Sample size reassessment in a clinical trial to demonstrate a clinical meaningful difference in the relief of back pain

Xin Zhang

promotor :
De heer Abel TILAHUN ESHETE, Mr. Eric LEMAIRE, Prof. dr. Paul JANSSEN, Bart GERRITSE, Mr. Javier CASTAÑEDA

Universiteit Maastricht

Eindverhandeling voorgedragen tot het bekomen van de graad master of Science in Biostatistics

universiteit hasselt

Hasselt University

Center for Statistics

Sample Size Reassessment in a Clinical Trial to Demonstrate a Clinical Meaningful Difference in the Relief of Back Pain

Zhang Xin

Internal Supervisor:

Abel Tilahun Eshete

Signature:……

Internal Supervisor:

Prof. dr. Paul Janssen

Signature:……

External Supervisors :

Javier Castaneda

Signature:……

External Supervisors :

dr. Bart Gerritse

Signature:……

A thesis submitted in the partial fulfillment of the requirements for the degree of Master of Science in Biostatistics

June 2008

Acknowledgments

First of all, I would like to thank Hasselt University for giving me the chance to do this thesis, and Medtronic for providing the dataset.

My sincere thank goes to Abel Tilahun Eshete, Javier Castaneda, dr. Bart Gerritse and Prof dr. Paul Janssen for their guidance, close attention, constructive comments and suggestions as my supervisors.

Finally, the appreciation and support from my fellow students were also immense. I thank you all!

Content

Abstract

The assessment of sample size in clinical trials comparing the means requires a variance estimate of the main efficacy variable. If there is uncertainty on the variance estimate when factors such as the difference between centers/cultures are unknown, the internal pilot study with EM algorithm and the conditional rejection probability (CRP) principles are appropriate to perform a sample size re-assessment (SSR) without unblinding the treatment status. These two methods fully preserve the Type I error rate. In order to compare these tow methods, both methodologies are applied in a randomized trial dataset and by a simulation study the mean squared error (MSE) of the variance estimator is used to quantify the amount by which the variance estimator differs from the true value of the variance and hence decide for the most appropriate method to perform a SSR. Finally, we conclude that the conditional rejection probability principle (CRP) is a better approach for re-calculation of the sample size in the considered clinical trial.

Key words: *sample size reassessment (SSR), internal pilot study, conditional rejection probability (CRP)*

Chapter 1

Introduction

Sample size is a key design input for any randomized clinical trial. Unfortunately, it is often computed in the face of inadequate knowledge about the variance (σ^2). Economic pressures, possibly combined with competition for patients, then encourage trial investigators to make optimistic estimates of this design parameter, a tendency that frequently results in underpowered studies. An underpowered trial is extremely undesirable; for it places human subjects at risk with a low probability of reaching a positive scientific conclusion and can result in abandoning an effective compound [1]. Therefore, in recent years there has been a considerable amount of research on clinical trials where the sample size is re-estimated after the clinical trial is underway, on the basis of updated information about σ^2.

Different methods for sample size re-calculation have been proposed. The internal pilot study design introduced by Stein [2] in 1945 use a pre-specified rule to re-calculate the sample size from the interim variance estimate at a pre-planned interim analysis. The method of Stein fully preserves the type I error rate, but it only uses the first-stage variance estimate for the final decision. More recently, internal pilot study designs which use the complete data for variance estimation in the final test have been considered for the t-test [3-6] and implemented the EM algorithm for carrying it out [7], while having a negligible effect on the type I error rate. All internal pilot study designs have in common that the sample size re-calculation is only based on the blinded or unblinded estimate for the unknown nuisance parameter which is inserted in a pre-specified rule. No pre-specification of a re-calculation rule is needed in the adaptive designs [12], which are based on conditional error functions. These designs allow all kinds of data-dependent modifications at pre-planned interim analyses while ensuring control of the type I error rate. Another alternative method to

adjust the sample size due to misspecified nuisance parameters is given by group-sequential designs with an error spending function and information time monitoring [8]. Since the group sequential testing is based on normal approximations for the test statistics, the type I error rate is approximately controlled for large sample sizes only. Muller and Schafer [10] have introduced the conditional rejection probability (CRP)-principle as a general theoretical instrument for design modifications at any time during the course of the trial. With this principle, all types of design modifications known from adaptive designs can also be implemented in every standard design without inflation of the type I error rate. In contrast to adaptive designs, where pre-specification of a conditional error function is a fundamental design element; the CRP-principle uses the "natural" conditional error function. This "natural" conditional error function is implicit in every design [9]. If the variance is unknown and t-distributed test statistics are used, the CRP depends on the unknown variance [11] as shown in the methodology part.

The data is described in Chapter 2 and the objective of this trial is mentioned in Chapter 3. In this report, the sample size re-calculation mainly depends on the unknown variance. Therefore, the possible methods for sample size reassessment based on the nuisance variance parameter are described in detail in Chapter 4; then a comparison of these methods will be reported. In Chapter 6, we extend the exploration of the methods by a simulation study. Finally, the conclusion and discussion will be made in Chapter 7.

Chapter 2

Data Description

To demonstrate a clinical meaningful difference in the relief of back pain between the control and treatment group, the Medtronic Bakken research Center (BRC) in Maastricht cooperates with medical specialists to perform a clinical trial. The data comes from a prospective, randomized, controlled and multi-center study on a Spinal Stabilization System. Data collection is performed at baseline (pre-operative visit) and 6 month post-surgery patient's back pain score on a visual analogue scale (VAS). In this study, patients with a complex lumbar disc disease indicated for a single level herniectomy are considered. The null-hypothesis and the alternative hypothesis of this study are formulated as follows, respectively.

$$H_0: \Delta_{\text{VAS Treatment}} = \Delta_{\text{VAS Control}}$$

$$H_A: \Delta_{\text{VAS Treatment}} \neq \Delta_{\text{VAS Control}}$$

Where $\Delta_{\text{VAS Treatment}}$ is the average change in VAS score (baseline - 6 months) in the treated patient group and $\Delta_{\text{VAS Control}}$ is the average change in the control group. The null hypothesis will be rejected in favor of the alternative if the average change in VAS score in the treatment group is determined to be different than the average change in VAS score in the control group at a significance level of 0.05.

Based on non-published single-center data, in a conservative way, it is expected that the average reduction between baseline and six months will be 3.5 (SD=2.5) points for the treatment group, and 2.5 (SD=3.0) points for the control group patients. A minimal sample size of 240 analyzable patients is required to demonstrate with 80% power a difference in back pain reduction that is significant at the 95% level, comparing treated and control groups.

Chapter 3

Objective

In the planning of a clinical trial, historical data from one center is used to determine the required sample size. Uncertainty on the variance estimate arises when factors such as the difference between centers/cultures are unknown. Sample size reassessment (SSR) is an increasingly popular strategy for designing and conducting clinical trials. In particular, SSR based on updating the variance estimate is a prudent practice accepted by the regulatory authorities to assure adequate power for a study.

The objective of this report is to evaluate and define the most appropriate methodology for sample size reassessment. Among several issues, the most appropriate time in the study to perform a SSR is expected to be defined, while considering that a blind procedure is preferred from the regulatory standpoint, because it better preserves the study integrity.

Chapter 4

Methodology

In this section, different methods for sample size re-calculation will be discussed. The best methods will be selected to apply into the real data by the discussion based on the literate review.

4.1 Internal Pilot Studies with the EM Algorithm for Sample Size Reassessment

Calculation of the required sample size is a key step in the design of a clinical trial. The sample size is determined by the type I and type II error rates, the minimum relevant clinical difference, and the variance of the primary outcome variable. When planning a clinical trial, the required value of the variability measure for the main efficacy variable is generally unknown. It is common practice to use an estimate of the variance from previous trials for sample size calculation. However, an estimation based on previous trials may not be representative of the trials being designed, e.g. different patients types, different treatment, different circumstances and so on. “Internal pilot studies” were recommended to overcome this problem by using data from the first “few” patients entered in the trial to estimate the variance of the main efficacy variable and thus to recalculate the required sample size. An EM algorithm is implemented to calculate an estimate of the within-group variance without unblinding the treatment status at the interim stage and hence to re-estimate the sample size [7].

