sample size for the two-stage design is n and is realized whenever a second stage is necessary. The expected sample size (ESS) of the two-stage design is given,

$$ESS(\theta) = n [1 - (1-p)P_s(\theta)] \quad (10)$$

where $P_s(\theta)$ denote the probability that the trial will be stopped at the first stage, and p is the rate of the number of patients at the first stage to the number of total patients at the second stage $p = n_1/n$. θ value can be computed for θ_0, θ_1 or $\theta_{\max}$, where θ_0 is the θ value when H_0 is true; θ_1 is the θ value when H_1 is true; and $\theta_{\max}$ is the maximum value of θ .

There are five unknown parameters in the two-stage design, namely: n_1 , n_2 , C_1 , C_2 and C_3 . The critical value at the second stage, C_3 , will be set to equal that of the fixed sample test

$$C_3 = \varphi^{-1}(1-\alpha) \quad (\text{or } \varphi^{-1}(1-\frac{\alpha}{2})) \quad (11)$$

where $\varphi(x)$ denotes the standard normal distribution function. The other four parameters of interest are chosen to satisfy the two equations:

$$\alpha = 1 - \Phi(C_2) + B(C_1, C_2; C_3, \infty; p) \quad (12)$$

$$1 - \beta = 1 - \Phi(C_2 - u\sqrt{p}) + B(C_1 - u\sqrt{p}, C_2 - u\sqrt{p}; C_3 - u, \infty; p) \quad (13)$$

where,

$$B(a, b, c, d, p) = \left(1/2\pi\sqrt{1-p}\right) \int_a^b \int_c^d \exp\left\{-(1/2)(1-p)(y^2 - 2\sqrt{pyz} + z^2)\right\} dy dz$$

and $u = \sqrt{n}(\theta_1 - \theta_0)/\sigma$.

So that the probability of rejecting H_0 at the first stage plus the probability of continuing the trial and rejecting H_0 at the second stage is equal to α , when assuming H_0 is true. The desired power of the trial $1-\beta$ is the same probability under the alternative hypothesis. Eq. (12) and Eq. (13) are solved iteratively by numerical

integration of the bivariate normal distribution using a double precision function [19, 21, 22].

With five parameters and only three constraints given by Eq. (11), (12), (13) optimality criteria are used to determine the parameter values. So, this test is called optimal restricted two-stage design. In this study, we have examined Minimax and Bayes criteria.

Minimax Criterion:

It is described as minimizing the maximum expected sample size with respect to the five design parameters ($\min[\max \text{ESS}(\theta)]$). For the two-stage design $\theta_{\max}$ is;

$$\theta_{\max} = \theta_0 + \frac{(C_1 + C_2)\sigma}{2\sqrt{n_1}} \quad (14)$$

When $\theta_{\max}$ is replaced into the $\text{ESS}(\theta)$ function, the function which will be minimized for one-sided test is:

$$\text{minimize } \text{ESS}(\theta_{\max}) = n \left[1 - (1-p) \left(\Phi \left(\frac{C_1 - C_2}{2} \right) + 1 - \Phi \left(\frac{C_2 - C_1}{2} \right) \right) \right] \quad (15)$$

Bayes Criterion:

Minimize a weighted average of the ESS under H_0 and the H_1 ,

$$\text{minimize } \text{ESS}_w(\theta) = (1-w)\text{ESS}(\theta_0) + w\text{ESS}(\theta_1) \quad (16)$$

Using a weight of 0 for this criterion gives the most efficient designs if the null hypothesis is true while a weight of 1 gives the most efficient designs if the specified alternative is true [19, 21].

The design parameters (C_1, C_2, C_3, p), the probabilities $p_s(\theta)$ and maximum (n) and expected sample sizes (ESS) obtained using the different criteria are given in Table 2 for $\alpha=0.05, 1-\beta=0.90$. The choice of p is sometimes determined by factors unrelated to optimal designs. For some studies it might be practical to choose equal samples at each stage. Therefore, if $p=0.50$, each stage has equal sizes. Table 2 gives the optimal design parameter for $p=0.50$. In tables, n_f denotes the sample size required using a fixed sample design.

Table 2. Optimal Restricted Two-Stage One- Sided Designs for Minimax and Bayes Criterion at given $\alpha = 0.05, 1-\beta = 0.90$.

