Sample Size in Adaptive Design with Treatment Selection

Zejiang Yang

Biostatistics Consultancy Group, Syneos Health™
3201 Beechleaf Court, Suite 600, Raleigh, NC 27604

Abstract
Treatment selection method is an important application of adaptive design in clinical studies. This paper will extend the simulation-based approach and Bonferroni-Holm test procedure in Posch *et al.* (2011) to a normally distributed endpoint assuming varying selection rules. We assume that the study has two active treatments and a control group with one interim analysis, and use the following treatment selection rules based on interim analysis results: (1) select one active treatment with a better response than other active treatment by a pre-specified threshold; (2) early termination is allowed if it meets the pre-specified criteria. How to determine critical boundaries to maintain Type I error under these scenarios, and how to calculate the sample size to achieve the target statistical power will be presented.

Key Words: treatment selection, sample size, simulation, Bonferroni, Holm

1. Introduction

In clinical studies, treatment selection method is a major application of adaptive design. It usually consists of two stages: a selection stage (first stage) and a confirmation stage (second stage). These studies start with several active treatment groups and a control group, and then use pre-specified rules to select treatment(s) for the second stage based on the interim analysis results.

The major statistical question and concern are how to control the overall Type I error. Posch *et al.* (2011) used simulation-based approach and adaptive Bonferroni-Holm (B-H) test procedure to control the overall Type I error rate for a binomial distributed endpoint. This paper will extend these methods to a normally distributed endpoint assuming varying selection rules. In this paper, we assume that the study has two active treatments and a control group with one interim analysis. Based on the results of interim analysis, Treatment Selection Rule A is to select one active treatment with a better response than other active treatment by a pre-specified threshold; (2) Treatment Selection Rule B allows the early termination if it meets the pre-specified criteria in addition to Treatment Selection Rule A. Using the simulation-based approach and adaptive B-H procedure with two different treatment selection rules, this paper will show how to determine critical boundaries to maintain Type I error under varying scenarios, and how to calculate the sample size to achieve the target statistical power with different study design characteristics.

2. Statistical Setting

In this paper, we assume that a clinical study has two active treatments with different dose levels: low dose (L) and high dose (H), and a control group (C). The primary

endpoint has a normal distribution, and denote them as $X_L \sim N(\mu_L, \sigma^2)$ for the low dose group, $X_H \sim N(\mu_H, \sigma^2)$ for the high dose group, $X_C \sim N(\mu_C, \sigma^2)$ for the control group respectively. Furthermore, we may also assume that a larger value for primary endpoint indicates a better condition. Thus the one-sided null hypotheses are:

$$H_{L0}: \mu_L=\mu_C \quad \text{and } H_{H0}: \mu_H=\mu_C$$

Against the one-sided alternatives:

$$H_{LA}: \mu_L>\mu_C , \qquad H_{HA}: \mu_H>\mu_C$$

At the first stage, the subjects are randomized into 3 treatments in 1:1:1 ratio. After the primary endpoint is observed for n_1 subjects per group, an interim analysis will be performed.

Let M_C, M_L, M_H be the sample means of primary endpoint at the interim analysis for control group, low dose group and high dose group respectively. The Z-statistics based on the first stage data are:

$$Z_{L1} = \frac{M_L - M_C}{\sigma\sqrt{\frac{2}{n_1}}} \qquad Z_{H1} = \frac{M_H - M_C}{\sigma\sqrt{\frac{2}{n_1}}}$$

In a clinical study, the pooled estimate of standard deviation can be used in the calculation of Z-statistics. Let f be the non-negative threshold, the decision-making will be based on the following treatment selection rules:

Selection Rule A:

(1) If $([M_H-M_L]/\sigma) > f \geq 0$, then the lower dose will be dropped for second stage;
(2) If $([M_L-M_H]/\sigma) > f \geq 0$, then the higher dose will be dropped for second stage;
(3) If the conditions (1) or (2) is not met, both active treatments will be kept in second stage;
(4) The control group will always be kept in second stage.

Selection Rule B:

In addition to Selection Rule A, if the larger Z-statistics $max(Z_{L1}, Z_{H1})$ based on the first stage data is less than the pre-specified value T_p, then the study is terminated.

In the second stage, n_2 subjects at each remaining treatment group will complete the study, and the primary endpoint will be observed for the additional n_2 subjects per group. Of note, when the threshold f=0, we will always select only one active treatment with the better response at the interim analysis.