To introduce this approach, let us consider clinical trials comparing two groups with a normally distributed outcome variable. Suppose the independent and identically distributed observations in group j, j=1, 2, be obtained from normal distributions with

unknown mean μ_j and common unknown variance σ^2. The pre-study estimate of the true variance σ^2 based on literature reviews, previous experience and so on, is denoted s_0^2. This initial estimate s_0^2 was used to determine the preplanned sample size N_0 required in each group to detect the difference $\Delta=\mu_A-\mu_B$ to be statistically significant at level α with power $1-\beta$. In order to reach the specified power, the sample size for per treatment is at least $2\sigma^2 (t_{2(n-1,1-\alpha/2)}+t_{2(n-1),1-\beta})^2/\Delta^2$, where $t_{v,1-P}$ denotes the value exceeded by a t_v random variable with probability P. Since the factor $(t_{2(n-1,1-\alpha/2)}+t_{2(n-1),1-\beta})^2$ depends on the sample size it is usually replace by $(z_{1-\alpha/2}+z_{1-\beta})^2$ where z_{1-P} denotes the value exceeded by a standard normal random variable with probability P. Hence the planned total sample size per group at the start of the trial could be obtained from

$$N_0=\frac{2s_0^2(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2} \qquad (1)$$

In order to determine the size of the internal pilot study, Sandvik *et al.* [13] proposed a method. We will give a brief description of that method. Let n_{int} denote the size of the sample from which the initial variance estimate s_0^2 is calculated and the probability of including more patients in the internal pilot study than is needed for the entire study ($2N_0$ patients) should not be exceed P. The probability P is determined by the investigators. Using the fact that $(n_{\text{int}}-1)s_0^2/\sigma^2$ has a $\chi^2_{n_{init}-1}$ distribution, from the inequality

$$P\left(\frac{(n_{init}-1)s_0^2}{\sigma^2}>\frac{n_{init}-1}{A}\right)<p \qquad (2)$$

We obtain $A=(n_{\text{int}}-1)/\chi^{2-1}_{n_{\text{int}}-1}(1-P)$. Therefore, the total sample size $2n_0=A2N_0$ of the internal pilot study could be obtained. Then these data from the internal pilot study are used to estimate the within-group variance σ^2 denoted s_1^2. With the EM

algorithm and the initial parameters proposed by Gould and Shih [14], we obtain a reliable estimate without unblinding the treatment group at the interim stage.

A brief description of the EM algorithm is given. Let $\tau_i = 1(0)$ denotes the treatment group membership indicator, if sample member *i* is in group 1 (group 2). And z_i, i=1,…, $2n_0$ denotes the interim observations. Given the independent random variables $\tau_1,\ldots,\ \tau_{2n_0}$ with $P(\tau_i = 1) = \theta$, z_i has a normal distribution with density

$$f(z_i \left| \tau_i, \mu_X, \mu_Y, \sigma^2 \right.) = \frac{1}{\sigma} e^{-1/(2\sigma^2)(\tau_i (z_i - u_X)^2 + (1-\tau_i)(z_i - u_Y)^2} \tag{3}$$

Therefore, the expectation of τ_i given z_i is

$$E(\tau_i \left| z_i \right.) = \frac{1}{1 + (1-\theta)/\theta e^{(\mu_X - \mu_Y)(\mu_X + \mu_Y - 2z_i)/(2\sigma^2)}} \tag{4}$$

and the log likelihood of the interim observations is

$$l = \frac{n}{2} \log \sigma^2 + \frac{\sum_{i=1}^{n} \tau_i (z_i - \mu_X)^2 + (1-\tau_i)(z_i - \mu_Y)^2}{2\sigma^2} \tag{5}$$

The EM algorithm procedure is an iterative algorithm that cycle between the so-called E- and M-steps. The E-step consists of substituting current estimates of μ_X, μ_Y and σ^2 into Equation (4) to obtain temporary values for the expectations of the τ_i. The M-step consists of obtaining maximum likelihood estimate of μ_X, μ_Y and σ^2 in Equation.(5) with provisional expectations. The two steps are repeated until the value of σ^2 stabilizes. The resulting value is the estimate s_1^2 of σ^2. The initial values of μ_X, μ_Y and σ^2 are chosen as recommended by Gould and Shih [14],

$$\hat{\mu}_{X,0} = a - \frac{b}{c},\ \hat{\mu}_{Y,0} = a + \frac{b}{c},\ \hat{\sigma}_0 = b$$

Where c=5.71, a and b denote the intercept and the slope obtained from a simple linear regression fitted to the points $(\Phi^{-1}((i-0.5)/(2n_0)), z_{(i)})$, i=1,…, $2n_0$. Φ^{-1} defines

the inverse of the standard normal distribution function and $z_{(i)}$ are the ordered data at the interim evaluation. When we fit a linear regression with the ordered data at the interim evaluation to obtain the initial value for EM algorithm, it is blinded which means we do not know the patient is from treatment group or control group.

After recalculation of the required sample size N_1 in each group, additional observations are taken to bring the total sample size in each group up to N_1, the smallest integer greater than or equal to

$$\max(n_0, \frac{2s_1^2(t_{2(n_0-1),1-\alpha/2} + t_{2(n_0-1),1-\beta})^2}{\Delta^2}) \quad (6)$$

The hypothesis H_0 is rejected if

$$\left(\frac{N_1}{2s^2}\right)^{1/2} \left|\hat{\mu}_X - \hat{\mu}_Y\right| > t_{2(N_1-1),1-\alpha/2} \quad (7)$$

where s^2 is the usual pooled estimate of the variance , $\hat{\mu}_X, \hat{\mu}_Y$ are calculated using all N_1 observations from each group respectively.

Gould & Shih [18] investigate the effect on Type I error rate using the approach discussed above. The details displayed in the Reference [18]. The investigation finds that there is a negligible effect on the type I error rate. And also the simulation studies in Shi [20], Gould & Shih [18] showed that there was high precision in estimating the true within-group variance, and that study power was attained to desired level while the type I error rate was not materially affected by the interim estimation of and updating the sample size [19].

Additionally, two situations were investigated by Shih and Long [15] which are likely to arise in practice. These are the occurrence of variance heterogeneity and of block effects, for example in multi-center trials. Simulations in this situation have shown that the EM procedure is robust against both [15]. It has been suggested that the EM procedure can be straightforwardly generalized to multi-center trials.

4.2 Conditional Rejection Probability (CRP) Principle for SSR

Another alternative method to adjust the sample size is the conditional rejection probability (CRP) proposed by Muller and Schafer [10] that is a general theoretical instrument for design modifications at any time during the course of the trial. With CRP principle, all types of design modifications such as the sample size modification, to include an interim analysis for early stopping when no formal rule for early stopping was foreseen, to increase or reduce the number of planned interim analysis, or to change test statistic, the outcome measure, …without inflation of the type I error rate.

To show the concept in simple terms, consider the case of a two-armed comparative study of an experimental treatment (E) and a control treatment (C). Let θ denote some measure of advantage of E over C, and suppose that we want to test the null-hypothesis H_0: $\theta = 0$ against H_1: $\theta \neq 0$ ($H_1^+ : \theta > 0$ and $H_1^- : \theta < 0$). Let t denote the number of observations at a given calendar time, or, more generally, an information time parameter such as the number of events observed in a survival study, and let $T(t)$ denote a suitable test statistic computed at time point t. Suppose that a group sequential test has been fixed in the study protocol with analyses at time points t_i, i=1,…,m, and with continuation intervals $[a_i, b_i]$ for $T(t_i)$. At the ith analysis, we plan to reject H_0 in favor of H_1, if $T(t_i) < a_i$ and $T(t_i) > b_i$, respectively.