Criterion	p	C ₁	C ₂	C ₃	P _s (θ)	n _f ^a	n ^a	ESS(θ ₀) ^a	ESS(θ ₁) ^a
Minimax	0.588	0.819	2.086	1.645	0.527	34.256	38.024	25.316	27.508
	0.500	0.667	2.130	1.645	0.464	34.256	39.088	24.152	27.472
Bayes	w = 0	0.382	0.474	2.168	1.645	0.697	34.256	41.280	23.500
		0.500	0.595	2.178	1.645	0.739	34.256	38.230	24.116
	w = 1	0.540	0.737	2.111	1.645	0.629	34.256	38.572	24.596
		0.500	0.700	2.109	1.645	0.609	34.256	39.497	24.185

^a Multiply each value by $(\sigma/\delta)^2$

Example:

In this section, we described a clinical trial to evaluate the potential chemoprevention agent tamoxifen in women with early stage breast cancer and showed how to determine the maximum sample size for group sequential and optimal restricted two-stage design.

The drug tamoxifen was evaluated at the University of Wisconsin-Madison as a potential chemoprevention agent in women with early stage breast cancer and compared with a placebo in a double blind clinical trial [23]. In addition to evaluating the recurrence of cancer, this study measured the effect the drug might have on the lipid cholesterol level. If the level were decreased, a reduction in cardiovascular risk would be induced. If the cholesterol level were increased, the drug might be increasing the risk of heart disease and thus diminish the potential benefit. For women eligible for this study, the mean cholesterol level on the placebo was expected to be 220 mg/dl with a standard deviation of 30 mg/dl. A clinically significant change induced by the drug was thought to be a 20 mg/dl decrease or increase in mean cholesterol level.

Suppose we wish to design this trial, using a 5 per cent significance level for a one-sided test of hypothesis with 90 per cent power to distinguish between mean cholesterol levels of 220 mg/dl and 200mg/dl. Assuming that the measurement of cholesterol level is normally distributed, the maximum sample size determination for group sequential design based on $\alpha_i^*(t)$ use function, $i=1,\dots,5$ and optimal restricted two-stage design are given below;

$\alpha^*(t)$ use function	n	Criteria	n
$\alpha_1^*(t)$	77.80	Minimax	87.94
$\alpha_2^*(t)$	85.86	Bayes	
$\alpha_3^*(t)$	83.36	w = 0	86.02
$\alpha_4^*(t)$	80.84	w = 1	88.87
$\alpha_5^*(t)$	79.53		

4. COMPARISON OF THE TEST DESIGNS AND RESULTS

In this section, efficiency of the designs regarding fixed sample size design has been examined. Therefore, the obtained expected sample sizes have been used for the cases in which the group sizes are equal for $t = 0.50, 1$ in group sequential design (Table 1) and $p = 0.50$ in optimal restricted two-stage design (Table 2).

The expected sample sizes of the designs relative to the fixed sample size are called the efficiency.

The efficiencies of the tests are computed as;

- Under H_1 hypothesis, $R_1 = \frac{ESS(\theta_1)}{n_f}$
- Under H_0 hypothesis, $R_0 = \frac{ESS(\theta_0)}{n_f}$
- The efficiency of the designs for equal group sizes have been presented in Table 3 at given $\alpha = 0.01, 0.05$ and $1 - \beta = 0.80, 0.90, 0.95$.

Table 3. Efficiencies for group sequential designs and two-stage designs^a
at given α , $1 - \beta$

α	$1-\beta$	$\alpha_1^*(t)$	$\alpha_2^*(t)$	$\alpha_3^*(t)$	$\alpha_4^*(t)$	$\alpha_5^*(t)$	T-S D.	w = 0	w = 1
0.01	0.80	0.948	0.867	0.862	0.866	0.879	0.853	0.861	0.853
		1.003	1.114	1.076	1.043	1.031	0.629	0.619	0.632
	0.90	0.912	0.789	0.789	0.799	0.814	0.822	0.828	0.821
		1.002	1.103	1.068	1.038	1.023	0.629	0.624	0.635
	0.95	0.875	0.730	0.733	0.744	0.760	0.785	0.789	0.783
		1.003	1.095	1.064	1.036	1.021	0.632	0.629	0.639
0.05	0.80	0.900	0.847	0.847	0.855	0.869	0.837	0.840	0.837
		1.008	1.111	1.075	1.047	1.032	0.699	0.696	0.700
	0.90	0.847	0.767	0.771	0.784	0.803	0.802	0.804	0.802
		1.006	1.098	1.066	1.040	1.026	0.705	0.704	0.706
	0.95	0.797	0.708	0.712	0.727	0.748	0.765	0.765	0.764
		1.006	1.088	1.058	1.035	1.022	0.712	0.711	0.713

^a Top number is R_1 (when $\theta = \theta_1$)

Bottom number is R_0 (when $\theta = \theta_0$)

When Table 3 is examined, we obtain the following results:

- Since Group Sequential designs mentioned in this study do not allow acceptance of H_0 until the final stage, their performance under the null hypothesis is poor. They perform much better under the alternative hypothesis.
- However, in each case, the ESS is the smallest for the optimal restricted two-stage designs under the null hypothesis.
- Moreover, the ESS under the alternative hypothesis for the optimal restricted two-stage designs is usually smaller than that of the Group Sequential designs.
- $\alpha_2^*(t)$ and Pocock's designs performance is better than other designs under the alternative hypothesis for large powers.
- If we examined the tables in second and third sections; the expected sample sizes are approximately the same in two-stage designs and group sequential designs under the alternative hypothesis. However, the expected sample sizes are rather small in two-stage designs under the null hypothesis. So, two-stage designs result approximately in an expected % 40 savings in the expected sample size under the null hypothesis.
- So, if we compare the optimal restricted two-stage designs with the group sequential designs in case of $N = 2$, we can say two-stage design is preferable in terms of sample sizes and performance.

- However, the test design to be used in a clinical trial is generally predetermined. The existence of parameters in an optimal two-stage design has some restrictions. Therefore, it is more advisable to use group sequential design in clinical trials where there is a probability of continuation.

REFERENCES

1. K. Kim, D.L. DeMets, Sample size determination for sequential clinical trials with immediate response, *Statistics in Medicine*, **11**, 1391-1399, 1992.
2. S.J. Pocock, Group sequential methods in the design and analysis of clinical trials, *Biometrika*, **64**, 191-199, 1977.
3. P.C.O'Brien, T.R.Fleming, A multiple testing procedure for clinical trials. *Biometrics*, **35**, 549-556, 1979.
4. P. Armitage, *Sequential medical trials*, 2nd ed. Blackwell/Oxford, 1975.
5. D.L. DeMets, J.H. Ware, Group sequential methods for clinical trials with a one-sided hypothesis, *Biometrika*, **67**, 651-660, 1980.
6. D.L. DeMets, J.H. Ware, Asymmetric group sequential boundaries for monitoring clinical trial, *Biometrika*, **69**, 661-663, 1982.
7. C.Jennison, Efficient group sequential test with unpredictable group sizes, *Biometrika*, **74** 155-65, 1987.
8. J.D. Eales, C. Jennison, An improved method for deriving optimal one-sided group sequential tests, *Biometrika*, **79**, 13-24, 1992.
9. S. Barber, C. Jennison, Optimal asymmetric one-sided group sequential tests, *Biometrika*, **89**, 49-60, 2002.
10. C. Jennison, B.W. Turnbull, *Group Sequential Methods with Application Clinical Trials*, Chapman and Hall/ Boca Raton, 2000.
11. S.S., Emerson, T.R., Fleming, Symmetric group sequential test design, *Biometrics*, **45**, 905-23, 1989.
12. S. Pampallona, A. A. Tsiatis. Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favour of the null hypothesis, *J. Statist Plan. Infer* **42**, 19-35, 1994.
13. K.K.G.Lan, D.L. DeMets, Discrete sequential boundaries for clinical trials, *Biometrika*, **70**, 659-663, 1982.
14. K. Kim, D.L. DeMets, Design and analysis of group sequential tests based on type-I error spending rate function, *Biometrika*, **74**, 149-154, 1987.
15. D.B. Owen, A double sample test procedure, *Ann. Math. Stat*, **24**, 449-457, 1953.
16. A. Hald, Optimum double sampling tests of given strength, The normal distribution, *J. Am. Stat. Assoc.* **70**, 451-456, 1975.
17. T. Calton, K. McPherson, Two-stage plans compared with fixed-sample size and Wald SPRT plans, *J. Am. Stat. Assoc.* **71**, 80-86, 1976.
18. C. DeWith, Two-stage plans for the testing of binomial parameters, *Contr. Clin. Trials*, **4**, 215-226, 1983.
19. L.D. Case, T.M. Morgan, C.E. Davis, Optimal restricted two-stage designs, *Contr. Clin. Trials*, **8**, 146-156, 1987.

20. D.M. Reboussin, D.L. DeMets, K. Kim, K.K.G. Lan, *Programs for computing group sequential bounds using the Lan-Demets Method*, Technical Report, **60**, University of Wisconsin Department of Biostatistics, 45 p.1992.
21. L.D. Case, T.M. Morgan, C.E. Davis, Three-stage designs for monitoring clinical trials, *Communications in Statistics-Theory and Methods*, **23 (7)**, 1875-1893, 1994.
22. K.Ö. Peker, S.Bacanlı, Optimal restricted two-stage design, *Anadolu University Journal of Science and Technology*, **1(1)**, 95-103.
23. R.R. Love, R.B. Mazess, H.S. Barden, S. Epstein, P.A. Newcomb, V.C. Jordan, P.P. Carbone, D.L. DeMets, Effects of tamoxifen on bone mineral density in postmenopausal women with breast cancer, *New England Journal of Medicine*, **326**, 852-856, 1992.