3. Simulation-Based Approach

3.1 Simulation-Based Critical Boundary

The underlying simulation method computes the adjusted critical boundary by simulating a large number of clinical trials assuming that the primary endpoint has the same mean for each treatment group. Each simulated trial follows the pre-determined decision rules based on the results of interim analysis. The adjusted critical boundary C_{SB} is computed such that the proportion of simulated trials, where at least one hypothesis (H_{L0} or H_{H0}) can be rejected, does not exceed α. That is, let Z_L and Z_H denote the test statistics to test H_{L0} or H_{H0} respectively, H_{L0} (or H_{H0}) is rejected if the corresponding $Z_L \geqslant C_{SB}$ (or $Z_H \geqslant C_{SB}$) for some critical boundary C_{SB}.

The following steps are used to calculate the adjusted critical boundary (C_{SB}):

(1) Since the test statistic depends only on the standardized treatment difference (effect size), without loss of generalization, we may assume $\mu_L = \mu_H = \mu_C = 0$ with the common standard deviation 1.0;
(2) At first stage of the study, generate the primary endpoint mean value from the normal distribution $N(0, 1/n_1)$; then use the decision rules in Section 2 for treatment selection;
(3) At second stage of the study, generate the primary endpoint mean value from the normal distribution $N(0, 1/n_2)$ for remaining treatment groups;
(4) Compute the test statistic based on the data from both stages. The test statistics Z_L (or Z_H) is the sample mean difference between active treatment and control group divided by its sample standard deviation;
(5) Set $Z_M = max(Z_L, Z_H)$;
(6) Repeat the above process a large number of times.

The adjusted critical boundary C_{SB} is defined as the 100(1-α)th percentile of the pool of all Z_M values. The following table shows the simulated critical boundary under some assumptions based on 100,000 simulations with $\alpha=0.025$. The critical boundary depends on the threshold f and the sample size per group in each stage (n_1 for stage 1 and n_2 for stage 2). For Treatment Selection Rule B, the simulated critical boundary also depends on the pre-specified value T_p. If the sample size is re-estimated based on the interim analysis results, the critical boundary will also be adjusted accordingly to maintain the overall Type I error rate.

Table 1: Adjusted Critical Boundaries

T_p	f	n_1	n_2	C_{SB}
	0	100	100	2.1717
	0.1	100	100	2.2035
	0.2	100	100	2.2042
	0.3	100	100	2.2154
	0	100	200	2.1494
	0.1	100	200	2.1879
	0.2	100	200	2.2120
	0.3	100	200	2.2090
0.5	0	100	100	2.1541
0.5	0.1	100	100	2.1867
0.5	0.2	100	100	2.2016
0.5	0.3	100	100	2.2135
0.5	0	100	200	2.0937
0.5	0.1	100	200	2.1500
0.5	0.2	100	200	2.1705
0.5	0.3	100	200	2.1797

3.2 Simulation-Based Statistical Power

Based on the simulated critical boundary, the statistical power can be calculated. Once again, without loss of generalization, we may assume the mean for control group $\mu_C=0$, and the common standard deviation=1.0. Let n_1 and n_2 be the sample size per group at stage 1 and 2 respectively. For threshold=0 or threshold=0.1, and $T_p=0.5$ (for Treatment Selection Rule B), Table 2 shows the following simulation results based on 100,000 simulations under difference scenarios:

(1) D_L: the probability that the low dose group is dropped at the interim analysis;
(2) D_H: the probability that the high dose group is dropped at the interim analysis;
(3) P_L: the power to reject null hypothesis H_{L0};
(4) P_H: the power to reject null hypothesis H_{H0};
(5) P_{any}: the power to reject any elementary null hypothesis (H_{L0} or H_{H0}).

