

## On Some Aspects of Conditional Power Evaluation In Two-phase Clinical Trials Under Linear Regression

A. S. Hedayat

Department of Mathematics, Statistics, and Computer Science  
University of Illinois at Chicago

Bikas K. Sinha

Bayesian & Interdisciplinary Research Unit (BIRU)  
Indian Statistical Institute, Kolkata  
bksinha@isical.ac.in

Li Wei

Global Biometric Science  
Bristol-Myers Squibb Company

### Abstract

We investigate the conditional power under the framework of linear regression models so that it can be applied to most actual clinical trials in which multiple treatment effects and covariate effects are included. It is well known that the standard power of a regular test for a treatment contrast depends on unknown parameters only through the contrast itself. However it is not true in general for conditional power. Conditions for this to happen are established here and some instances are illustrated. We also show that similar arguments can be made about the sufficient statistics for the conditional power.

**Keywords:** Interim analysis, Regression models, Conditional probability.

### 1. Introduction

For ethical and financial reasons, early modifications or in rare cases early termination of the study is often desirable in a long-term clinical trial if the accumulated data during the study has already shown overwhelming evidence of efficacy or the null result seems inevitable. This concern has motivated many statisticians to investigate when and how the accrued data should be examined and under what conditions the clinical trial should be terminated. Such a statistical inferential process performed at interim stages in a trial is called interim analysis. It has attracted more and more attention of both academic and

---

Research was primarily sponsored by the National Science Foundation Grants DMS-0603761 and DMS-0904125, and the NIH Grant P50-AT00155 (jointly supported by the National Center for Complementary and alternative Medicine, the Office of Dietary Supplements, the Office for Research on Women's Health, and the National Institute of General Medicine). The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. Professor Hedayat is now retired from ISI wef April 1, 2012.

industrial researchers over the past decades, and a number of statistical approaches have been developed and are made available in the literature.

One class of these general methods is identified as group-sequential procedures. Armitage, Mcpherson and Rowe (1969) first introduced a sequential method using repeated significance testing. The idea is that after every observation, a two-sided significance test is conducted. If a nominal significance level  $\alpha'$  is obtained, stop the trial in favor of the alternative hypothesis. Here  $\alpha'$  and the maximum number of observations are chosen so that the overall significance level  $\alpha$  and the power are maintained at the pre-specified levels. However, this method is criticized for its need for continuous unblinding of the data and assessment after every observation, which is very difficult to meet in practical clinical trials. To circumvent this disadvantage, Pocock (1977) modified Armitage's scheme and proposed a group sequential method in which the data are tested at equally-spaced intervals, that is, after every equally divided group of patients enrolled in the trial. The sample size of each group and the nominal significance are determined to control the overall type I and type II error rate as required. Later, DemMts and Ware (1980) extended Pocock's results to trials with a one-sided hypothesis. Gould (1982) further generalized these procedures by bringing two-end stopping boundaries to allow early termination of a trial with acceptance of null hypothesis. Following the above fundamental contributions on this topic, additional theory has also been developed for more complex models in recent years. Lee and DeMets (1991) obtained group sequential tests for a linear mixed-effects model. Tsiatis (1982) worked on parametric survival models. Gu and Ying(1995) investigated Cox's proportional hazards regression model. Gange and DemMets (1996) considered the sequential analysis of correlated response with non-normal distributions using the generalized estimating equations. For a more comprehensive review of these developments, the interested reader may refer to Jennison and Turnbull (2000).

In most cases, sequential approaches will require less patients to be enrolled than a fixed sample size design. However, the complexity in design and realization as well as the administrative burden obscured this advantage so that their application are relatively limited in many clinical trials, especially those conducted in multiple centers. Practitioners prefer methods which are simple to be implemented and interpreted. We will not discuss the sequential family further in this paper, turning our main interest to another important class of strategies adopted in interim analysis.

The second cluster of statistical tools arise from a very natural question: "Based on the data accrued so far, what is the likelihood of a significant result if the trial is completed?" Unlike those sequential procedures which are purely driven by interim data, these latter approaches combine the current data and the potential future outcome together. The key consideration lies in the calculation of the conditional power, which is the probability of rejecting the null hypothesis at the planned end of study given the existing data at the interim stages, along with certain speculation about future data. If this probability equals to or exceeds a pre-specified threshold, a termination of the trial is made in favor of or against the null hypothesis. These procedures are called conditional power methods, also well known as stochastic curtailment methods termed by Lan, Simon and Halperin (1982), who initiated studies on this class.

In their thought-provoking paper, Lan, Simon and Halperin (1982) developed an elegant theory on stochastic curtailment for Brownian motion. The test statistic in a fixed-sample-size trial for testing  $H_0: \theta = 0$  versus  $H_1: \theta = \theta_1 (\theta_1 > 0)$  with significance level  $\alpha$  and power  $1 - \beta$  is expressed as  $S(t) = B(t) + \theta t, 0 \leq t \leq 1$ , where  $B(t)$  is a standard Brownian motion process. It immediately follows that the conditional distribution of  $S(1)$  given  $S(t)$  is normal with mean  $S(t) + \theta(1-t)$  and variance  $1-t$ , and the conditional probability of rejecting  $H_0$  upon completion of the trial given  $S(t)$ , expressed as  $P_\theta\{S(1) > z_\alpha | S(t)\}$ , is equal to  $\Phi\left(\frac{S(t) + \theta(1-t) - z_\alpha}{\sqrt{1-t}}\right)$ , where  $\Phi$  is the c.d.f of the standard normal and  $\Phi(z_\alpha) = 1 - \alpha$ . A stopping rule based on the above probability is then defined as: accept  $H_0$  if  $P_{\theta_1} < 1 - \gamma$ , which results in a type II error rate bounded by  $\beta / \gamma$ . A similar rule in favor of  $H_1$  is obtained as well but less utilized due to its practical demands.

The simplicity as well as flexibility of this method is very attractive to practitioners, but it also raises a problem on how to choose the value of the tested parameter  $\theta$  under which the conditional power is calculated. As Pepe and Anderson (1992) has argued, it is debatable for Lan, Simon and Halperin (1982) to compute the conditional power only under the alternative hypothesis  $\theta = \theta_1$ , which is often chosen in discretion. Jennison and Turnbull (1990) suggested replacing  $\theta_1$  with  $\hat{\theta}$  which is equal to  $S_n / n$ , the sample mean of the interim data. Pepe and Anderson (1992) considered a similar approach but used a more pessimistic value of  $(S_n + \sqrt{n}) / n$  for  $\theta$ . Finally, Betensky (1997) modified  $(S_n + \sqrt{n}) / n$  to  $(S_n + a\sqrt{n}) / n$  to obtain a less conservative result. In these three procedures, the conditional probability is independent of the alternative hypothesis. Although many procedures based on conditional power have been developed for interim analysis, possibly for the purpose of convenience, only trials of one or two arms are considered.

In this article, we will discuss the conditional power under the framework of linear regression models so that it can be applied to most actual clinical trials in which multiple treatment effects and covariate effects are included. It is well known that the standard power of a regular test for a treatment contrast depends on unknown parameters only through the contrast itself. However it is not true in general for conditional power. Conditions for this to be true for conditional power are established here and some instances are illustrated. We also show that similar arguments can be made about the sufficient statistics for the conditional power.

## **2. Models without covariate effects**

### **2.1 Model description and the full analysis plan**

Let us start with a commonly used clinical trial discussed by many researchers as listed in the previous section. It consists of two arms, one arm for treatment A and another arm for treatment B. We are interested in comparison of main effects of treatment A and

treatment B. The responses obtained from subjects are supposed to follow the model described below:

$$Y_{ij} = \mu_i + e_{ij} \quad (1)$$

where  $i = A, B$ , index of treatment;  $j = 1, \dots, n_i$ , index of subject;  $Y_{ij}$  is the response of the  $j^{th}$  subject assigned to the  $i^{th}$  treatment;  $\mu_i$  is the main effect of the  $i^{th}$  treatment;  $e_{ij}$  are i.i.d random variables from a  $N(0, \sigma^2)$ ,  $\sigma$  being known.

At the scheduled end of the study, a z-test is conducted for testing  $H_0: \mu_A - \mu_B \leq d_0$  versus  $H_1: \mu_A - \mu_B > d_0$ . Let  $\bar{Y}_A$  and  $\bar{Y}_B$  be the mean responses based on treatment A and B respectively. Then, we reject the null hypothesis if  $\bar{Y}_A - \bar{Y}_B > c_v$ ; otherwise, we accept it. Here  $c_v$  is the critical value and will be discussed next.