Suppose that at a time point τ one decides to include an interim look and that τ lies in the interval $t_{l-1} \le \tau < t_l$. At the interim data inspection, first of all the value of the test statistic $T(\tau)$ is determined from the data. Let x denote this value. Next, let R^- and R^+ denote the events that H_0 will be rejected in favor of H_1^- and H_1^+ respectively, at one of the future interim analyses $t_l, t_{l+1},\ldots, t_{m-1}$ or at the final analysis according to the group sequential design of the trial. The event R^+ can be written as the union of

disjoint events $R^+ = \sum_{i=1}^{m} R_I^+$, where $R_i^+ = \{T(t_i) > b_i$ and $T(t_j) \in [a_j, b_j]$ for all j with $l \le j < i\}$. The conditional probabilities of these events given that $T(\tau) = x$ are called conditional rejection probabilities (CRP). They are calculated under the null hypothesis and are then denoted by $\varepsilon_0^+(x) = P_0(R^+ | T(\tau) = x)$ and $\varepsilon_0^-(x) = P_0(R^- | T(\tau) = x)$. The index 0 means $\theta = 0$. Now, after the values of $\varepsilon_0^-(x)$ and $\varepsilon_0^+(x)$ have been determined at the interim look, one is allowed to change the design to any design with conditional type I error rates equal to $\varepsilon_0^-(x)$ and $\varepsilon_0^+(x)$ for testing the null hypothesis H_0 against H_1^- and H_1^+ respectively. The CRPs are the only information, which is carried over into the second part of the study. This process will preserve the intended type I error rate. A proof that the procedure holds the pre-specified Type I error level is showed in the Reference [10].

A SAS/IML program for the computation of the conditional rejection probability can be obtained from the Muller and Schafer [10]. The upper and lower characteristic drift values denoted by $\delta^{\pm}$ can be calculated from the SAS macro. These drift values are used to calculate the additional sample size for each group

$$n = \delta^2 \sigma^2 / \Delta^2 (1 + r^2) / r$$

where δ is a characteristic drift value, Δ is the mean difference between two group, and r is the randomization ratio. σ^2 is the variance estimate. Since this is a nuisance variable, one may use the interim variance estimate of s_1^2 to replace it.

The proposed method allows the researcher to include an interim look at any time during the study in progress, to inspect the data of all patients who have completed their follow-up so far, including data group by treatment, and to change the design taking into account the results of the interim look and any other external or internal information. Design changes may include increase or reduction of the sample size and of the number and the time points of interim analyses, and even changes of the type of

statistical test, of the outcome variable, or of the null-hypothesis. Also the method does not require the necessity to pre-specify a sample size rule. And the pre-specified type I error will be maintained. However, to re-calculate the sample size we need to command an unplanned interim analysis. Before starting the trial, an efficient design for the experiment has to be set up, including the interim analyses based on ethical and economic consideration. This always needs careful consideration.

4.3 Information Based Interim Monitoring

We have known that sample size is a key design input for any randomized clinical trial. At the design phase of a randomized clinical trial the total number of participants needed to achieve a certain level of significance and power depends frequently on nuisance parameters like variance, baseline response rate, or regression coefficients other than the main effect. In practical applications, nuisance parameter values are often unreliable guesses founded on little or no available past history. As a result, if the initial guesses for the nuisance parameters are far from the truth, then the study may be under or over powered to detect the desired treatment difference. Therefore, in recent years there has been a considerable amount of research on more flexible clinical trials where the sample size is re-estimate after the clinical trial is underway, on the basis of updated information about variance and effect size. The 'Information Based Interim Monitoring' in which the sample size is recalculated to achieve desired information can be applied here [16, 17].

To give the description of this approach, we start with the discussion of a single unified nuisance-parameter-free information-based approach for designing and monitoring clinical trials. This single unified nuisance-parameter-free information-based approach is applicable to studies involving dichotomous response, continuous response, time to event response and longitudinal response, in the presence of one or more nuisance parameters, including covariates. To show the

concept in simple terms, consider the case of a two-armed comparative study of an experimental treatment (E) and a control treatment (C). Let δ denote some measure of advantage of E over C, and suppose that we want to test the null-hypothesis

$$H_0: \ \delta = 0$$

at the α level of significance. Suppose a two-sided test is to be conducted and is required to have power equal to $1-\beta$ against the alternative hypothesis

$$H_1: \ \delta \neq 0$$

A typical way to carry out a two-sided fixed-information test of the null hypothesis is to fit the underlying model to all the data available at the end of the study, compute the Wald test statistic

$$T = \frac{\hat{\delta}}{se(\hat{\delta})} \tag{12}$$

and reject H_0 if

$$|T| \geq z_{\alpha/2} \tag{13}$$

where z_u is the *(1 - u)*th quantile of the standard normal distribution and se($\hat{\delta}$) is an estimate of the standard error of $\hat{\delta}$. The term information (or Fisher information) has a strict technical meaning which is related to the variance of $\hat{\delta}$ [16]. Now the true variance of $\hat{\delta}$ is usually not known. For all practical purposes, however, the Fisher information available at the end the study can be well approximated by the inverse of the estimated variance of $\hat{\delta}$, or $[\text{se}(\hat{\delta})]^{-2}$. Throughout this development, this approximation should be used as though it were the actual Fisher information [16].

The amount of Fisher information (*I*), needed in order for the test inequality (13) to achieve a power of $1-\beta$ can be derived by standard statistical methods as

$$I = \left[\frac{z_{\alpha/2} + z_{\beta}}{\delta_a} \right]^2 \tag{14}$$

where I is approximated by

$$I \approx \left[se(\hat{\delta}) \right]^{-2} \tag{15}$$

Thus, in a fixed information study one would gather data until the inverse square of the standard error of the estimate of δ equaled the right hand side of equation (14), and would then perform the hypothesis test.

For the normal response with unknown variance, we continue to enroll subjects into the clinical trial until we have gathered a sufficient number. Let us say n_E on treatment group and n_C on control group, so as to satisfy the fixed information requirement

$$\left[se(\hat{\delta}) \right]^{-2} \equiv \left[\frac{\hat{\sigma}^2}{n_E} + \frac{\hat{\sigma}^2}{n_C} \right]^{-1} \geq \left[\frac{z_{\alpha/2} + z_\beta}{\delta_a} \right]^2 \tag{16}$$

We then perform the hypothesis test using the test statistic in which $\hat{\delta} = u_E - u_C$ as long as (16) is satisfied, the study will have the desired $1-\beta$ power. In the case of re-estimation based on the revised estimates of nuisance parameters like the variance of the response variable, the traditional group sequential methodology is utilized without any modification, with Fisher information playing the role of sample size. The maximum information required to provide $1-\beta$ power for a group sequential level-α test to detect a difference δ_a is determined at the design stage by the formula

$$I_{\max} = \left[\frac{z_\alpha + z_\beta}{\delta_a} \right]^2 * IF \tag{17}$$

where IF is the inflation factor whose value depends on α, β, the number of interim looks, and a pre-specified α-spending function. Observe that the computation of I_{max} by Equation (17) does not involve any unknown nuisance parameters whereas the computation of N_{max} requires specification of the nuisance parameter σ^2.

$$N_{\max} = 4\sigma^2 \left[\frac{z_{\alpha/2} + z_{\beta}}{\delta_a} \right]^2 * IF \qquad (18)$$

The study is monitored at administratively convenient times with information at the *jth* interim look being estimated by

$$I_j = [se(\hat{\delta}_j)]^{-2} \qquad (19)$$

The information fraction at look j is thus $t_j = I_j / I_{max}$, the cumulative amount of type I error that may be spent by look j is given by $\alpha(t_j)$, and the corresponding stopping boundary is derived by inverting this cumulative error as discussed in Reference [16]. In a maximum information trial the maximum sample size, N_{max}, need not be fixed in advance. The study remains open with a floating sample size until either I_{max}, the primary determinant of statistical power, is attained or a stopping boundary is crossed [17]. And the final total sample size will be the $\max(n_0, N_{man})$, where n_0 is the planned sample size for our trial.

