Clinical Trials: What's behind?

Aïcha Martijn
BMI paper: A literature study

Supervisor:
Prof. A van der Vaart

October, 2009

VU University
Faculty of Exact Sciences
Business Mathematics and Informatics
De Boelelaan 1081a
1081 HV Amsterdam

Preface

In the last year of the course Business Mathematics & Informatics each student has to write a paper we called the BMI paper. It is a literature study on a subject of choice.

The main subject of this paper is clinical trials. My gratitude goes to prof. van der Vaart for facilitating me a subject that turned out to be very interesting. The knowledge gained with this study is very valuable.

Clinical trials is a hot topic in the medical research field. This literature study explores the statistics, stopping rules and ethical aspect of clinical trials.

Amsterdam, October 2009

Contents

1 Introduction

1.1 The research

The dictionary records the following explanation for a clinical trial. "A clinical trial is any research project that assigns human subjects prospectively to comparison groups and compares the relationships between a medical intervention and a health outcome" [15]. The NRC Handelsblad (newspaper) published on January 23, 2008 an article on clinical trials that was entitled "New test drug leads to death".

The article records that a number of Dutch patients with pancreatitis died during an experimental treatment with probiotica, which was part of a nationwide study. The adverse effects were not expected [14]. People may ask, how come? Or, why did not the researchers stop the experiment earlier? Could the statisticians not see this in advance? What about ethics? This study gives a glimpse of the different issues behind clinical trials.

1.2 The Scope

The purpose of this document is to give a clear and simple explanation of the issues behind clinical trials. It presents one method proposed by Snapinn [7] for monitoring clinical trials based on the ethical dilemmas. The choice for exposition of only one method has to do with delimiting purposes.

1.3 The Outline

Section 2 gives the general aspect of clinical trials and section 3 presents the ethical dilemmas behind clinical trials. Section 4 presents the statistics and stopping rules behind clinical trials based on one method. The choice for this method has to do with the ethical dilemmas presented in Section 3. Section 5 gives some results based on Snapinn's [7] method. Section 6 discusses the advantages and disadvantages of the chosen method. This report also provides a glossary of some statistical terms.

2 Clinical Trials

This section gives an insight in the general structure of clinical trials.

2.1 The benefit-to-harm aspects

The 1962 amendments to the U.S. Federal Food, Drug and Cosmetic Act require that for a new drug to be approved for marketing, there need to be substantial evidence of both safety and efficacy when the drug is prescribed for its intended indication [11]. These are known as the benefit-to-harm aspect of clinical research. Clinical trials are necessary to achieve that standard. Furberg [11] defines the following main goals of clinical trials:

- To make the patient feel better.
- To reduce the risk of future disease complications.
- To improve survival.
- Economic benefit.

2.2 Phases of Clinical trials

The types of clinical trials conducted in human subjects can be divided into four phases [10]. These phases were originally developed for drug development purposes. Exact parallels are not necessarily applicable for other studies, for example a phase III study can be conducted without going through the first two phases. (See Figure (1)). This literature study focuses on Phase III trials based on two groups for this variant seems to be more common. The phase III trials mostly fall under what is called Randomized Controlled Trial (RCT). Randomization is when the patients are divided in two groups and the treatment for each group is randomly selected. Controlled in RCT, stands for the use of comparable groups, which enables a more reliable estimate of both favorable and unfavorable treatment effects. A Randomized, controlled and double-blinded clinical trial is called the "gold standard" for assessing the treatment effects [11]. A double-blinded trial, where both patients and attending physicians are blinded to the actual treatment given to the individuals, is used to reduce any potential bias to a minimum.

2.3 Meta-analysis

Meta-analysis is the process or technique of synthesizing research results, by using various statistical methods to retrieve, select, and combine results from previous separate but related studies. [16]. Furberg [10] define meta-analysis as a database-oriented publication that uses formal statistical methods to combine outcome results from multiple studies of related interventions. Such an analysis increase the statistical power for evaluating treatment effects but at the other hand leads to pooling data from both sound and poor conducted trials.

2.4 Data Monitoring Committee

The Data Monitoring Committee (DMC) are committees established to review interim data and efficacy outcomes in clinical trials. The findings of these committees are used in deciding

whether a trial should be continued as designed, changed, or terminated [17]. A DMC is put into practice when conducting clinical trials based on the "gold standard".