Let us denote by  $N_A$  and  $N_B$ , the respective sample sizes needed for treatment A and B in order to ensure the prescribed significance level  $\alpha$  and power  $1 - \gamma$  (at pre-specified  $d_1$ ). Then we have

$$c_v = \frac{d_1 z_\alpha + d_0 z_\gamma}{z_\alpha + z_\gamma} \quad (2)$$

$$\frac{1}{N_A} + \frac{1}{N_B} = \frac{(d_1 - d_0)^2}{\sigma^2 (z_\alpha + z_\gamma)^2}$$

$$\text{where } z_\alpha = \Phi^{-1}(1 - \alpha) \text{ and } z_\gamma = \Phi^{-1}(1 - \gamma). \quad (3)$$

### Remark 2.1.1

- Both type I and type II errors are controlled for any  $n_A, n_B$  as long as

$$\frac{1}{n_A} + \frac{1}{n_B} \leq \frac{1}{N_A} + \frac{1}{N_B}.$$

- If  $\frac{1}{n_A} + \frac{1}{n_B} \neq \frac{1}{N_A} + \frac{1}{N_B}$ , then  $c_v$  is given by

$$c_v = z_\alpha \sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} + d_0. \quad (4)$$

The actual power will be greater than  $1 - \gamma$  if  $\frac{1}{n_A} + \frac{1}{n_B} < \frac{1}{N_A} + \frac{1}{N_B}$ .

- If  $N_A = N_B = N$ , then

$$N = \frac{2}{\frac{(d_1 - d_0)^2}{\sigma^2 (z_\alpha + z_\gamma)^2}} = \frac{2\sigma^2 (z_\alpha + z_\gamma)^2}{(d_1 - d_0)^2} \quad (5)$$

which minimizes  $N_A + N_B$  given  $\frac{1}{N_A} + \frac{1}{N_B}$  as in (3).

**Note:**  $d_0$  is set to 0 in the rest of this paper.

## 2.2 Conditional power

At a certain stage during the trial, the interim analysis is performed. Without loss of generality, we assume that the whole study consists of just two stages, stage 1 and stage 2.

### Notations:

Let  $n_{i1}, n_{i2}, n_i$  be the number of observations under treatment  $i$  at stage 1, stage 2 and the full study respectively.

Let  $\bar{Y}_{i1}, \bar{Y}_{i2}, \bar{Y}_i$  be the average response for treatment  $i$  at stage 1, stage 2 and the full study respectively.

Let  $\hat{\mu}_{i1}, \hat{\mu}_{i2}, \hat{\mu}_i$  be the estimators for  $\mu_i$  at stage 1, stage 2 and the full study respectively.

Let  $r_A = \frac{n_{A1}}{n_A}, r_B = \frac{n_{B1}}{n_B}$ .

Let  $\delta = \mu_A - \mu_B$ .

Let  $D_1$  and  $D_2$  be the set of responses at stage 1 and stage 2 respectively.

The conditional power,  $P_c$ , which is the probability of rejection of the null hypothesis at the end of the study conditional on the data  $D_1$  obtained at stage 1 is formulated as

$$P_c = P(\bar{Y}_A - \bar{Y}_B > c_v | D_1). \quad (6)$$

Note that for  $i = A, B$ ,  $\bar{Y}_i = r_i \bar{Y}_{i1} + (1 - r_i) \bar{Y}_{i2}$ . So (6) can be written as

$$\begin{aligned} P_c &= P(r_A \bar{Y}_{A1} - r_B \bar{Y}_{B1} + (1 - r_A) \bar{Y}_{A2} - (1 - r_B) \bar{Y}_{B2} > c_v | D_1) \\ &= P((1 - r_A) \bar{Y}_{A2} - (1 - r_B) \bar{Y}_{B2} > c_v - r_A \bar{Y}_{A1} + r_B \bar{Y}_{B1} | D_1). \end{aligned} \quad (7)$$

Note that  $D_1$  and  $D_2$  are independent. The following theorem is immediate.

**Theorem 2.2.1** Under normality assumption,  $P_c$ , the probability of rejecting the null hypothesis upon completion of the study conditional on the interim data  $D_1$  is given by

$$P_c = \Phi\left(\frac{r_A \bar{Y}_{A1} + (1 - r_A) \mu_A - r_B \bar{Y}_{B1} - (1 - r_B) \mu_B - c_v}{\sigma \sqrt{\frac{1 - r_A}{n_A} + \frac{1 - r_B}{n_B}}}\right) = \Phi(T), \text{ say.} \quad (8)$$

Further, the following corollary follows immediately.

**Corollary 2.2.1** The conditional power,  $P_c$ , depends on  $\mu_A, \mu_B, \bar{Y}_{A1}$  and  $\bar{Y}_{B1}$  only through  $\mu_A - \mu_B$  and  $\bar{Y}_{A1} - \bar{Y}_{B1}$  if and only if  $r_A = r_B$ . And in that case,

$$P_c = \Phi\left(\frac{\delta + (v - \delta)r - c_v}{\sigma \sqrt{\frac{1 - r_A}{n_A} + \frac{1 - r_B}{n_B}}}\right) \quad (9)$$

where  $r = r_A = r_B$ ,  $\delta = \mu_A - \mu_B$  and  $v$  is the value of  $\bar{Y}_{A1} - \bar{Y}_{B1}$ .

### 2.3 Conditional probability on the observed difference

The purpose of the interim analysis is to explore the possibility of early termination of an experiment. This is usually done by examining the difference  $\bar{Y}_{A1} - \bar{Y}_{B1}$  and carrying out a test for  $H_0$  at that stage. We, therefore, take the view that the conditional probability  $P(\bar{Y}_A - \bar{Y}_B > c_v | D_1)$  also needs to be evaluated when  $D_1$  provides information only on  $\bar{Y}_{A1} - \bar{Y}_{B1}$  above, and we denote it by  $P_{od}$ . To simplify the notation, let  $U = \bar{Y}_A - \bar{Y}_B$ ,  $V = \bar{Y}_{A1} - \bar{Y}_{B1}$  and  $W = \bar{Y}_{A1}$ , then the joint distribution of U and V is bivariate normal.

$$\begin{bmatrix} U \\ V \end{bmatrix} \sim N_2\left(\begin{bmatrix} \delta \\ \delta \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & \frac{1}{r^*} \end{bmatrix} \sigma^{*2}\right) \quad (10)$$

where,

$$r^* = \frac{\frac{1}{n_A} + \frac{1}{n_B}}{\frac{1}{n_{A1}} + \frac{1}{n_{B1}}}$$

$$\sigma^{*2} = \left(\frac{1}{n_A} + \frac{1}{n_B}\right) \sigma^2.$$

The conditional distribution of U given V is also normal with

$$\begin{aligned} E(U | V = v) &= \delta + (v - \delta)r^* \\ \text{Var}(U | V = v) &= (1 - r^*)\sigma^{*2}. \end{aligned}$$

So we have

$$\begin{aligned} P_{od} &= P(U > c_v | V = v) \\ &= 1 - \Phi\left(\frac{c_v - \delta - (v - \delta)r^*}{\sqrt{1 - r^*} \sigma^*}\right) \\ &= \Phi\left(\frac{\delta + (v - \delta)r^* - c_v}{\sqrt{1 - r^*} \sigma^*}\right). \end{aligned} \quad (11)$$

Now, we may ask "what is the connection between  $P_c$  and  $P_{od}$ ?" This is answered by the following theorem.