Such a strategy is well suited for use in conjunction with a group sequential approach where the data are routinely monitored anyway [17]. We have the ability to estimate the nuisance parameters during the interim analysis and, if it seems that the original design will not meet the goals of the study, we may extend the study, increase the sample size. Although in this approach one can control the level of the test, the subsequent power of the test may be greatly affected if the values of the nuisance parameters are guessed incorrectly. It is also reasonable to adopt information based monitoring for sample size adjustment without any intention of early stopping [16]. However, we have to specify the interim boundaries which is the α in the each stage and therefore the interim analysis need to be commanded in the protocol.

4.4 Discussion of the Methods

In the planning of a clinical trial, historical data from one center is used to determine

the required sample size. Uncertainty on the variance estimate arises when factors such as the difference between centers/cultures are unknown. The SSR methods we considered here both have pros and cons. Therefore, the discussion needs to be done.

First of all, we start with the discussion of the "Information Based Interim Monitoring" approach. As we discussed before that subsequent power of the test may be greatly affected if the values of the nuisance parameters are guessed incorrectly. And we do not only intent to increase the total sample size, but also want to know if the sample size is big enough for our trial, then we may decrease the sample size. However the "Information Based Interim Monitoring" always increase the total sample size, it is not suitable in this case.

Secondly, based on the discussion of the internal pilot study with EM algorithm, this approach suggested that it could be straightforwardly generalized to multi-center trials. The simulations in the situation, which include variance heterogeneity and of block effects such as multi-center trials have shown that the EM procedure is robust against both situations, see Reference [15]. Furthermore the re-estimated final sample size can be not only increased but also decreased. Therefore, this approach is suitable in our case. However, for this approach there is a negligible effect on the Type I error.

Thirdly, another alternative method to adjust the sample size is the conditional rejection probability (CRP). The proposed method allows the researcher to include an interim look at any time during the study in progress. Design changes include increase or reduction of the sample size. At the same time, the method does not require the necessity to pre-specify a sample size rule. And the pre-specified type I error will be maintained [10]. The procedure for the sample size adjustment based on interim estimates of nuisance parameter like variance is proposed before. Therefore, to re-estimate the sample size we recommend performing an unplanned interim look. The interim analysis will be arranged not for the purpose of considering early stopping but to recalculate the same size. This always needs careful considerations,

which is the interim analysis for a decision to change the design requires high data quality, i.e. complete and valid data. However, this method still can be considered in our case.

Since we do not always want to increase the sample size, we also want to use the reduction of the sample size due to the economic consideration. Therefore based on the discussion before, we could conclude that the internal pilot study with EM algorithm and CRP principle can be applied in the real data.

4.5 Missingness

Missing or incomplete data are a common scenario occurring in many studies. An observation is considered as incomplete case if the value of any of the variables is missing. Even with the best design and monitoring, the observations can be incomplete usually due to the following possible reasons: missing by design; censoring and drop-out; or non-response etc. Most statistical packages exclude incomplete cases from analysis by default. This approach is easy to implement but has serious problems. Firstly, the loss of any information on incomplete cases may lower the desired efficiency in the study. Secondly; they may lead to substantial biases in analyses. Thus, missing data are important to consider in the analyses.

Missingness frequently complicates the analysis of longitudinal data. In many clinical trials and other setting, the standard methodology used to analyze incomplete longitudinal data is based on such methods as complete case analysis (CC), simple form of imputation (unconditional or conditional mean imputation). This is often done without questioning the possible influence of these assumptions on the final results [21].

In this subsection, we intend to review briefly some simple methods used in practice

in handling missingness problems, though some are invalid statistically. Validity of many of these methods revolves round MCAR assumptions or more strict assumptions. A popular solution for dealing with incomplete longitudinal data is the use of likelihood-based methods due to their validity under the assumption of missing at random (MAR). Therefore, multiple imputations will be introduced in this subsection.

4.5.1. Complete Case Analysis (CC)

In the CC analysis, the analysis is restricting to those subjects with no missing data on variable of interest and assumes MCAR. This method thus uses the entire subjects which have complete observations. This method has clear advantages. It is simple to describe and almost any software computer package can be used to analyze it since there is no missing data. The major disadvantage of this method is that it ignores the possible systematic differences between complete cases and incomplete cases hence leads to substantial loss of information and getting biased results especially when the missingness mechanism is MAR rather than MCAR.

4.5.2. Imputation Techniques

An alternative way to obtain a data set on which the complete data method can be used is to fill in rather than delete [21]. Among those methods used are Unconditional Mean Imputation (UMI) and Buck's method or Conditional mean imputation. The ideal behind a missing value unconditional mean imputation is to replace a missing value with the average of the observed values on the same variable over the other subjects. The term unconditional refers to the fact that one does not use information on the subject for which an imputation is generated. The method is useful for the continuous outcome but problematic in binary outcome.

4.5.3. Multiple Imputation Techniques

Multiple imputation methods proposed by Rubin [21] is a technique to replace missing

values with a set of *M* plausible values, that is, values generated from the distribution of one's data. This is an alternative technique to direct likelihood and Weighted GEE and, at least in its basic form, requires the missing mechanism to be MAR. However, the technique can equally be applied under the MNAR assumption [21]. The multiple imputation technique has three basic phases:

1) The missing values are filled in M times to obtained M complete data sets;
2) The M complete data sets are analyzed by using standard procedure;
3) The results from M analyses are combined to a single inference.

Chapter 5

Result

5.1 Internal Pilot Studies with the EM Algorithm for SSR

As we discussed before, there are two methods suitable in this case. Therefore, we applied the two methods into the dataset. A randomized clinical trial should be planned to compare two groups: one is control group and the other one is treatment group. We suppose that the independent and identically distributed observations be obtained from normal distribution with mean values $\Delta_{\text{VAS Treatments}}$, $\Delta_{\text{VAS Control}}$ and unknown but common standard deviation (S.D.) σ. It was decided that a two-sample *t*-test should be used to compare the two groups with a significant level of α=0.05. The null hypothesis H_0: $\Delta_{\text{VAS TREATMENT}} = \Delta_{\text{VAS Control}}$ should be tested against the two-sided alternative H_A: $\Delta_{\text{VAS TREATMENT}} \neq \Delta_{\text{VAS Control}}$. The sample size should be high enough to detect a clinically relevant difference of at least Δ_{diff}=1 point of the efficacy variable between the two groups, with a power of at least $1-\beta=0.8$. However the value of σ is general unknown. From a previous study with $n_{\text{int}}=52$ patients an estimated S.D. of treatment group and control group are s_1=1.43 and s_2=1.63 respectively. Using this information, it was calculated that 240 patients should be included into the trial, with N_0=120 patients in each group. Then the investigators set *P*=0.1, 0.05, 0.03 and 0.01 to determine the different sample size for the internal pilot study. The *P* is the probability of including more patients in the internal pilot study than is needed for the entire study ($2N_0$ patients). Then from internal pilot study data, the estimate S.D. is obtained at the interim stage and with that s_1 the total sample size is re-estimated. The result is shown in the Table 1.