Table 2. Statistical Power Using Simulation-Based Approach

T_p	f	μ_L	μ_H	n_1	n_2	D_L (%)	D_H (%)	P_S (%)	P_L (%)	P_H (%)	P_{any} (%)
	0	0	0	100	100	50.3	49.7		1.3	1.3	2.5
	0	0.2	0.2	100	100	50.3	49.7		33.8	33.8	55.2
	0	0.2	0.4	100	100	92.1	7.9		29.2	93.8	95.6
	0	0.2	0.2	100	200	50.3	49.7		42.8	42.9	71.3
	0	0.2	0.4	100	200	92.1	7.9		33.9	96.6	98.6
	0.1	0.2	0.2	100	100	23.8	24.3		37.9	37.6	57.6
	0.1	0.2	0.4	100	100	75.9	1.7		29.2	95.5	96.4
	0.1	0.2	0.2	100	200	23.8	24.3		51.4	50.9	74.7
	0.1	0.2	0.4	100	200	75.9	1.7		34.7	98.7	99.5
0.5	0	0.2	0.2	100	100	54.0	53.7	7.6	33.9	34.1	55.4
0.5	0	0.2	0.4	100	100	92.3	8.4	0.7	29.6	93.5	95.3
0.5	0	0.2	0.2	100	200	54.0	53.7	7.6	43.3	43.7	71.0
0.5	0	0.2	0.4	100	200	92.3	8.4	0.7	35.7	96.2	98.2
0.5	0.1	0.2	0.2	100	100	30.3	30.6	7.5	38.1	38.1	57.8
0.5	0.1	0.2	0.4	100	100	76.4	2.4	0.7	29.6	95.4	96.2
0.5	0.1	0.2	0.2	100	200	30.3	30.6	7.5	50.9	50.6	73.3
0.5	0.1	0.2	0.4	100	200	76.4	2.4	0.7	35.7	98.1	98.9

As can be seen from the first row of Table 2 that the Type I error is maintained. With the increase of threshold f, the likelihood of dropping one active treatment decreases. Thus the total sample size increases, and the statistical power also tends to increase.

4. Adaptive Bonferroni-Holm (B-H) Procedure

The adaptive Bonferroni-Holm (B-H) procedure (Posch *et al.* , 2011) is an alternative to the simulation-based approach. If no treatment is dropped at the interim analysis, the classical Bonferroni-Holm procedure will be performed. Only if a treatment is dropped, the adaptive Bonferroni-Holm procedure based on the partial conditional error rate will be used.

The classical Bonferroni-Holm procedure is equivalent to a closed test procedure when applying to the two elementary hypotheses. The intersection hypothesis of H_{L0} and H_{H0} is

rejected if *max* (Z_L, Z_H)≥$z_{1-\alpha/2}$. If the intersection hypothesis is rejected, each elementary null hypothesis is tested at significance level of α.

The adaptive Bonferroni-Holm test is based on partial conditional error (PCE) rate:

$$A_i^{(\gamma)} = 1 - \Phi\left(\frac{z_{1-\gamma} - w_1 z_{i1}}{w_2}\right)$$

Where z_{i1} (i = L or H) is the Z-statistics based on the first stage data and

$$W_i = \sqrt{n_i/(n_1 + n_2)}$$

If the p-value based on the second stage data for the continued group $p_{2,C} < A_L^{(\alpha/2)} + A_H^{(\alpha/2)}$ and $p_{2,C} < A_C^{(\alpha)}$, then reject null hypothesis H_{C0} (C=L or H). Posch *et al.* (2011) shows that Type I error can be controlled using adaptive B-H procedure. Table 3 reports simulation results (100,000 simulations) using adaptive B-H procedure for Treatment Selection Rule B.

Table 3: Statistical Power Using Adaptive B-H Procedure

T_p	f	μ_L	μ_H	n_1	n_2	D_L (%)	D_H (%)	P_S (%)	P_L (%)	P_H (%)	P_{any} (%)
0.5	0	0	0	100	100	77.1	77.3	54.4	1.1	1.2	2.3
0.5	0	0.2	0.2	100	100	54.0	53.7	7.6	9.7	27.8	37.5
0.5	0	0.2	0.4	100	100	92.3	8.4	0.7	2.8	89.3	92.1
0.5	0	0	0	100	200	77.1	77.3	54.4	1.1	1.1	2.2
0.5	0	0.2	0.2	100	200	54.0	53.7	7.6	7.2	35.2	42.4
0.5	0	0.2	0.4	100	200	92.3	8.4	0.7	2.0	91.4	93.4
0.5	0.1	0	0	100	100	68.0	68.2	54.6	1.3	1.3	2.3
0.5	0.1	0.2	0.2	100	100	30.3	33.6	7.5	28.7	37.7	47.4
0.5	0.1	0.2	0.4	100	100	76.4	2.4	0.7	16.0	94.5	95.3
0.5	0.1	0	0	100	200	68.0	68.2	54.6	1.2	1.2	2.1
0.5	0.1	0.2	0.2	100	200	30.3	30.6	7.5	36.0	49.6	57.8
0.5	0.1	0.2	0.4	100	200	76.4	2.4	0.7	18.4	97.4	97.9