Phases of Clinical Trials

I	II
The earliest types of studies that are carried out in humans. They are typically done using small numbers of healthy subjects and are to investigate the amount and frequency of a drug.	Carried out in patients, usually to find the best dose of drug and to investigate safety.
III	**IV**
Generally major trials aimed at conclusively demostrating efficacy. They are sometimes called confirmatory trials and, in context of pharmaceuticals, typically are the studies on which registration of a new product will be based.	Studies carried out most of the time after registration of a product. They are often for marketing purposes as well as to gain broader experience with using the new product.

Figure 1: Phases Clinical Trials

3 Ethical dilemmas behind randomized Clinical Trials

This section presents the common ethical dilemmas of clinical trials.

3.1 The ethical dilemma

The basic ethical issue behind clinical trials is to balance the individual ethics and collective ethics. Individual ethics has to do with the interests of patients within the trials, while collective ethics deals with reliable results (added value) for future treatments.
Section 2.1 presented the goals of clinical research, implying that each treatment or research has to have an added value to the medical world. Clinical trials concentrating exclusively on individual ethics will be out of balance for it requires the next patient to receive full information including findings. All this makes randomization difficult. Unbiased and precise comparison of treatment would become impossible, and the development of new treatments would be chaotic and unscientific [9]. Extreme use of individual ethics may also cause a trial to stop early, what Whitehead [8] called underrunning. In this way the collective ethics are compromised and the results will be unreliable. Beside the fact that adverse results cannot be excluded from any treatment, it is equally true that underrunning may enhance the probability for adverse results when recorded in the meta-analysis. Pocock [9] presents some disadvantages for underrunning a trial:

- Lack of credibility - small trials are not convincing.
- Lack of realism - dramatic treatment difference are implausible.
- Imprecision - wide confidence interval for treatment effect.
- Bias - trial liable to stop on a "random high".
- Speed - insufficient time and information to consider overall balance of costs and benefits.
- Pressure - unduly enthusiastic and extrapolated recommendations may follow
- Mistakes - risk of false positive results.

There is also a possibility that the stopping criterion is reached but that the data on patients treated on protocol continues to accumulate, what Whitehead [8] called overrunning. This probably happens in cases that the individual ethics are neglected which is again not in harmony with the goals presented in Section 2.1. Whitehead [8] records that overrunning can be prevented when there is a continuous administration from randomization to the time of assessment. Pocock [9] classifies a randomized, controlled and double-blinded clinical trial that is supervised by a Data Monitoring Committee (DMC), that apply a good balance between individual and collective ethics, as a "good" clinical trial.

4 Statistics behind Clinical trials

Pocock [2] and later O'Brien-Fleming [1] present a group sequential procedure for monitoring clinical trials. Many other authors [3-7] has been analyzing Pocock [2] and O'Brien-Fleming [1] methods and presenting new methods. Between all the authors the most appealing study and solution for monitoring clinical trials is the one proposed by Snapinn [7]. This section presents the statistical approach proposed by Snapinn [7].

4.1 Reason for the choice of Snapinn's method

As already mentioned before, there is a wide variety of methods for monitoring clinical trials. The reasons in which Snapinn [7] outweighs Pocock [2] and O'Brien-Fleming [1] are presented in the following sub-sections.

4.1.1 Sequential analysis

The method presented by Snapinn [7] is based on sequential analysis. Sequential analysis is an analysis where the sample size is not fixed in advance and are usually performed due to ethical reasons. This reduces the expected sample size and thus spare study resources.

4.1.2 Fixed sample design

The Snapinn's [7] method does not require a fixed sample size or pre-specified number of Interim analyses. Most of the time it is inconvenient or impossible to specify accurately the number of an interim analysis [7]. This gives a certain flexibility to the procedure and diminish the possibility of underrunning the trial.

4.1.3 Stopping rules

The stopping rules of Snapinn [7] not only makes it possible to stop when a significant level is reached, but also gives the possibility to stop when a significant difference is not longer expected [12]. The stopping boundaries for the p-values depend on the fraction of the completely evaluated patients. This enforces an ongoing assessment control and diminishes the possibility for overrunning the trial.