Table 1: *The Results of SSR. Using Internal Pilot Study with EM*

P	A	Internal Pilot SS($2n_0$)	The Estimate S.D.	The Final SS ($2N_1$)
0.01	0.65903	158	3.2347165	334
0.03	0.71218	170	3.1895157	324
0.05	0.74269	178	3.1858782	324
0.1	0.79321	190	3.1911841	326

From Table 1, we could see that when the sample size of the internal pilot study increased, the estimated S.D. decreased. That is because the more information we obtained the variability will be smaller. Therefore the final total sample size decreased. However since there is 10% of patients with missing information in the dataset for internal pilot study (composed by 190 patients), so in actually there is only 170 patients for the internal pilot study, and the estimated S.D. goes up to 3.2011841. To detect the effect on the type I error rate, the adjust type I error was calculated. The results are displayed in Table 2. It can confirm that this approach has a negligible effect on the type I error and the study power was attained to the desired level.

Table 2: *The Result for Internal Pilot Study with EM and Adjust Alpha*

P	A	Internal Pilot SS($2N_0$)	The Final SS ($2N_1$)	α_{adj}	Power
0.01	0.65903	158	334	0.0484	0.8
0.03	0.71218	170	324	0.0490	0.8
0.05	0.74269	178	324	0.0493	0.8
0.1	0.79321	190	326	0.0496	0.8

5.2 Conditional Rejection Probability (CRP) Principle for SSR

The second method which is the conditional rejection probability principle can be applied to the problem of the sample size recalculation using interim estimates of parameters. In order to re-estimate the sample size, we have to specify an unplanned interim analysis. Suppose we have 158 patients, and then we need to specify an

unplanned interim analysis at time point of 0.658. The boundaries of this design are asymmetric and can be obtained from the SAS/IML program SPEND. The lower nominal alpha levels at information 0.658 and 1 are 0.01 and 0.025, respectively, and the upper nominal alpha level is 0.01 and 0.025, respectively. The lower and upper characteristic draft values which is calculated from the SAS/IML program SPEND for $1-\beta=0.8$ are δ^{-} = -2.846686 and $\delta^{+}=2.8466857$. To adapt the sample size to the interim variance estimate of $s_1^2=10.844671$, one may insert the value $\sigma^2=10.844671$, $\Delta=1.0$ and $\delta=2.8466857$ into the sample size formula $n=\delta^2\sigma^2/\Delta^2(1+r^2)/r$, resulting in $n=80$ patients per group. This means that 160 additional patients will have to be randomized to achieve a conditional power ≥ 0.8 under the design. And the different interim analyses were conducted at the different time points. The result was shown in Table 3.

Table 3: *The Results of SSR. Using CRPs*

SS of Interim Analysis	**Time of Interim Analysis**	**The Estimate S.D.**	**Additional SS**	**The Final SS($2N_1$)**
158	0.658	3.2931249	80*2	318
170	0.708	3.2459700	76*2	322
178	0.742	3.2419075	75*2	328
190	0.792	3.2480500	76*2	342

From Table 3, the estimated S.D. decreased when the sample size for the interim analyses increased resulting in the additional sample size for the design decreased. However, the total sample size is sum of the sample size for the interim analyses and additional sample size for the design, even if the additional sample size decreased, since the sample size for the interim analyses increased, the total sample size still increased for later interim analyses. One thing we have to notice is that there is 10% missingness in the final dataset, therefore the estimated S.D. increase to 3.2480500.

5.3 Missingness

The missing or incomplete data are an important factor in our study. In statistical terminology, missingness in the data is assumed to be three types: 1) Missing completely at Random (MCAR); 2) Missing at random and 3) Missing not at random (MNAR). In order to check the missing data mechanism, the complete case, simple imputation and multiple imputations are applied here.

The data used in this analysis suffer some missingness which is the main objective of this analysis where attention is basically given to handling of missingness. In longitudinal data, missing values are inevitable, so much attention have being given to series of approaches that can be used in analyzing such data. Among several approaches that are found widely in applied statistics are complete case analysis (CC), unconditional mean imputation (UMI). Most standard techniques that are available today and supported by literatures is multiple imputation [21].

From the pattern of the missingness presented in Table 4 we observe that different missing pattern. The most of the profiles are complete (90%), only 0.5% exhibit missingness at the baseline and follow up 6 months. The remaining 9.5% representing the patients have one missing value. Since only one patient misses both two measurements, this patient can be said to be relatively negligible.

Table 4: *Overview of Missingness Pattern and Frequency with Which They Occur*

	Measure Occasion			
Pattern	Baseline	6M	Freq.	Percentage
Complete	O	O	170	90%
Missingness	M	M	1	0.5%
	O	M	16	8%
	M	O	3	1.5%

- 'O' indicates observed, and 'M' indicate missing

The analyses start with simple methods and extend to standard methods of handling missingness such as multiple imputations since CC and UMI based on the assumption of MCAR, while MI is based on MAR assumption. The results of the simple methods and multiple imputation for internal pilot study with EM are shown in Table 5.

In the simple methods, it is observed that estimated S.D. for CC is higher than that of UMI. This may be as a result of less observation where information is reduced by deleting the patients with missing values. The UMI produce a parameter estimate that is pretty close to the MI. This may arise as a result of dealing with complete observations where information is exaggerated by adding unavailable values. Though this method is not totally suitable to handle missingness problem, however, it gives insights into what could be expected when using methods like multiple imputation.

Table 5: *The Results of SSR. Using CC,UMI and MI for Internal Pilot Study with EM*

P	**Internal Pilot SS ($2N_0$)**	**The Final SS Using CC (S.D.)**	**The Final SS Using UMI (S.D.)**	**The Final SS Using MI (S.D.)**
0.1	190	326(3.201184)	316(3.1461626)	316(3.1461626)

The same procedure was used to deal with the missingness in CPRs. The results of the simple methods and multiple imputation of for CRPs are shown in Table 6 blow. The similar result was obtained, in the simple methods; it is observed that estimated S.D. for CC is higher than that of UMI. The estimated S.D. for the UMI is close to the MI.

Table 6: *The Results of SSR. Using CC,UMI and MI for CRPs*

SS of Interim Analysis	**Time of Interim Analysis**	**The Final SS Using CC (S.D.)**	**The Final SS Using UMI (S.D.)**	**The Final SS Using MI (S.D.)**
190	0.792	342(3.2580500)	336(3.2011117)	340(3.23112)

The mean profiles of the VAS score for two groups after using CC, UMI, MI methods were drawn in Figure1. The results obtained from Figure 1 shows that these analyses

are strongly believed to be valid under MCAR for CC and simple imputations and MAR for MI, since the mean for two groups are no difference using CC, UMI and MI when dealing the missingness.

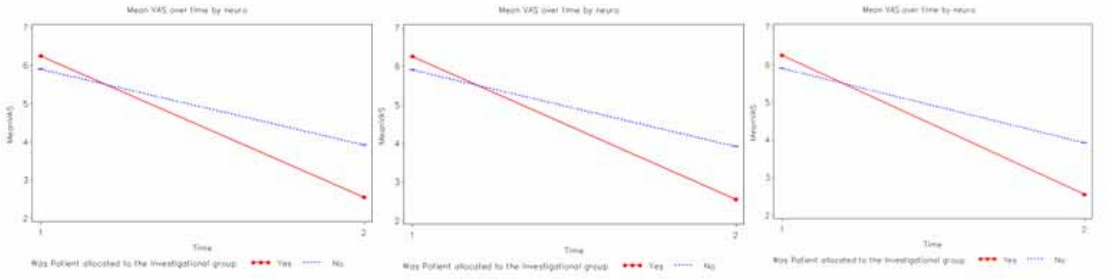

Figure 1*: Mean Profiles :(a) CC (b) UMI (c) MI*

To detect the type of the missingness, model on dropout is fitted with previous observation as a covariate. Table 7 shows that the probability of dropout is a function of previous observation. This implies that the probability of patient dropping out from the study depend on his/her previous observations (observed values). In conclusion, we may say that there is evidence for Missing at Random (MAR) of this dropout pattern.