5. Sample Size Determination

We may assume that an interim analysis will be performed when the primary endpoint is observed for 50% planned subjects. Under various scenarios, one active treatment termination at the interim analysis is possible. Thus we will calculate the total sample size instead of the sample size per treatment group at each stage. Table 4 reports the total sample size needed to achieve the target statistical power (for rejecting null hypothesis H_{L0} or H_{L0}) under different scenarios based on simulation-based approach (100,000 simulations). Table 5 provides the total sample size needed to achieve the target statistical power under different scenarios based on the adaptive Bonferroni-Holm procedure.

Table 4: Sample Size Results Using Simulation-Based Approach

Selection Method	T_p	f	μ_L	μ_H	Critical Boundary	Statistical Power	Total Sample Size
Rule A		0	0.2	0.2	2.1618	80%	888
		0	0.2	0.2	2.1618	90%	1176
		0	0.2	0.4	2.1618	80%	300
		0	0.2	0.4	2.1618	90%	396
		0.1	0.2	0.2	2.1957	80%	948
		0.1	0.2	0.2	2.1944	90%	1284
		0.1	0.2	0.4	2.1712	80%	300
		0.1	0.2	0.4	2.1824	90%	408
		0.2	0.2	0.2	2.2043	80%	996
		0.2	0.2	0.2	2.2072	90%	1344
		0.2	0.2	0.4	2.1893	80%	312
		0.2	0.2	0.4	2.1839	90%	420
Rule B	0.5	0	0.2	0.2	2.1207	80%	888
	0.5	0	0.2	0.2	2.1207	90%	1188
	0.5	0	0.2	0.4	2.1207	80%	300
	0.5	0	0.2	0.4	2.1207	90%	408
	0.5	0.1	0.2	0.2	2.1617	80%	936
	0.5	0.1	0.2	0.2	2.1802	90%	1296
	0.5	0.1	0.2	0.4	2.1548	80%	312
	0.5	0.1	0.2	0.4	2.1552	90%	408
	0.5	0.2	0.2	0.2	2.1934	80%	996
	0.5	0.2	0.2	0.2	2.2073	90%	1356
	0.5	0.2	0.2	0.4	2.1802	80%	324
	0.5	0.2	0.2	0.4	2.1807	90%	432

Table 5: Sample Size Results Using Adaptive B-H Procedure

T_p	f	μ_L	μ_H	Statistical Power	Total Sample Size
0.5	0	0.2	0.2	80%	2148
0.5	0	0.2	0.2	90%	3708
0.5	0	0.2	0.4	80%	312
0.5	0	0.2	0.4	90%	456
0.5	0.1	0.2	0.2	80%	1068
0.5	0.1	0.2	0.2	90%	1476
0.5	0.1	0.2	0.4	80%	288
0.5	0.1	0.2	0.4	90%	396
0.5	0.2	0.2	0.2	80%	888
0.5	0.2	0.2	0.2	90%	1188
0.5	0.2	0.2	0.4	80%	288
0.5	0.2	0.2	0.4	90%	384

As can be seen from Table 4, the total sample sizes using simulation-based approach for two treatment selection rules are very close under different assumptions. Compared with Table 5, the total sample size using simulation-based approach and adaptive B-H

procedure are also quite close except for the case when the threshold f=0 and two active treatments have the same mean value of 0.2.

Though there is an obvious difference in statistical power between the two approaches when f=0, μ_C=0, μ_L=0.2 and μ_H=0.2, the following figures show that simulation-based approach and adaptive B-H procedure provide very similar results in all other situations.