4.2 Principles of the stopping rules of Snapinn

The stopping rule proposed by Snapinn [7] is based on the conditional probability of rejecting the null hypothesis. This is based on a research between two treatments. The significance between the two treatments is based on two probability boundaries (lower and upper). (See figure (2)). The lower bound is the probability for early acceptance p_{acc} and upper bound the probability for early rejection p_{rej}. In the next subsection we present how Snapinn [7] formulates this, using the one-tailed method.

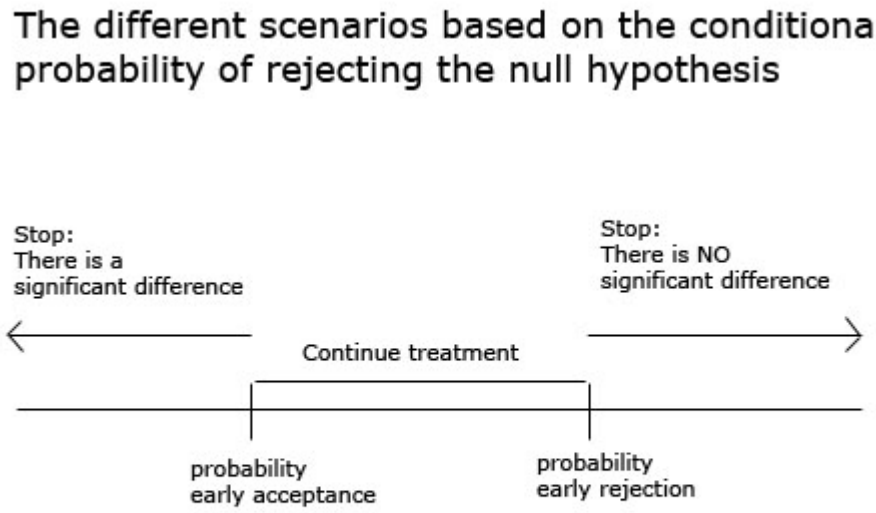

Figure 2: The different scenarios of the stopping rule

4.3 The statistics based on the one-tailed test

Let us compare two (independent) treatments A and B, with respect to a normally distributed random variable, X, assuming that the variance σ^2 is known. The null hypothesis is that the μ of X in group A (μ_A) is equal to the μ of X in group B (μ_B). The one-tailed alternative hypothesis that is considered here is $\mu_A > \mu_B$.

$$H_0: \quad \mu_A = \mu_B \quad \text{or} \quad \mu_A - \mu_B = 0$$
$$H_1: \mu_A > \mu_B$$

Suppose that the planned total sample size of a trial is n subjects, $\frac{n}{2}$ in each group, and that the mean difference between the groups with respect to X is denoted by D [7]. Lets rephrase this all in equations. There is a set of independent identically distributed random variables Z_i that are normally distributed with mean $\mu_A - \mu_B$ and variance $2\sigma^2$.

$$Z_1, \cdots, Z_{\frac{n}{2}} \quad \text{i.i.d} \quad \text{N}(\mu_A - \mu_B, 2\sigma^2)$$

The mean D at the end of the trial is normally distributed with mean zero (under the null hypothesis) and variance $\frac{2\sigma^2}{\frac{n}{2}}$. The mean D can be formulated as:

$$\text{D} = \frac{1}{\frac{n}{2}} \sum_{i=1}^{\frac{n}{2}} Z_i, \quad \text{N}(0, \frac{2\sigma^2}{\frac{n}{2}})$$

Reduce the test of $\mu_A = \mu_B$ to know when to reject H_0. Reject the H_0 if:

$$\frac{X-\mu}{\sigma} = \frac{D-0}{\sqrt{\frac{2\sigma^2}{\frac{n}{2}}}} = \frac{\sqrt{n}D}{2\sigma} > z_{1-\alpha} \tag{1}$$

Accept H_0 if otherwise.