**Theorem 2.3.1** The conditional expectation of  $P_c$  given  $V$  is  $P_{od}$ , namely,

$$P_{od} = E(P_c | V) \quad (12)$$

*Proof.* Since conditioning on  $(Y_{A1}, Y_{B1})$  is equivalent to conditioning on  $(W, V)$ , we have

$$P_c = P(U > c_v | Y_{A1}, Y_{B1}) = P(U > c_v | W, V).$$

Thus,

$$\begin{aligned} E(P_c | V) &= \int P(U > c_v | W, V) f(w | v) dw \\ &= \int_{-\infty}^{\infty} \int_{c_v}^{\infty} f(u | v, w) du f(w | v) dw \\ &= \int_{c_v}^{\infty} \int_{-\infty}^{\infty} f(u | v, w) f(w | v) dw du \\ &= \int_{c_v}^{\infty} \int_{-\infty}^{\infty} f(u, v, w) f(v, w) f(w, v) f(v) dw du \\ &= \int_{c_v}^{\infty} 1 f(v) \int_{-\infty}^{\infty} f(u, v, w) dw du \\ &= \int_{c_v}^{\infty} f(u, v) f(v) du \\ &= \int_{c_v}^{\infty} f(u | v) du \\ &= P(U > c_v | V = v) \\ &= P_{od} \end{aligned} \quad (13)$$

In Corollary 2.2.1, we have observed that  $P_c$  coincides with  $P_{od}$  when  $r_A = r_B$ . Further in Theorem 2.3.1, we have established that  $E(P_c | V) = P_{od}$ . It remains to be seen how close is  $P_c$  to  $P_{od}$  when  $r_A \neq r_B$ . Towards this end, we first examine the properties of  $T$  such that  $P_c = \Phi(T)$ , given  $\bar{Y}_{A1} - \bar{Y}_{B1}$ .

From (8), we can rewrite  $T$  in terms of  $V$  and  $W$  as

$$T = \frac{(r_A - r_B)(W - \mu_A) + r_B(V - \delta) + \delta - c_v}{\sigma \sqrt{\frac{1 - r_A}{n_A} + \frac{1 - r_B}{n_B}}}. \quad (14)$$

Then the following useful observations can be made.

**Fact 2.3.1** Conditional on  $V$ ,  $T$  has normal distribution with mean  $\mu_T$  and variance  $\sigma_T^2$ , where  $\mu_T$  and  $\sigma_T^2$  are given by

$$\mu_T = E(T | V = v) = \frac{(v - \delta)n_{A1}r_B + n_{B1}r_A n_{A1} + n_{B1} + \delta - c_v}{\sigma \sqrt{\frac{1 - r_A}{n_A} + \frac{1 - r_B}{n_B}}} \quad (15)$$

and,

$$\sigma_T^2 = \text{Var}(T | V = v) = (r_A - r_B)^2 n_B (1 - r_A) + n_A (1 - r_B) \cdot n_A n_B n_{A1} + n_{B1} \quad (16)$$

*Proof.* Note that the joint distribution of W and V is bivariate normal.

$$\begin{bmatrix} W \\ V \end{bmatrix} \sim N_2 \left( \begin{bmatrix} \mu_A \\ \delta \end{bmatrix}, \begin{bmatrix} \frac{1}{n_{A1}} & \frac{1}{n_{A1}} \\ \frac{1}{n_{A1}} & \frac{1}{n_{A1}} + \frac{1}{n_{B1}} \end{bmatrix} \sigma^2 \right) \quad (17)$$

with

$$E(W | V = v) = \frac{\mu_A + (v - \delta)n_{B1}}{n_{A1} + n_{B1}} \quad (18)$$

$$\text{Var}(W | V = v) = \frac{\sigma^2}{n_{A1} + n_{B1}} \quad (19)$$

Then, (15) and (16) follow.

**Fact 2.3.2** If  $n_A = n_B = n$ , then (15) and (16) can be simplified as

$$\mu_T = \frac{2r_A r_B (v - \delta)r_A + r_B + \delta - c_v}{\sigma \sqrt{\frac{2 - r_A - r_B}{n}}} \quad (2)$$

and,

$$\sigma_T^2 = (r_A - r_B)^2 / (2 - r_A - r_B)(r_A + r_B) \quad (3)$$

**Fact 2.3.3** If  $n_A = n_B = n$ , the variance of  $P_c$  conditional on  $V$  can be approximated by

$$\text{Var}(P_c | V) \approx \frac{(r_A - r_B)^2}{(2 - r_A - r_B)(r_A + r_B)} e^{\frac{-\mu_T^2}{2\pi}} \quad (4)$$

where  $\mu_T$  is given by (20).

This can be obtained by the standard delta-method. For completeness, the proof is given in the Appendix.

**Theorem 2.3.2** If  $n_A = n_B = n$ , given  $\varepsilon$  and  $\eta (< 1)$ , there exists a choice of  $d_r$  such that if  $|r_A - r_B| < d_r$ , then  $P(|P_c - P_d| > \varepsilon) < \eta$ , and  $d_r$  is given by:

$$d_r = \varepsilon \sqrt{\eta} e^{\frac{\mu_T^2}{2}} \sqrt{2\pi(2 - r_A - r_B)(r_A + r_B)} \quad (5)$$

where  $\mu_T$  is as in (20).



### 3. Model containing a covariate with common coefficient

#### 3.1 Model description and the full analysis plan

The trial we consider here still consists of two arms, one arm for treatment A and another for treatment B; but the response differs from (1) and is expressed below:

$$Y_{ij} = \mu_i + \beta x_{ij} + e_{ij} \quad (6)$$

where  $i = A, B$ , index of treatment;  $j = 1, \dots, n_i$ , index of subjects;  $Y_{ij}$  is the response of the  $j^{\text{th}}$  subject in the  $i^{\text{th}}$  treatment;  $\mu_i$  is the main effect of the  $i^{\text{th}}$  treatment;  $\beta$  is the common coefficient of covariate;  $x_{ij}$  is the covariate value on the  $j^{\text{th}}$  subject in treatment  $i$ ;  $e_{ij}$ 's are i.i.d random variables from a  $N(0, \sigma^2)$ ,  $\sigma$  being known.

As in Section 2.1, a z-test for testing the null hypothesis  $H_0 : \mu_A - \mu_B \leq d_0$  is conducted at the planned end of the study. But the test statistic is  $\hat{\mu}_A - \hat{\mu}_B$  rather than  $\bar{Y}_A - \bar{Y}_B$ .  $\hat{\mu}_A$  and  $\hat{\mu}_B$  are the least square estimators for  $\mu_A$  and  $\mu_B$  respectively and are computed as

$$\begin{aligned} \hat{\mu}_A &= \bar{Y}_A - \hat{\beta} \bar{x}_A \\ \hat{\mu}_B &= \bar{Y}_B - \hat{\beta} \bar{x}_B \end{aligned} \quad (7)$$

$$\hat{\beta} = \frac{\sum_{j=1}^{n_A} (x_{Aj} - \bar{x}_A) Y_{Aj} + \sum_{j=1}^{n_B} (x_{Bj} - \bar{x}_B) Y_{Bj}}{\sum_{j=1}^{n_A} (x_{Aj} - \bar{x}_A)^2 + \sum_{j=1}^{n_B} (x_{Bj} - \bar{x}_B)^2},$$

where  $\hat{\beta}$  is the least square estimator for  $\beta$ ;  $\bar{Y}_A$  and  $\bar{Y}_B$  are the mean responses for treatments A and B respectively;  $\bar{x}_A$  and  $\bar{x}_B$  are the means of covariates for treatments A and B respectively. For convenience in notation, we let  $SS_X$  denote the denominator in (25), namely,

$$SS_X = \sum_{j=1}^{n_A} (x_{Aj} - \bar{x}_A)^2 + \sum_{j=1}^{n_B} (x_{Bj} - \bar{x}_B)^2$$

which is the sum squares of covariates within arms.

#### Remark 3.1.1

1. Recall that,  $\hat{\mu}_A$ ,  $\hat{\mu}_B$  and  $\hat{\beta}$  are the BLUEs for  $\mu_A$ ,  $\mu_B$  and  $\beta$  respectively and their variance expressions are as follows:

$$Var(\hat{\mu}_i) = \sigma^2 \left( \frac{1}{n_i} + \frac{\bar{x}_i^2}{SS_X} \right) \quad i = A, B$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{SS_X}.$$

It is easily seen that the variances of  $\hat{\mu}_A$ ,  $\hat{\mu}_B$  under this model are larger than those under model without covariates, and the two expressions will be closer if the C.V of covaritates goes to  $\infty$ . Further, the variance of  $\hat{\beta}$  will shrink to 0 as  $SS_X$  becomes sufficiently large.