Table 7: *Parameter Estimates (Standard Errors) of the Dropout Model*

Parameter	Estimate (s.e.)	Pr > ChiSq
Intercept	-2.6559 (0.7108)	0.0002
PREV	-0.0521(0.02508)	0.0422

Chapter 6

Simulation Study

From the result showed in the last section, both internal pilot study with EM and CRP methods are believed to be valid .In this section, the simulations were done in order to select the best method in this particular trial. We simulate a longitudinal data that consists of the baseline and follow up 6 months of the patient's back-pain score on a Visual Analogue Scale (VAS). The score was conducted on the beginning and 6 months after surgical operation simultaneously. Therefore, the difference of the baseline and 6 months of the patient's back-pain score on a VAS is equal to $VAS_{baseline}$-VAS_{FU6M}. And the true variance of this difference will be $Var(V_B - V_F) = Var(V_B) + Var(V_F) - 2Cov(V_B, V_F)$, therefore, given the different values of variance for $VAS_{baseline}$ and $VAS_{FU6M,}$ we will obtain the different values of the variance for that difference VAS. Based on the true variance, the planned the sample can be calucated using the equation (1). 1000 runs were made based on the given variance for $VAS_{baseline}$ and VAS_{FU6M}, then two methods is used to re-estimate the variance. Using that re-estimated variance, the total sample size is re-calculated. Table 8 summarizes the results of the planned sample size and final total sample size respectively and re-estimated variance based on the simulated data using two methods.

Table 8: *The Results of Estimated S.D. for Internal Pilot Study with EM and CRPs*

SS of Interim Analysis	**The Planned SS (True σ)**	**The Final SS Using EM(S.D.)**	**The Final SS Using CRPs(S.D.)**
190	538(4.14108681)	520 (4.0755494)	530(4.12835462)
190	350(3.33889203)	312(3.07732632)	346(3.3106852)
190	226(2.67361179)	210(2.4804382)	274 (2.6453087)
190	100(1.77431677)	190(1.6413256)	230 (1.7643982)

The simulation has shown that the difference between re-estimated variance and true value in CRPs is smaller than the values in internal pilot study with EM. However, we could not conclude the CRPs is better than internal pilot study with EM, since when we re-estimate the variance, there is variability for this parameter estimator. Consequently the mean squared error (MSE) of an estimator, which is one of many ways to quantify the amount by which an estimator differs from the true value of the quantity being estimated, is considered to compare in our case. The results are displayed in Table 9.

Table 9*: The Results of MSE for Internal Pilot Study with EM and CRPs*

SS of Interim Analysis	**The Planned the SS (True σ)**	**The Final SS Using EM(MSE)**	**The Final SS Using CRPs(MSE)**
190	538(4.14108681)	520 (1.1288848)	530 (1.107625)
190	350(3.33889203)	312 (0.7941365)	346 (0.716325)
190	226(2.67361179)	210 (0.4843295)	274 (0.3716128)
190	100(1.77431677)	190 (0.103782)	230 (0.0858109)

From Table 9, we could find the values of MSE for CRPs is smaller than the MSE value in internal pilot study with EM. Hence, we could conclude the conditional rejection principle is better to use for recalculating the sample size in our trial.

Once the unplanned interim analysis is recommended, an important issue is to determine the most appropriate time to perform the interim analysis. A simulation analysis with 1000 runs is made based on a given variance for $VAS_{baseline}$ and VAS_{FU6M}, and with 240 patients in each simulation. Since an unplanned interim analysis is recommended, the study power is also one thing we need to specify here which equals the desired power of 80%. Moreover as we discussed before the type I error rate will not be inflated in this approach. The result for the estimated S.D. based on the interim analysis and the final sample size is shown Table 11.

Table 11: *The Result for Different Interim Analysis Based on the Different Variance*

True σ	Time of Interim Analysis	SS of Interim Analysis	The Estimate S.D.	MSE.	Additional SS	The Final SS($2N_1$)
4.14108681	0.1	24	3.5570415	2.400608	110*2	244
	0.2	48	3.7865983	2.123282	110*2	276
	0.3	72	3.8313465	2.007334	114*2	292
	0.4	96	3.9968759	1.745249	121*2	336
	0.5	120	4.0567875	1.466706	140*2	400
	0.6	144	4.0887694	1.383880	140*2	424
	0.7	168	4.1065841	1.178655	165*2	498
	0.8	192	4.1283546	1.107625	169*2	530
	0.9	216	4.1353987	1.067421	168*2	550
1.77431677	0.1	24	1.6121088	0.335972	22*2	58
	0.2	48	1.6822311	0.196261	26*2	100
	0.3	72	1.7053298	0.158579	26*2	124
	0.4	96	1.7205747	0.133036	26*2	148
	0.5	120	1.7453701	0.117292	24*2	168
	0.6	144	1.7470832	0.107148	20*2	184
	0.7	168	1.7547216	0.092197	23*2	214
	0.8	192	1.7643982	0.085074	20*2	232
	0.9	216	1.7701385	0.076534	21*2	258

From Table 11, we can see, the more patients you take into account in the interim analysis, the estimated S.D. will be closer to the true σ. At the time point 0.5 to including an unplanned interim analysis, the estimated S.D. is closer to the true σ. Moreover the estimated S.D. is closer to the true σ for the later interim analysis times. And MSE is also smaller for the later interim analysis. Therefore, to choose the most appropriate time to include an unplanned interim analysis, there is no strict criterion. However, it is better to consider this unplanned interim analysis after the time point 0.5.

Chapter 7

Conclusion and Discussion

An internal pilot study with EM algorithm and a general principle for statistical design adaptations during the course of an experiment have been presented respectively. The first proposed method used the data from the first 'few' patients entered in the trial to estimate the variance and thus to recalculate the required sample size. And this method is robust the occurrence of variance heterogeneity and of block effects, for example in multi-center trials. However, this approach has a negligible effect on the Type I error rate. The other proposed method is called CRPs, which allows the researcher to include an interim look at any time during the study in progress, to re-estimate the sample size. The pre-specified Type I error level will be maintained. The method can be applied repeated during the course of the trial. It is based upon calculation of the conditional rejection probability (CRP) of the initial study design under the null hypothesis. Both methods can allow the increase or reduction of the sample size.

In longitudinal data, missing values are inevitable, so much attention have being given to series of approaches that can be used in analyzing such data. The data used in this trial suffer some missingness. Therefore, to handle the missingness is important in our trial. The two approaches widely used are complete case analysis (CC) and unconditional mean imputation (UMI). Most standard techniques that are available today and supported by literatures are multiple imputations [21]. However, the results obtained from these analyses are strongly believed to be valid under MCAR for CC and simple imputations and MAR for standard methods such multiple imputations. To detect the type of the missingness, model on dropout is fitted with previous observation as a covariate. From the model, we may say that there is evidence for Missing at Random (MAR) of this dropout pattern.

Finally, in order to compare the two methods for SSR, a simulation analysis was studied here. By comparing the MSE of the variance estimator, we could conclude the conditional rejection probability principle (CRPs) is better to use for recalculating the sample size in our trial. From the simulations, even we cannot give the strict criterion to choose the most appropriate time to include an unplanned interim analysis. We may suggest that it is better to consider this unplanned interim analysis after half of the initial sample size has been reached. Therefore, in our trial, the recalculated sample size would be 318, which means the additional 78 patients should be include into the trials. However, one thing we should notice is that design changes during the course of the trial always need careful considerations, especially when an unplanned interim look is preformed. We recommend that before an interim look is performed the exact procedure should be fixed in the study protocol or in an amendment. If a design modification is made, a protocol amendment should be made immediately including the reasons for changing the design and the complete description of the new design.