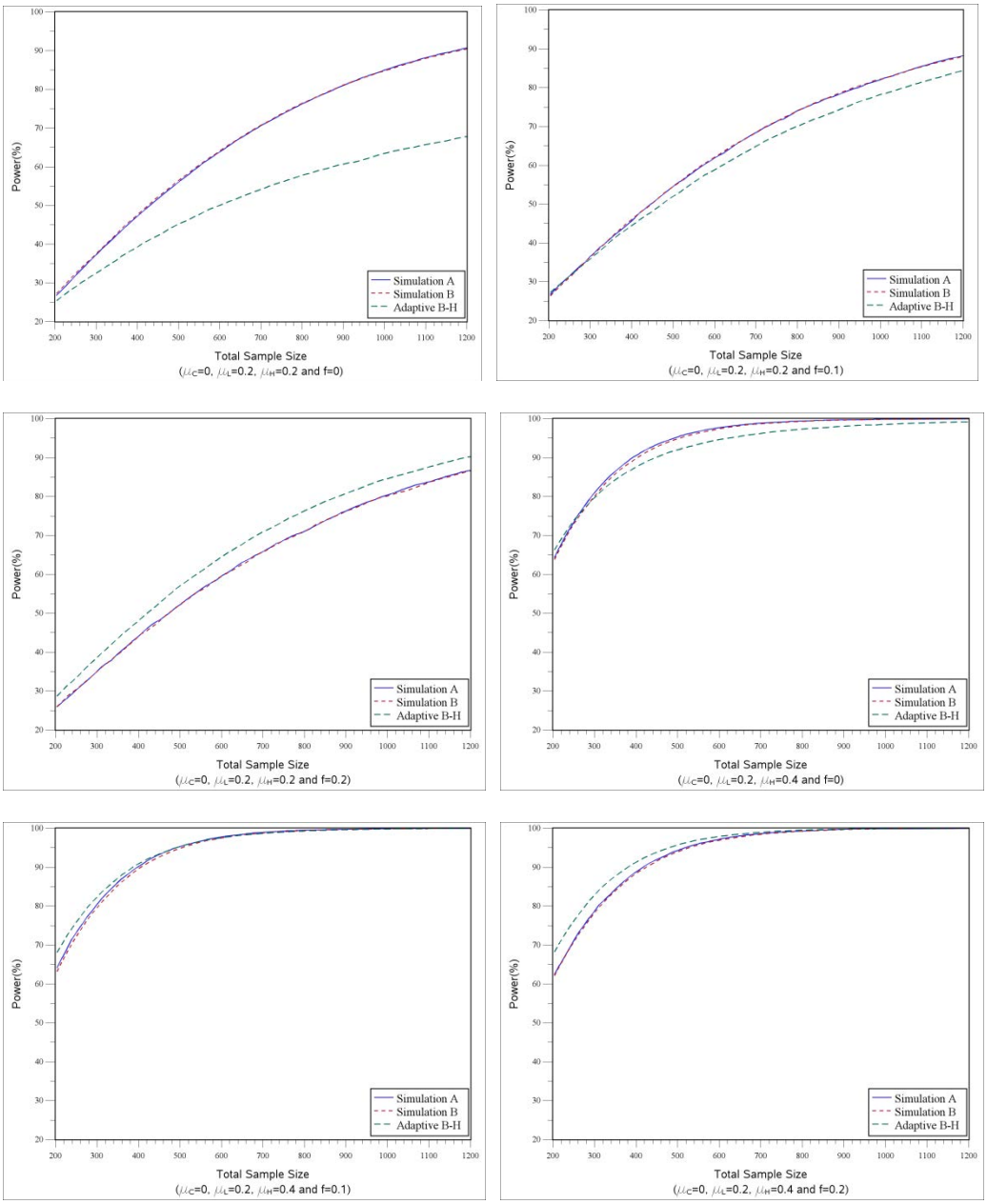

Figure 1: Statistical Power versus Total Sample Size (μ_C=0, μ_L=0.2, μ_H=0.2 or 0.4, f=0, 0.1 or 0.2)

6. Summary

In this paper, we discuss the simulation-based approach and adaptive B-H procedure for a study with two active treatment groups and a control group. Based on the results at interim analysis, two treatment selection rules are used: (1) select one active treatment with a better response than other active treatment by a pre-specified threshold; (2) early termination of study is allowed if Z-statistics at the interim analysis meet the pre-specified criteria. This paper presented how to determine critical boundaries to maintain Type I error under different scenarios, and how to calculate the sample size to achieve the target statistical power. In most situations, the simulation-based approach and adaptive B-H procedure provide very similar results. These approaches can be easily extended to those studies with more than two active treatments.

Acknowledgements

I am very grateful to my colleagues in Biostatistics Consultancy Group in our company for their careful reviews, constructive comments and very helpful suggestions, in particular to Jill Marsden, Donna Levy, Ülker Aydemir, Nick Wells and Anne Esler.

Reference

Posch, M., Maurer, W., and Bretz, F. 2011. Type I error rate control in adaptive designs for confirmatory clinical trials with treatment selection at interim. Journal of Pharmaceutical Statistics. 10: 96-104.