2.  $\hat{\mu}_A$  and  $\hat{\beta}$  as well as  $\hat{\mu}_B$  and  $\hat{\beta}$  are independent, but  $\hat{\mu}_A$  and  $\hat{\mu}_B$  are correlated, and the covariance is given by

$$\text{Cov}(\hat{\mu}_A, \hat{\mu}_B) = \bar{x}_A \bar{x}_B \frac{\sigma^2}{SS_X}.$$

3. The variance of  $\hat{\mu}_A - \hat{\mu}_B$  is given by

$$\text{Var}(\hat{\mu}_A - \hat{\mu}_B) = \sigma^2 \{1/n_A + 1/n_B + (\bar{x}_A - \bar{x}_B)^2 / SS_X\}. \quad (8)$$

Obviously, this variance does not depend on  $\beta$  and is minimized when  $\bar{x}_A = \bar{x}_B$ , and remains the same as the one under the model without the covariate.

With (26), the critical value  $c_v$  and required sample size under this setup can be obtained as

$$c_v = \frac{d_1 z_\alpha + d_0 z_\gamma}{z_\alpha + z_\gamma} \quad (27)$$

$$\frac{1}{N_A} + \frac{1}{N_B} = \frac{(d_1 - d_0)^2}{\sigma^2 (z_\alpha + z_\gamma)^2} - \frac{(\bar{x}_A - \bar{x}_B)^2}{SS_X}. \quad (28)$$

### Remark 3.1.2

1. Both type 1 and type 2 errors are controlled for any  $n_A, n_B$  as long as

$$\frac{1}{n_A} + \frac{1}{n_B} \leq \frac{1}{N_A} + \frac{1}{N_B}.$$

2. When  $1/n_A + 1/n_B \neq 1/N_A + 1/N_B$ , then  $c_v$  is given by

$$c_v = z_\alpha \sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{x}_A - \bar{x}_B)^2}{SS_X}} + d_0. \quad (29)$$

3. Given  $1/n_A + 1/n_B$ ,  $N_A + N_B$  is minimized when  $N_A = N_B = N$ , and  $N$  is given by

$$N = \frac{2}{\frac{(d_1 - d_0)^2}{\sigma^2 (z_\alpha + z_\gamma)^2} - \frac{(\bar{x}_A - \bar{x}_B)^2}{SS_X}}. \quad (30)$$

4. By comparing (4) with (29) and (5) with (30), we see that both  $c_v$  and  $N$  here are greater than or equal to those under model without covariate and further, the two are equal if and only if  $\bar{x}_A = \bar{x}_B$ . Therefore, in design of this trial, it is prudent to adopt a design where the variance of the covariate within treatments remain as large as possible while the difference of the means of the covariate values between treatments remains as small as possible.

5. If  $(\bar{x}_A - \bar{x}_B)^2$  is appreciably large, then no matter what the sample size is, the expected power will be hard to achieve.

### 3.2 Conditional power

As in the previous section, the conditional power is evaluated when  $n_{A1}$  and  $n_{B1}$  patients are enrolled in arms A and B respectively at stage 1.

#### Notations:

Let  $n_{i1}, n_{i2}, n_i$  be the number of observations under treatment  $i$  at stage 1, stage 2 and full study.

Let  $\bar{Y}_{i1}, \bar{Y}_{i2}, \bar{Y}_i$  be the average response for treatment  $i$  at stage 1, stage 2 and full study.

Let  $\bar{x}_{i1}, \bar{x}_{i2}, \bar{x}_i$  be the average of covariate for treatment  $i$  at stage 1, stage 2 and full study.

Let  $\hat{\mu}_{i1}, \hat{\mu}_{i2}, \hat{\mu}_i$  be the estimators for  $\mu_i$  at stage 1, stage 2 and full study.

Let  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}$  be the estimators for  $\beta$  at stage 1, stage 2 and full study.

Let  $D_1, D_2$  be the set  $(Y_{ij})$  in stage 1 and stage 2 respectively.

Let  $J_1, J_2$  be the set of indices of subjects in stage 1 and stage 2 respectively.

Let  $SS_{X1} = \sum_{j=1}^{n_{A1}} (x_{Aj} - \bar{x}_{A1})^2 + \sum_{j=1}^{n_{B1}} (x_{Bj} - \bar{x}_{B1})^2$ .

Let  $SS_{X2} = \sum_{j=1}^{n_{A2}} (x_{Aj} - \bar{x}_{A2})^2 + \sum_{j=1}^{n_{B2}} (x_{Bj} - \bar{x}_{B2})^2$ .

The conditional power,  $P_c$ , is formulated as

$$P_c = P(\hat{\mu}_A - \hat{\mu}_B > c_v | D_1). \quad (31)$$

Note that for  $i = A, B$ ,  $\hat{\mu}_i = \bar{Y}_i - \hat{\beta}\bar{x}_i$  and  $\bar{Y}_i = r_i\bar{Y}_{i1} + (1-r_i)\bar{Y}_{i2}$ . So we have

$$\begin{aligned} P_c &= P(r_A\bar{Y}_{A1} - r_B\bar{Y}_{B1} + (1-r_A)\bar{Y}_{A2} - (1-r_B)\bar{Y}_{B2} - \hat{\beta}(\bar{x}_A - \bar{x}_B) > c_v | D_1) \\ &= P((1-r_A)\bar{Y}_{A2} - (1-r_B)\bar{Y}_{B2} - \hat{\beta}(\bar{x}_A - \bar{x}_B) > c_v - r_A\bar{Y}_{A1} + r_B\bar{Y}_{B1} | D_1). \end{aligned} \quad (32)$$

Before further simplification for (32), let us first introduce the following facts.

**Fact 3.2.1** The conditional expectation and variance of  $\hat{\beta}$  on  $D_1$  are given by

$$\begin{aligned} E(\hat{\beta} | D_1) &= \frac{1}{SS_X} \{ n_{A2}\mu_A(\bar{x}_{A2} - \bar{x}_A) + \beta \sum_{j \in J_2} x_{Aj}(x_{Aj} - \bar{x}_A) + \sum_{j \in J_1} (x_{Aj} - \bar{x}_A)Y_{Aj} \\ &\quad + n_{B2}\mu_B(\bar{x}_{B2} - \bar{x}_B) + \beta \sum_{j \in J_2} x_{Bj}(x_{Bj} - \bar{x}_B) + \sum_{j \in J_1} (x_{Bj} - \bar{x}_B)Y_{Bj} \} \end{aligned} \quad (33)$$

$$Var(\hat{\beta} | D_1) = \frac{SS_{X2} + n_{A2}(\bar{x}_{A2} - \bar{x}_A)^2 + n_{B2}(\bar{x}_{B2} - \bar{x}_B)^2}{SS_X^2} \sigma^2.$$

*Proof.* Notice that,

$$\hat{\beta} = \frac{1}{SS_X} \left\{ \sum_{j \in J_1} (x_{Aj} - \bar{x}_A) Y_{Aj} + \sum_{j \in J_1} (x_{Bj} - \bar{x}_B) Y_{Bj} + \sum_{j \in J_2} (x_{Aj} - \bar{x}_A) Y_{Aj} + \sum_{j \in J_2} (x_{Bj} - \bar{x}_B) Y_{Bj} \right\}. \quad (34)$$

and that for  $j \in J_1$ ,  $Y_{Aj}$  and  $Y_{Bj}$  are constants given  $D_1$ ; for  $j \in J_2$ , because of independence,

$$\begin{aligned} E(Y_{Aj}, j \in J_2 | D_1) &= E(Y_{Aj}, j \in J_2) = \mu_A + \beta x_{Aj} \\ E(Y_{Bj}, j \in J_2 | D_1) &= E(Y_{Bj}, j \in J_2) = \mu_B + \beta x_{Bj} \\ Var(Y_{Aj}, j \in J_2 | D_1) &= Var(Y_{Bj}, j \in J_2 | D_1) = \sigma^2. \end{aligned} \quad (35)$$

### Remark 3.2.1

- From (33), we see that  $E(\hat{\beta} | D_1)$  is a function of  $\beta$  and  $D_1$  as well as of  $\mu_A$  and  $\mu_B$ . However,  $E(\hat{\beta} | D_1)$  depends only on  $\mu_A - \mu_B$  if and only if  $n_{A2}(\bar{x}_{A2} - \bar{x}_A) = n_{B2}(\bar{x}_{B2} - \bar{x}_B)$ . Moreover,  $E(\hat{\beta} | D_1)$  is a function of just  $\beta$  and  $D_1$  if and only if  $\bar{x}_{A1} = \bar{x}_{A2} = \bar{x}_A$  and  $\bar{x}_{B1} = \bar{x}_{B2} = \bar{x}_B$ .