Reference

1. Cyrus R. Mehta1 and Nitin R. Patel. Adaptive, group sequential and decision theoretic approaches to sample size determination. *Statistics in Medicine* 2006;**25**: 3250-3269
2. Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* 1945; **16**:243–258.
3. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* 1990; **9**(1–2): 65–71.
4. Birkett MA, Day SJ. Internal pilot studies for estimating sample size. *Statistics in Medicine* 1994; **13**(23–24): 2455–2463.
5. Kieser M, Friede T. Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine* 2000; **19**(7): 901–911.
6. Denne JS, Jennison C. Estimating the sample size for a t-test using an internal pilot. *Statistics in Medicine* 1999;**18**(13):1575–1585.
7. Dietmar Zellner,Gunter E. Zellner, Frieder Keller. A SAS marco for sample size re-estimation. *Computer Methods and Programs in Biomedicine* 2001; **65**:183-190.
8. Mehta CR, Tsiatis AA. Flexible sample size considerations using information-based interim monitoring. *Drug Information Journal* 2001; **35**(4): 1095–1112.
9. Nina Timmesfel,Helmut Schafer and Hans-Helge Muller. Increasing the sample size during clinical trials with t-distributed test statistics without inflating the type I error rate. *Statistics in medicine* 2007; **26:**2449-2464
10. Muller HH, Schafer H. A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine* 2004; **23**(16): 2497–2508. doi: 10.1002/sim.**1852**.
11. Posch M, Timmesfeld N, Konig F, Muller HH. Conditional rejection probabilities of student's t-test and design adaptations. *Biometrical Journal* 2004; **46**(4):

389–403.