- (33) can also be written as

$$\begin{aligned} E(\hat{\beta} | D_1) &= \frac{SS_{X1}}{SS_X} \hat{\beta}_1 + \frac{SS_{X2}}{SS_X} \beta + \frac{n_{A1}(\bar{x}_{A1} - \bar{x}_A) \bar{Y}_{A1}}{SS_X} + \frac{n_{B1}(\bar{x}_{B1} - \bar{x}_B) \bar{Y}_{B1}}{SS_X} \\ &+ n_{A2}(\bar{x}_{A2} - \bar{x}_A)(\mu_A + \beta \bar{x}_{A2}) SS_X + n_{B2}(\bar{x}_{B2} - \bar{x}_B)(\mu_B + \beta \bar{x}_{B2}) SS_X. \end{aligned} \quad (36)$$

So from (36), it follows that:

- $E(\hat{\beta} | D_1) \neq \hat{\beta}_1$ ;
- $E(\hat{\beta} | D_1) \neq \beta$ .

However, as expected,  $E(\hat{\beta}) = \beta$  regardless of the values assumed by the other parameters. Moreover, in case  $\bar{x}_{A1} = \bar{x}_{A2} = \bar{x}_A$ ,  $\bar{x}_{B1} = \bar{x}_{B2} = \bar{x}_B$ , (36) simplifies to

$$E(\hat{\beta} | D_1) = SS_{X1} SS_X \hat{\beta}_1 + SS_{X2} SS_X \beta. \quad (37)$$

- $Var(\hat{\beta} | D_1)$  does not depend on  $D_1$ ,  $\mu_A$ ,  $\mu_B$  and  $\beta$ .

**Fact 3.2.2** The covariance between  $\hat{\beta}$  and  $\bar{Y}_{A2}$ ,  $\hat{\beta}$  and  $\bar{Y}_{B2}$  conditional on  $D_1$  are given by:

$$\begin{aligned} Cov(\hat{\beta}, \bar{Y}_{A2} | D_1) &= (\bar{x}_{A1} - \bar{x}_A) \sigma^2 \\ Cov(\hat{\beta}, \bar{Y}_{B2} | D_1) &= (\bar{x}_{B1} - \bar{x}_B) \sigma^2. \end{aligned}$$

The proof readily follows from equation(35).

**Lemma 3.2.1** Let  $T = (1-r_A)\bar{Y}_{A2} - (1-r_B)\bar{Y}_{B2} - \hat{\beta}(\bar{x}_A - \bar{x}_B)$  so that we rewrite (32) as  $P_c = P(T > c_v - r_A\bar{Y}_{A1} + r_B\bar{Y}_{B1} | D_1)$ . Then we have

$$E(T | D_1) = (1-r_A)\mu_A - (1-r_B)\mu_B + \{(1-r_A)\bar{x}_{A2} - (1-r_B)\bar{x}_{B2}\}\beta - E(\hat{\beta} | D_1)(\bar{x}_A - \bar{x}_B) \quad (38)$$

$$Var(T | D_1) = \sigma^2 \left\{ \frac{1-r_A}{n_A} + \frac{1-r_B}{n_B} \right\} + (\bar{x}_A - \bar{x}_B)^2 Var(\hat{\beta} | D_1) + \sigma^2 (\bar{x}_A - \bar{x}_B) \{ (1-r_A)(\bar{x}_{A2} - \bar{x}_A) - (1-r_B)(\bar{x}_{B2} - \bar{x}_B) \} \quad (39)$$

where  $E(\hat{\beta} | D_1)$  and  $Var(\hat{\beta} | D_1)$  are as in (33) and (34) respectively.

Now it is easy to conclude the following.

**Theorem 3.2.1** Under the model (24), with normality assumption,  $P_c$ , the probability of rejecting null hypothesis upon completion of study conditional on the interim data  $D_1$  is given by

$$P_c = \Phi\left(\frac{r_A\bar{Y}_{A1} - r_B\bar{Y}_{B1} + E(T | D_1) - c_v}{\sqrt{Var(T | D_1)}}\right) \quad (40)$$

where  $E(T | D_1)$ ,  $Var(T | D_1)$  are given by (38) and (39) respectively.

**Corollary 3.2.1** The conditional power,  $P_c$ , depends on  $\mu_A$  and  $\mu_B$  through their difference  $\mu_A - \mu_B$  if and only if

$$r_A - r_B = \frac{n_{A2}(\bar{x}_{A2} - \bar{x}_A) + n_{B2}(\bar{x}_{B2} - \bar{x}_B)}{SS_X} (\bar{x}_A - \bar{x}_B). \quad (41)$$

**Remark 3.2.4** If  $\bar{x}_A = \bar{x}_B$ , then condition (41) will reduce to  $r_A = r_B$ .

$P_c$  is a function of  $\mu_A$ ,  $\mu_B$ ,  $\beta$  and  $D_1$ . In practice, we may simplify (40) by making some assumptions or controlling the covariates.

**Case 1** If  $\bar{x}_A = \bar{x}_B$ , (40) is simplified to:

$$P_c = \frac{\Phi(r_A\bar{Y}_{A1} - r_B\bar{Y}_{B1} + (1-r_A)\mu_A - (1-r_B)\mu_B + \{(1-r_A)\bar{x}_{A1} - (1-r_B)\bar{x}_{B1}\}\beta - c_v)}{\sigma\sqrt{1-r_An_A + 1-r_Bn_B}}. \quad (42)$$

**Case 1.1** If  $\bar{x}_A = \bar{x}_B$  and  $r_A = r_B = r$ ,  $P_c$  is given by

$$P_c = \frac{\Phi(r(\bar{Y}_{A1} - \bar{Y}_{B1}) + (1-r)(\mu_A - \mu_B) + (1-r)(\bar{x}_{A1} - \bar{x}_{B1})\beta - c_v)}{\sigma\sqrt{1-rn_A + 1-rn_B}}, \quad (43)$$

which depends on  $D_1$  only through  $\bar{Y}_{A1} - \bar{Y}_{B1}$  and on  $(\mu_A, \mu_B)$  only through  $\mu_A - \mu_B$ .

**Case 1.2** If  $\bar{x}_A = \bar{x}_B$  and  $(1-r_A)\bar{x}_{A2} = (1-r_B)\bar{x}_{B2}$ ,  $P_c$  is given by

$$P_c = \frac{\Phi(r_A \bar{Y}_{A1} - r_B \bar{Y}_{B1} + (1-r_A)\mu_A - (1-r_B)\mu_B - c_v)}{\sigma \sqrt{1-r_A n_A + 1-r_B n_B}}, \quad (44)$$

which is independent of  $\beta$ .

**Case 2** If  $r_A = r_B = r$ ,  $\bar{x}_{A1} = \bar{x}_{A2} = \bar{x}_A$ ,  $\bar{x}_{B1} = \bar{x}_{B2} = \bar{x}_B$ ,  $P_c$  is given by

$$P_c = \Phi\left(\frac{r(\bar{Y}_{A1} - \bar{Y}_{B1}) - \frac{SS_{X1}}{SS_X}(\bar{x}_{A1} - \bar{x}_{B1})\hat{\beta}_1 + (1-r)\delta + \beta(\bar{x}_{A1} - \bar{x}_{B1})(1-r - \frac{SS_{X1}}{SS_X}) - c_v}{\sigma \sqrt{1-r_A n_A + 1-r_B n_B + (\bar{x}_A - \bar{x}_B)^2 \frac{SS_{X2}}{SS_X^2}}}\right). \quad (45)$$

This result can also be obtained by noting that, since  $\bar{X}_{A1} = \bar{x}_A$ ,  $\bar{x}_{B1} = \bar{x}_B$ , we have

$$\begin{aligned} E(\hat{\beta} | D_1) &= \frac{SS_{X1}}{SS_X} \hat{\beta}_1 + \frac{SS_{X2}}{SS_X} \beta \\ \text{Var}(\hat{\beta} | D_1) &= \frac{SS_{X2}}{SS_X^2} \sigma^2 \\ \text{Cov}(\hat{\beta}, \bar{Y}_{A2} - \bar{Y}_{B2} | D_1) &= 0. \end{aligned} \quad (46)$$

**Case 2.1** In (45), with additional constraints  $SS_{X1}r = SS_{X2}1-r = SS_X$ ,  $P_c$  is given by

$$P_c = \Phi\left(\frac{r(\hat{\mu}_{A1} - \hat{\mu}_{B1}) + (1-r)(\mu_A - \mu_B) - c_v}{\sigma \sqrt{(1-r)^2\left(\frac{1}{n_{A2}} + \frac{1}{n_{B2}}\right) + \frac{r(1-r)(\bar{x}_A - \bar{x}_B)^2}{SS_{X1}}}}\right) \quad (47)$$

which depends on  $D_1$  only through  $\hat{\mu}_{A0} - \hat{\mu}_{B0}$  and on  $(\mu_A, \mu_B)$  only through  $\mu_A - \mu_B$ . This is a fairly reasonable assumption in practice.

**Case 2.2** In (45), with additional constraints  $\bar{x}_A = \bar{x}_B$ ,  $P_c$  is given by

$$P_c = \Phi(r(\bar{Y}_{A1} - \bar{Y}_{B1}) + (1-r)(\mu_A - \mu_B) - c_v \sigma (1-r) \sqrt{\frac{1}{n_{A2}} + \frac{1}{n_{B2}}}).$$

which is the same as  $P_c$  under model without covariates' effect.

### 3.3 Extension to multiple treatments

In practical clinical trials, there are often more than two treatments involved, as in dose-response finding trials. Therefore, in this section, we extend our work to these situations. We still use the model (24), but  $k$  treatments are considered. The model is as follows:

$$Y_{ij} = \mu_i + \beta x_{ij} + e_{ij}. \quad (48)$$

All the notations have the same meaning as in section 3.2, except  $i = 1, \dots, k, k > 2$ .

Without loss of generality, we are interested in comparing the main effects of treatment 1 and treatment 2. As in previous section, the final test will be based on the LS-estimators for  $\mu_i$  and  $\beta$  as follows:

$$\hat{\mu}_i = \bar{Y}_i - \hat{\beta}\bar{x}_i \quad (49)$$

$$\hat{\beta} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} (x_{ij} - \bar{x}_i) Y_{ij}}{\sum_{i=1}^k \sum_{j=1}^{n_k} (x_{ij} - \bar{x}_i)^2} \quad (50)$$

Note that the difference between (50) and (25) is that  $\hat{\beta}$  comes from the response in all treatment groups rather than just treatment 1(A) and treatment 2(B), which results in the changes in  $E(\hat{\beta} | D_1)$  and  $Var(\hat{\beta} | D_1)$  as follows:

$$\begin{aligned} E(\hat{\beta} | D_1) &= \frac{1}{SS_X} \sum_{i=1}^k \{n_{i1} \mu_i (\bar{x}_{i1} - \bar{x}_i) + \beta \sum_{j \in J_2} x_{ij} (x_{ij} - \bar{x}_i) \\ &\quad + \sum_{j \in J_1} (x_{ij} - \bar{x}_i) Y_{ij}\} \end{aligned} \quad (51)$$

$$Var(\hat{\beta} | D_1) = \frac{SS_{X2} + \sum_{i=1}^k n_{i1} (\bar{x}_{i1} - \bar{x}_i)^2}{SS_X^2} \sigma^2. \quad (52)$$

With similar arguments as in section 3.2, we have the following theorem.

**Theorem 3.3.1** Under model (48),  $P_c$ , the probability of rejecting null hypothesis  $H_0 : \mu_A - \mu_B \leq 0$  upon completion of study, conditional on the interim data  $D_1$ , is given by

$$P_c = \Phi(r_A \bar{Y}_{A1} - r_B \bar{Y}_{B1} + E(T | D_1) - \frac{c_v}{\sqrt{Var(T | D_1)}}) \quad (9)$$

where  $E(T | D_1)$ ,  $Var(T | D_1)$  are given by

$$\begin{aligned} E(T | D_1) &= (1 - r_A) \mu_A - (1 - r_B) \mu_B + \{(1 - r_A) \bar{x}_{A2} - (1 - r_B) \bar{x}_{B2}\} \beta \\ &\quad - E(\hat{\beta} | D_1) (\bar{x}_A - \bar{x}_B) \end{aligned} \quad (54)$$

$$\begin{aligned} Var(T | D_1) &= \sigma^2 \left\{ \frac{1 - r_A}{n_A} + \frac{1 - r_B}{n_B} \right\} + (\bar{x}_A - \bar{x}_B)^2 Var(\hat{\beta} | D_1) \\ &\quad + \sigma^2 (\bar{x}_A - \bar{x}_B) \{(1 - r_A) (\bar{x}_{A2} - \bar{x}_A) - (1 - r_B) (\bar{x}_{B2} - \bar{x}_B)\}, \end{aligned} \quad (55)$$

in which  $E(\hat{\beta} | D_1)$  and  $Var(\hat{\beta} | D_1)$  are given by (51) and (52) respectively.

It is obvious that  $P_c$  depends on  $\mu_1(\mu_A), \mu_2(\mu_B), \mu_3, \dots, \mu_k, \beta$  and  $D_1$ .

**Corollary 3.3.1** The conditional power,  $P_c$ , depends on  $\mu_A$  and  $\mu_B$  through their difference  $\mu_A - \mu_B$  if and only if

$$r_A - r_B = \frac{n_{A2}(\bar{x}_{A2} - \bar{x}_A) + n_{B2}(\bar{x}_{B2} - \bar{x}_B)}{SS_X} (\bar{x}_A - \bar{x}_B) \quad (56)$$

and,  $P_c$  is independent of  $\mu_3, \dots, \mu_k$  if and only if  $\bar{x}_{i2} = \bar{x}_i, i = 3, \dots, k$  or  $\bar{x}_A = \bar{x}_B$ .

#### 4. Model containing a covariate with possibly different coefficients

##### 4.1 Model description and the full analysis plan

In the previous section, we assumed that all treatment groups share the same coefficient, which may not hold in practice. In this section, we do not make that assumption and the model of interest is described below:

$$Y_{ij} = \mu_i + \beta_i x_{ij} + e_{ij} \quad (57)$$

where all notations have the same meaning as in (24) except that  $\beta_i$  denotes the coefficient of covariate for treatment  $i$ . For simplicity, we first consider a trial consisting of two arms, A and B. At the planned end of the study, the test is based on  $\hat{\mu}_A$  and  $\hat{\mu}_B$  given by

$$\begin{aligned} \hat{\mu}_A &= \bar{Y}_A - \hat{\beta}_A \bar{x}_A \\ \hat{\mu}_B &= \bar{Y}_B - \hat{\beta}_B \bar{x}_B \\ \hat{\beta}_A &= \frac{\sum_{j=1}^{n_A} (x_{Aj} - \bar{x}_A) Y_{Aj}}{\sum_{j=1}^{n_A} (x_{Aj} - \bar{x}_A)^2} \\ \hat{\beta}_B &= \frac{\sum_{j=1}^{n_B} (x_{Bj} - \bar{x}_B) Y_{Bj}}{\sum_{j=1}^{n_B} (x_{Bj} - \bar{x}_B)^2}. \end{aligned}$$

It is readily seen that  $\hat{\mu}_A, \hat{\beta}_A, \hat{\mu}_B, \hat{\beta}_B$  are mutually independent and  $Var(\hat{\mu}_A - \hat{\mu}_B) = \sigma^2 \{1n_A + 1n_B + \bar{x}_A^2 SS_{XA} + \bar{x}_B^2 SS_{XB}\}$ . The critical value and the required sample size are given by

$$c_v = \frac{d_1 z_\alpha + d_0 z_\gamma}{z_\alpha + z_\gamma} \quad (58)$$

$$\frac{1}{n_{A2}} + \frac{1}{n_{B2}} = \frac{(d_1 - d_0)^2}{\sigma^2 (z_\alpha + z_\gamma)^2} - \frac{\bar{x}_A^2}{SS_{XA}} - \frac{\bar{x}_B^2}{SS_{XB}} \quad (59)$$

##### Remark 4.1

1. Both type 1 and type 2 errors are controlled as long as  $1n_A + 1n_B \leq 1N_A + 1N_B$ .



2. If  $1n_A + 1n_B \neq 1N_A + 1N_B$ , then  $c_v$  is given by

$$c_v = z_\alpha \sigma \sqrt{\frac{1}{n_{A2}} + \frac{1}{n_{B2}} + \frac{\bar{x}_A^2}{SS_{XA}} + \frac{\bar{x}_B^2}{SS_{XB}}} + d_0, \quad (60)$$

which is larger than the corresponding  $c_v$  in (4) and (29).