12. Jay Bartroff1 and Tze Leung Lai. Efficient adaptive designs with mid-course sample size adjustment in clinical trials. *Statistics in Medicine* 2008; **27:1593**-1611
13. L. Sandvik, J. Erikssen, P. Mowinckel, E.A. Rodland. A method for determining the size of internal pilot studies. *Statistics in Medicine* 1996:**15**:1587-1590
14. A.L.Gould, W.J. Shih. Sample size re-estimation without unbilinding for normally distributed outcomes with unknown variance. *Communications in Statistics Theory* 1992; **21**:2833-2853
15. Shih WJ, Long J. Blinded sample size re-estimation with unequal variances and center effects in clinical trials. *Communications in Statistics* (A) 1998; **27**(2): 395-408.
16. Cyrus R. Mehta and Anastasios A. Tsiatis. Flexible Sample Size Considerations Using Information Based Interim Monitoring. *Drug Inf. J.* 2001; **35**: 1095-1112
17. Cyrus R. Metha and Nitin R. Patel. Adaptive, group sequential and decision theoretic approaches to sample size determination. *Statistics in Medicine* 2006; **25:**3250-3269
18. Gould and Shih. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Commun.Satatist.*1992; **21(10):**2833-2853
19. Long and Shih. Blinded sample size re-estimation with unequal variances and center effects in clinical trials. *Commun.Satatist.*1998; **27(2):**395-408
20. Gould,A.L..Interim analses for monitoring clinical trials that do not affect the type I error rate. *Statistics in Medicin 1992*;**11(10):**55-56
21. Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Appendix

```
data varcov1;/*The D matrix of the random slope and intercept*/
         input m1-m2;
         cards;
        5.7787 4.2046
        4.2046 5.7787
        ;
       /* Store the mean vector in a data set */
       data means1;
         input m1;
         cards;
       0
       0
       ;
run;

%macro generate(means=,varcov=,means1=,varcov1=,nsam=,indi=);/*Macro
to generate sample*/

%do sample=1 %to &nsam;

proc iml worksize=100;
   use &varcov1;          /* read variance-covariance matrix */
   read all into cov;
   use &means1;           /* read means */
   read all into mu;
   v=nrow(cov);          /* calculate number of variables */
   n=240;
   seed = &sample;
   l=t(root(cov));        /* calculate cholesky root of cov matrix */
   z=normal(j(v,n,seed));/* generate nvars*samplesize normals */
   x=l*z;               /* premultiply by cholesky root */
   x=repeat(mu,1,n)+x;    /* add in the means */
   tx=t(x);
   create samplee from tx;  /* write out sample data to sas dataset */
   append from tx;

data samplee;
set samplee;
sample=&sample;
base=col1;
fu6m=col2;
```

```
id=_n_;
drop col1 col2;
run;

          PROC APPEND BASE= summaryw DATA=samplee  FORCE;run;
         proc datasets nolist;
     delete sample;
     quit;
%end;
data verttlevels;
set summaryw;
array yy(2) base fu6m;
do i=1 to 2;
time=i;
error=yy(i);
output;
end;
run;
data verttlevels1;
set verttlevels;
trt=0;
if id>95 then trt=1;
beta0= 10.7969;
beta1=-1.5307;
beta2=-5.2014;
beta3=1.5509;
response=beta0+beta1*trt+(beta2)*time+beta3*trt*time+error;
keep response sample id time trt;
run;
data base;
set verttlevels1;
baseline=response;
where time=1;
drop response;
run;
proc sort data=base;
by sample id;
run;
data foub;
set verttlevels1;
folloup=response;
where time=2;
drop response;
run;
```

```
proc sort data=foub;
by sample id;
run;
data final&indi;
merge base foub;
by sample id ;
diff=baseline-folloup;
drop time;
run;
proc datasets nolist;
delete base foub verttlevels verttlevels1 samplee summaryw;
quit;
%mend;
%generate(means=means,varcov=Ar1rho,means1=means1,varcov1=varcov1,nsa
m=5,indi=2);
/*EM SSR*/
%macro ssem(nsim=,dsname=,simulate=,varname=);
   /*Defining local macro variables*/
   %local maxdiffs       /*Max difference for std between iterations*/
          maxdiffm      /*Max difference for means between iterations*/
          maxiter       /*Max number of iterations*/
             c         /*Constant to ealeulate initial values*/
          theta        /*Probability sample member in treatment groupX*/
         simulate      /*Simulate data or read data froma SAS dataset*/
          dsname       /*Data set name if data were reading from a SAS
dataset*/
          varname      /*variable name to analyse if data were reading
from a SAS dataset*/
           nsim        /*Number of simulations*/
          ngroup       /*Sample size Per group*/
           myx         /*Mean for treatment group X*/
           myy         /*Mean for treatment group Y*/
           std         /*Standard deviation for simulated data*/
          pages        /*Specifies the page size of SAS output*/
         lines;       /*Specifies the line size of SAS output*/
/*Initialisation*/
%let maxdiffs=0.001;      /*Reeommended by Gould and Shih*/
%let maxdiffm=0.001;      /*Reeommended by Gould and Shih*/
%let maxiter=1000;         /*Reeommended  by Gould and Shih*/
%let  c=5.71;             /*Reeommended  by Gould and Shih*/
%let theta=0.5;           /*P(sample member in treatment group X)=theta
*/
%let pages=80;
```

```
%let lines=120;

proc  iml;
  do nsim=1 to  &nsim;
     /*Generating variates for simulation or reading from SAS dataset*/
     %if &simulate=1 %then %do;
                            ngroup=&ngroup;
                             z=j(2*ngroup,1,.)
                             do j=1 to ngroup;
                                z[j,1]=&myx+&std*rannor(-1);
                                end;
                             do j=ngroup+1 to 2*ngroup;
                                z[j,1]=&myy+&std*rannor(-1);
                                end;
                            %end;
                     %else %do;
                                use &dsname;
                                read all var{&varname} into z
where(&varname^=.);
                                ngroup=nrow(z)/2;
                          %end;
 /*Sorting z*/
  help=z;
  z[rank(z),1]=help;
/*Initial valueo for the EM algorithm from linear regression*/
 q=j(2*ngroup,1,.);
 expect=j(2*ngroup,1,.);
 iter=1;  /*First iteration*/
do i=1 to 2*ngroup;
       q[i,1]=probit((i-0.5)/(2*ngroup));
     end;
     qbar=q[:];
     zbar=z[:];
     b=(t(q)*z-2*ngroup*qbar*zbar)/(t(q)*q-2*ngroup*qbar**2);
     a=zbar-b*qbar;

     s_e=b;
     var_e=s_e**2;
     myx_e=a-b/&c;
     myy_e=a+b/&c;
/*Start EM algorithm*/
     do until(iter>&maxiter|(diffs<=&maxdiffs & diffmyx<=&maxdiffm &
diffmyy<=&maxdiffm));
```

```
/*E step*/
do i=1 to 2*ngroup;

expect[i,1]=1/(1+(1-&theta)/&theta*exp((myx_e-myy_e)*(myx_e+myy_e-2*z
[i,1])/(2*var_e)));
        end;
/*M step*/
/*Maximum likelihood estimates*/
myx_m=(t(z)*expect)/sum(expect);
myy_m=(t(z)*(j(2*ngroup,1)-expect))/(2*ngroup-sum(expect));
var_m=(t((z-myx_m*j(2*ngroup,1))##2)*expect+t((z-myy_m*j(2*ngroup,1))
##2)*(j(2*ngroup,1)-expect))/(2*ngroup);
s_m=sqrt(var_m);
/*Initialisation for the next iteration*/
diffs=abs(s_m-s_e);
diffmyx=abs(myx_m-myx_e);
diffmyy=abs(myy_m-myy_e);
s_e=s_m;
var_e=var_m;
myx_e=myx_m;
myy_e=myy_m;
iter=iter+1;
end;    /*End EM algorithm,do until*/
iter=iter-1;  /*Number of iterations*/
/*Creating a SAS file*/
if nsim=1 then create output var{nsim s_m diffs diffmyx diffmyy iter
ngroup};
 append;
end;
quit; /*quit proc iml*/
/*Set print options*/
options pageno=1 nodate ls=&lines ps=&pages;
title4 'VARIANCE ESTIMATION FOR SAMPLE SIZE RE_ESTIMATION';
title7;
%if &simulate=1
   %then %do;
             footnote "SIMULATED DATA:MEAN GROUP X=&myx,
                       MEAN GROUP Y=&myy,STD=&std";
        %end;
   %else %do;
            footnote "SAS DATASET:%upcase(&dsname),
                     VARIABLE:%upcase(&varname)";
        %end;
```

```
/*Printing results*/
proc report data=outPut nowindows headline headskip split='!' spacing=2
ls=&lines  ps=&pages;
  column nsim s_m('_DIFFERENCE OF_' diffs diffmyx diffmyy)iter ngroup;
  define nsim        /order width=10 left'SIMULATION';
  define  s_m        /width=10    'STD DEV';
  define diffs      /width=10 f=8.4  'STD DEV';
  define diffmyx    /width=12 f=8.4  'MEAN GROUP X';
  define diffmyx    /width=12 f=8.4  'MEAN GROUP Y';
  define iter       /width=10  'NO.OF!ITERATIONS';
  define ngroup     /width=16  "'PATIENTS'!PER GROUP AT!INTERIM
ANALYSIS";
run;

%if &nsim>1
       %then %do;
               /*Statistics of the estimated standard deviation*/
                 proc means data=output noprint;
                   var s_m;
                   output out=stat n=nsd mean=mwsd std=stdsd min=minsd
max=maxsd;
                    run;
                 title7 'SUMMARIES OF THE ESTIMATED STANDARD DEVIATION';
                 proc report data=stat center nowindows headline headskip
split='!' spacing=2 ls=&lines ps=&pages;
                 column nsd mwsd stdsd minsd maxsd;
                 define nsd      /width=13 left'SIMULATIONS';
                 define mwsd     /width=15 'MEAN OF STD DEV';
                 define stdsd    /width=8  'STD';
                 define minsd    /width=8  'MIN';
                 define maxsd    /width=8  'MAX';
              run;
           %end;
     title;
   footnote;
data output1;
set output;
keep S_M;
run;
%mend ssem;
%ssem(nsim=1,dsname=sample10,simulate=2,varname=diff)
PROC IML;
START SPEND(time, spend_low, spend_upp, beta, alpha_low, alpha_upp,
delta_low, delta_upp);
```

```
epsilon=0.00000001;
Faktor=1.05;

m=NCOL(time);
time=(1/time[1])*time;
cv_low=SQRT(time)#PROBIT(spend_low);
cv_upp=SQRT(time)#PROBIT(1-spend_upp);

DO k=2 TO m;
 tscale=time[1,2:k]-time[1,1:k-1];
 size=0;
 cl=SQRT(time[k])*PROBIT(spend_low[k]-spend_low[k-1]);
 cu=cv_low[k];
 DO UNTIL (spend_low[k]-epsilon<size & size<=spend_low[k]);
  cv_low[k]=(cl+cu)/2;
  bound=cv_low[1,1:k]//cv_upp[1,1:k];
  CALL SEQ(prob,bound) TSCALE=tscale EPS=epsilon;
  size=SUM(prob[1,1:k]);
  IF size<spend_low[k] THEN cl=cv_low[k]; ELSE cu=cv_low[k];
 END;
 size=0;
 cl=cv_upp[k];
 cu=SQRT(time[k])*PROBIT(1-(spend_upp[k]-spend_upp[k-1]));
 DO UNTIL (spend_upp[k]-epsilon<size & size<=spend_upp[k]);
  cv_upp[k]=(cl+cu)/2;
  bound=cv_low[1,1:k]//cv_upp[1,1:k];
  CALL SEQ(prob,bound) TSCALE=tscale EPS=epsilon;
  size=1-prob[2,k]-SUM(prob[1,1:k-1]);
  IF size<spend_upp[k] THEN cu=cv_upp[k]; ELSE cl=cv_upp[k];
 END;
END;

bound=cv_low//cv_upp;
size=0;
dl=(PROBIT(1-spend_low[m])+PROBIT(1-beta))/SQRT(time[m]);
du=dl;
fdl=1;
fdu=1;
DO UNTIL (beta-epsilon<size & size<=beta);
 delta_low=(dl+du)/2;
 bound_delta=bound+delta_low*(time//time);
 CALL SEQ(prob,bound_delta) TSCALE=tscale EPS=epsilon;
 size=1-SUM(prob[1,1:m]);
```

```
 IF size<beta
  THEN DO; du=delta_low; fdu=0; END;
  ELSE DO; dl=delta_low; fdl=0; END;
 IF fdl=1 THEN dl=du/Faktor;
 IF fdu=1 THEN du=dl*Faktor;
END;
size=0;
dl=(PROBIT(1-spend_upp[m])+PROBIT(1-beta))/SQRT(time[m]);
du=dl;
fdl=1;
fdu=1;
DO UNTIL (beta-epsilon<size & size<=beta);
 delta_upp=(dl+du)/2;
 bound_delta=bound-delta_upp*(time//time);
 CALL SEQ(prob,bound_delta) TSCALE=tscale EPS=epsilon;
 size=prob[2,m]+SUM(prob[1,1:m-1]);
 IF size<beta
  THEN DO; du=delta_upp; fdu=0; END;
  ELSE DO; dl=delta_upp; fdl=0; END;
 IF fdl=1 THEN dl=du/Faktor;
 IF fdu=1 THEN du=dl*Faktor;
END;
cv_low=(time##(-0.5))#cv_low;
cv_upp=(time##(-0.5))#cv_upp;
alpha_low=PROBNORM(cv_low);
alpha_upp=1-PROBNORM(cv_upp);
delta_low=-SQRT(time[m])*delta_low;
delta_upp=SQRT(time[m])*delta_upp;
FINISH;
time={0.791 1};
spend_low={0.01 0.025};
spend_upp={0.01 0.025};
beta=0.2;

CALL SPEND(time, spend_low, spend_upp, beta, alpha_low, alpha_upp,
delta_low, delta_upp);

PRINT alpha_low alpha_upp delta_low delta_upp;

QUIT;
```