3. Given  $1N_A + 1N_B$ ,  $N_A + N_B$  is minimized when  $N_A = N_B = N$ , and  $N$  is given by

$$N = \frac{2}{\frac{(d_1 - d_0)^2}{\sigma^2(z_\alpha + z_\gamma)^2} - \frac{\bar{x}_A^2}{SS_{XA}} - \frac{\bar{x}_B^2}{SS_{XB}}}, \quad (61)$$

which is also larger than  $N$  in (5) and (30).

## 4.2 Conditional Power

To obtain the conditional power, let us first state the following facts.

$$\begin{aligned} E(\hat{\beta}_A | D_1) &= \frac{1}{SS_{XA}} \{n_{A2}\mu_A(\bar{x}_{A2} - \bar{x}_A) + \beta_A \sum_{j \in J_2} x_{Aj}(x_{Aj} - \bar{x}_A) \\ &\quad + \sum_{j \in J_1} (x_{Aj} - \bar{x}_A)Y_{Aj}\} \end{aligned} \quad (62)$$

$$\begin{aligned} E(\hat{\beta}_B | D_1) &= \frac{1}{SS_{XB}} \{n_{B2}\mu_B(\bar{x}_{B2} - \bar{x}_B) + \beta_B \sum_{j \in J_2} x_{Bj}(x_{Bj} - \bar{x}_B) \\ &\quad + \sum_{j \in J_1} (x_{Bj} - \bar{x}_B)Y_{Bj}\} \end{aligned} \quad (63)$$

$$Var(\hat{\beta}_A | D_1) = \frac{SS_{XA2} + n_{A2}(\bar{x}_{A2} - \bar{x}_A)^2}{SS_{XA}^2} \sigma^2 \quad (64)$$

$$Var(\hat{\beta}_B | D_1) = \frac{SS_{XB2} + n_{B2}(\bar{x}_{B2} - \bar{x}_B)^2}{SS_{XB}^2} \sigma^2 \quad (65)$$

$$Cov(\hat{\beta}_A, Y_{A2} | D_1) = (\bar{x}_{A1} - \bar{x}_A) \sigma^2 \quad (66)$$

$$Cov(\hat{\beta}_B, Y_{B2} | D_1) = (\bar{x}_{B1} - \bar{x}_B) \sigma^2 \quad (67)$$

$$Cov(\hat{\beta}_A, \bar{Y}_{B2} | D_1) = Cov(\hat{\beta}_B, \bar{Y}_{A2} | D_1) = 0. \quad (68)$$

### Remark 4.2.1

- From (62), we see that  $E(\hat{\beta}_A | D_1)$  is not only a function of  $\beta_A$  and  $D_1$  but also of  $\mu_A$ . The same holds for  $E(\hat{\beta}_B | D_1)$  as well.
- Expressions (62) and (63) can also be written as

$$\begin{aligned} E(\hat{\beta}_A | D_1) &= \frac{SS_{XA1}}{SS_{XA}} \hat{\beta}_{A1} + \frac{SS_{XA2}}{SS_{XA}} \beta_A + \frac{n_{A1}(\bar{x}_{A1} - \bar{x}_A)\bar{Y}_{A1}}{SS_{XA}} \\ &\quad + \frac{n_{A2}(\bar{x}_{A2} - \bar{x}_A)(\mu_A + \beta_A \bar{x}_{A2})}{SS_{XA}} \end{aligned} \quad (69)$$

$$E(\hat{\beta}_B | D_1) = \frac{SS_{XB1}}{SS_{XB}} \hat{\beta}_{B1} + \frac{SS_{XB2}}{SS_{XB}} \beta_B + \frac{n_{B1}(\bar{x}_{B1} - \bar{x}_B) \bar{Y}_{B1}}{SS_{XB}} + \frac{n_{B2}(\bar{x}_{B2} - \bar{x}_B)(\mu_B + \beta_B \bar{x}_{B2})}{SS_{XB}}. \quad (70)$$

So from (69) and (70), it follows that:

- (i)  $E(\hat{\beta}_A | D_1) \neq \hat{\beta}_{A1}$ ,  $E(\hat{\beta}_B | D_1) \neq \hat{\beta}_{B1}$ ;
- (ii)  $E(\hat{\beta}_A | D_1) \neq \beta_A$ ,  $E(\hat{\beta}_B | D_1) \neq \beta_B$ .

However, as expected,  $E(\hat{\beta}_A) = \beta_A$ ,  $E(\hat{\beta}_B) = \beta_B$  regardless of the values assumed by the other parameters. Moreover, in case  $\bar{x}_{A1} = \bar{x}_{A2} = \bar{x}_A$ ,  $\bar{x}_{B1} = \bar{x}_{B2} = \bar{x}_B$ , (69) and (70) respectively simplify to

$$E(\hat{\beta}_A | D_1) = \frac{SS_{XA1}}{SS_{XA}} \hat{\beta}_{A1} + \frac{SS_{XA2}}{SS_{XA}} \beta_A \quad (71)$$

$$E(\hat{\beta}_B | D_1) = \frac{SS_{XB1}}{SS_{XB}} \hat{\beta}_{B1} + \frac{SS_{XB2}}{SS_{XB}} \beta_B. \quad (72)$$

**Lemma 4.2.1** Let  $T = (1-r_A)\bar{Y}_{A2} - (1-r_B)\bar{Y}_{B2} - \hat{\beta}_A \bar{x}_A + \hat{\beta}_B \bar{x}_B$  so that we rewrite  $P_c = P(\hat{\mu}_A - \hat{\mu}_B > c_v | D_1)$  as  $P_c = P(T > c_v - r_A \bar{Y}_{A1} + r_B \bar{Y}_{B1} | D_1)$ . Then we have

$$\begin{aligned} E(T | D_1) &= (1-r_A)\mu_A - (1-r_B)\mu_B + (1-r_A)\bar{x}_{A2}\beta_A - (1-r_B)\bar{x}_{B2}\beta_B \\ &\quad - E(\hat{\beta}_A | D_1)\bar{x}_A + E(\hat{\beta}_B | D_1)\bar{x}_B \\ Var(T | D_1) &= \left(\frac{1-r_A}{n_A} + \frac{1-r_B}{n_B}\right)\sigma^2 + \bar{x}_A^2 Var(\hat{\beta}_A | D_1) + \bar{x}_B^2 Var(\hat{\beta}_B | D_1) \\ &\quad + (1-r_A)\bar{x}_A Cov(\hat{\beta}_A, \bar{Y}_{A2}) - (1-r_B)\bar{x}_B Cov(\hat{\beta}_B, \bar{Y}_{B2}). \end{aligned} \quad (73)$$

where  $E(\hat{\beta}_A | D_1)$ ,  $E(\hat{\beta}_B | D_1)$ ,  $Var(\hat{\beta}_A | D_1)$ ,  $Var(\hat{\beta}_B | D_1)$ ,  $Cov(\hat{\beta}_A, \bar{Y}_{A2} | D_1)$ ,  $Cov(\hat{\beta}_B, \bar{Y}_{B2} | D_1)$  are as given in (62) through (68).

The following theorem is immediately obtained.

**Theorem 4.2.1** Under model (57), with normality assumption,  $P_c$ , the probability of rejecting null hypothesis upon completion of the study conditional on the interim data  $D_1$ , is given by

$$P_c = \Phi\left(\frac{r(\bar{Y}_{A1} - \bar{Y}_{B1}) + E(T | D_1) - c_v}{\sqrt{Var(T | D_1)}}\right) \quad (74)$$

where  $E(T | D_1)$ ,  $Var(T | D_1)$  are given by (73) and (74) respectively.

**Corollary 4.2.1** The conditional power,  $P_c$ , depends on  $\mu_A$  and  $\mu_B$  through their difference  $\mu_A - \mu_B$  if and only if

$$r_A - r_B = \frac{\bar{x}_A n_{A2} (\bar{x}_{A2} - \bar{x}_A)}{SS_{XA}} - \frac{\bar{x}_B n_{B2} (\bar{x}_{B2} - \bar{x}_B)}{SS_{XB}}. \quad (76)$$

**Remark 4.2.2** If  $\bar{x}_{A1} = \bar{x}_A$  and  $\bar{x}_{B1} = \bar{x}_B$ , then the condition (76) reduces to  $r_A = r_B$ .

**Corollary 4.2.2** If  $\bar{x}_{A1} = \bar{x}_{A2} = \bar{x}_A$ ,  $\bar{x}_{B1} = \bar{x}_{B2} = \bar{x}_B$ ,  $\frac{SS_{XA1}}{SS_{XA}} = r_A$ ,  $\frac{SS_{XB1}}{SS_{XB}} = r_B$ ,  $P_c$  is given by

$$P_c = \Phi\left(\frac{r_A \hat{\mu}_{A1} - r_B \hat{\mu}_{B1} + (1-r_A)\mu_A - (1-r_B)\mu_B - c_v}{\sigma \sqrt{\frac{1-r_A}{n_A} + \frac{1-r_B}{n_B} + \frac{r_A(1-r_A)\bar{x}_{A1}^2}{SS_{XA1}} + \frac{r_B(1-r_B)\bar{x}_{B1}^2}{SS_{XB1}}}}\right), \quad (77)$$

which depends on  $D_1$  only through  $r_A \hat{\mu}_{A1} - r_B \hat{\mu}_{B1}$ .

**Corollary 4.2.3** In corollary 4.2.2, if  $r_A = r_B = r$ ,  $P_c$  is given by

$$P_c = \Phi\left(\frac{r(\hat{\mu}_{A1} - \hat{\mu}_{B1}) + (1-r)(\mu_A - \mu_B) - c_v}{\sigma \sqrt{(1-r)\left(\frac{1}{n_{A2}} + \frac{1}{n_{B2}} + \frac{r\bar{x}_{A1}^2}{SS_{XA1}} + \frac{r\bar{x}_{B1}^2}{SS_{XB1}}\right)}}\right). \quad (78)$$

### 4.3 Extension to multiple arms

In case the trial consists of more than two arms, the results about  $P_c$  remain the same for all pairwise comparisons, taken separately. This is because the existence of other arms have no effects on expressions (62) through (68).

## 5. Concluding Remarks

In this paper we have studied some aspects of evaluation of conditional power in the context of interim analysis, involving two stages of analysis, under a simple linear regression model. We have extended available results in the framework of linear regression with / without a common regression parameter. Additional studies involving more than one regressor / other design-set are to be found in the Doctoral Dissertation of Li (UIC, Unpublished Thesis, 2007).

## References

1. Armitage, P., Mcpherson, C., and Rowe, B.: Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society, Series A*, **132**, 235-244, 1969.
2. Betensky, R.: Conditional power calculations for early acceptance of  $H_0$  embedded in sequential tests. *Statistics in Medicine*, **16**, 456-477, 1997.

3. DeMets, D. and Ware, J.: Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika*, **67**, 651-660, 1980.
4. Gange, S.J. and DeMets D.L.: Sequential testing in clinical trials with repeated measurements. *Biometrika*, **75**, 311-318, 1996.
5. Gould, A.L. and Pecore V.J: Group sequential methods for clinical trials allowing early acceptance of  $H_0$  and incorporating costs. *Biometrika*, **69**, 75-80, 1982.
6. Gu, M.G. and Ying, Z.: Group sequential methods for survival data using partial likelihood score processes with covariate adjustment. *Statistica Sinica*, **5**, 793-804, 1995.
7. Jennison, C. and Turnbull, B.: Group sequential methods with applications to clinical trials. *Chapman & Hall*, 2000.
8. Lan, K.K.G., Simon, R., and Halperin, M.: Stochastically curtailed tests in long-term clinical trials. *Communications in Statistics - Sequential Analysis*, **1**, 207-219, 1982.
9. Lee, J.W. and DeMets, D.L.: Sequential comparison of changes with repeated measurement data. *Journal of the American Statistical Association*, **86**, 757-762, 1991.
10. Pepe, M. and Anderson, G.: Two-stage experimental designs: early stopping with a negative result. *Applied Statistics*, **41**, 181-190, 1992.
11. Pocock, S.J.: Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191-199, 1977.
12. Siatis, A.A.: Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association*, **77**, 855-861, 1982.

## Appendix

Let  $Z_1$  be a random variable, having the normal distribution of mean  $\mu$  and variance  $\sigma^2$ .  
Let  $Z_2 = \Phi(Z_1)$ .

**Theorem A.1** The expectation of  $Z_2$  is given by:

$$E(Z_2) = E(\Phi(Z_1)) = \frac{\mu}{\sqrt{1+\sigma^2}} \quad (79)$$

*Proof.* Let  $Z$  be a random variable which follows  $N(0,1)$  and is independent of  $Z_1$ . Then we have

$$\begin{aligned} E(\Phi(Z_1)) &= P(Z < Z_1) \\ &= P(Z - Z_1 < 0) \\ &= P\left(Z - Z_1 - \frac{(0-\mu)}{\sqrt{1+\sigma^2}} < \frac{\mu}{\sqrt{1+\sigma^2}}\right) \\ &= \Phi\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right) \end{aligned}$$

**Corollary A.1** If  $Z_3 = \Phi(a + bZ_1)$ , where  $a$  and  $b$  are constants, then the expectation of  $Z_3$  is given by

$$E(Z_3) = \Phi(a + b\mu\sqrt{1 + b^2\sigma^2}) \quad (80)$$

**Theorem A.2** Let  $f_{Z_1}$  be the density of  $Z_1$ , then the density of  $Z_2$  is given by

$$f(z_2) = \frac{f_{Z_1}(\Phi^{-1}(z_2))}{\phi(\Phi^{-1}(z_2))} \quad (81)$$

*Proof.* Taking the first derivative of both sides of equation  $z_2 = \Phi(z_1)$  with respect to  $z_1$ , we have

$$\frac{dz_2}{dz_1} = \phi(z_1),$$

so

$$\frac{dz_1}{dz_2} = \frac{1}{\phi(z_1)}$$

and

$$\begin{aligned} f(z_2) &= f_{Z_1}(z_1) \left| \frac{dz_1}{dz_2} \right| \\ &= \frac{f_{Z_1}(\Phi^{-1}(z_2))}{\phi(\Phi^{-1}(z_2))} \end{aligned}$$

**Theorem A.3** The density  $f(z_2)$  has the following property:

If  $\sigma^2 > 1$ ,  $\lim_{z_2 \rightarrow 0} f(z_2) = \lim_{z_2 \rightarrow 1} f(z_2) = \infty$ , and  $f(z_2)$  is minimized when  $z_2 = \Phi(\mu(1 - \sigma^2))$ .

If  $\sigma^2 < 1$ ,  $\lim_{z_2 \rightarrow 0} f(z_2) = \lim_{z_2 \rightarrow 1} f(z_2) = 0$ , and  $f(z_2)$  is maximized when  $z_2 = \Phi(\mu(1 - \sigma^2))$ .

if  $\sigma^2 = 1$  and  $\mu = 0$ ,  $f(z_2) \equiv 1$ .

if  $\sigma^2 = 1$  and  $\mu > 0$ ,  $\lim_{z_2 \rightarrow 0} f(z_2) = 0$ ,  $\lim_{z_2 \rightarrow 1} f(z_2) = \infty$ , and  $f(z_2)$  keeps increasing.

if  $\sigma^2 = 1$  and  $\mu < 0$ ,  $\lim_{z_2 \rightarrow 0} f(z_2) = \infty$ ,  $\lim_{z_2 \rightarrow 1} f(z_2) = 0$ , and  $f(z_2)$  keeps decreasing.

**Theorem A.4** The variance of  $Z_2$  is given by

$$\begin{aligned} \text{Var}(Z_2) &= \int \Phi^2(z_1) f_{Z_1}(z_1) dz_1 - E(Z_2)^2 \\ &= \int \Phi^2(z_1) f_{Z_1}(z_1) dz_1 - \Phi^2\left(\frac{\mu}{\sqrt{1 + \sigma^2}}\right) \end{aligned} \quad (82)$$

and, by the standard delta-method, it can be approximated by

$$\begin{aligned} \text{Var}(Z_2) &\approx \text{Var}(Z_1) \left( \frac{dz_2}{dz_1} \Big|_{z_1=E(Z_1)} \right)^2 \\ &= \frac{\sigma^2 e^{-\mu^2}}{2\pi}. \end{aligned} \quad (83)$$