Next, split randomly the total sample into two subsamples with sizes n_1 and n_2, so that $n_1 + n_2 = n$. Assume that the sizes are equal within each subsample and that D_1 and D_2 are the observed mean differences between the two groups in the two subsamples. The situation is changed and needs some readjustment.

$$Z_1, \cdots, Z_{\frac{n_1}{2}}, Z_{\frac{n_1}{2}+1} \cdots, Z_{\frac{n}{2}} \quad \text{i.i.d} \quad \mathrm{N}(\mu_A - \mu_B, 2\sigma^2),$$

with:

$$\mathrm{D}_1 = \frac{1}{\frac{n_1}{2}} \sum_{i=1}^{\frac{n_1}{2}} Z_i, \quad \mathrm{N}(0, \frac{2\sigma^2}{\frac{n_1}{2}})$$

$$\mathrm{D}_2 = \frac{1}{\frac{n_2}{2}} \sum_{i=\frac{n_1}{2}+1}^{\frac{n}{2}} Z_i, \quad \mathrm{N}(0, \frac{2\sigma^2}{\frac{n_2}{2}})$$

Next, express D in D_1 and D_2:

$$\mathrm{D} = \frac{\frac{n_1}{2}\mathrm{D}_1 + \frac{n_2}{2}\mathrm{D}_2}{\frac{n}{2}} \tag{2}$$

The H_0 is now rejected if :
(Replace the D in equation (1) by the D of equation (2))

$$\mathrm{D}_2 > \frac{2\sqrt{n}\sigma z_{1-\alpha} - n_1 D_1}{n_2}$$

Consider an interim analysis performed on the first n_1 subjects, with the remaining n_2 yet to be observed [7]. D_2 is normally distributed with mean $E(D_2)$ and variance $\frac{4\sigma^2}{n_2}$. It is now possible to calculate the probability of eventually rejecting the H_0.

$$P\{\mathrm{D}_2 > \mathrm{X}\} = P\{\text{reject} \quad H_0|D1\} = 1 - \Phi\left(\frac{\mathrm{X} - E(\mathrm{D}_2)}{\sqrt{\frac{2\sigma^2}{\frac{n}{2}}}}\right) = \Phi\left(\frac{E(\mathrm{D}_2) - \mathrm{X}}{\sqrt{\frac{2\sigma^2}{\frac{n}{2}}}}\right),$$

with:

$$\mathrm{X} = \frac{2\sqrt{n}\sigma z_{1-\alpha} - n_1 \mathrm{D}_1}{n_2},$$

yields:

$$Pr\{\text{reject} \quad \mathrm{H}_0|\mathrm{D}_1\} = \Phi\left(\frac{n_1\mathrm{D}_1 + n_2 E(\mathrm{D}_2) - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \tag{3}$$

The trial stop when the conditional probability ($Pr\{\text{reject } \mathrm{H}_0|\mathrm{D}_1\}$) for rejecting the H_0 is smaller then the lower bound p_{acc} or greater then the upper bound p_{rej}. The trial continues if in between. Before continuing for the determination of an adequate p_{acc} and p_{rej}, Snapinn [7] presents two assumptions for $E(D_2)$. One for the purpose of determining the early rejection boundary and the other for the purpose of determining early acceptance boundary.

Assumption 1:
In case of early rejection, the assumption is that the distribution of the future data will be a weighted average of the observed data and zero, with weights equal to the observed samples sizes, respectively [7]. This yields:

$$E(D_2) = fD_1, \tag{4}$$

where f is the fraction of the total sample already observed ($\frac{n_1}{n}$).

Assumption 2:
In the case of early acceptance the assumption is, that the distribution of the future data will be a weighted average of the observed data and δ. δ is the estimated mean difference between the treatment groups upon which the power calculation is based [7]:

$$E(D_2) = fD_1 + (1-f)\delta. \tag{5}$$

Snapinn [7] said that their are other reasonable formulas for $E(D_2)$ but uses (4) and (5) to keep it simple. Notice that at the beginning of the trial ($n_1 = f = 0$) equation (3) equals α if $E(\mathrm{D}_2)$ comes from (4) and is equal to $1-\beta$ if $E(\mathrm{D}_2)$ comes from (5). We also know that $f = \frac{n_1}{n}$ and that $n_2 = n - fn$ since $n_1 + n_2 = n$. Thus, the power calculation can be formulated as follow:

$$z_{1-\beta} = \frac{n_1\mathrm{D}_1 + n_2(f\mathrm{D}_1 + (1-f)\delta) - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}.$$

Applying $n_1 = f = 0$ and $n_2 = n - fn$ to the power equation yields:

$$z_{1-\beta} = \frac{0\cdot\mathrm{D}_1 + (n - 0\cdot n)(0\cdot\mathrm{D}_1 + (1-0)\delta) - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{(n-0\cdot n)}\sigma} = \frac{n\delta}{2\sqrt{n}\sigma} - z_{1-\alpha},$$

$$z_{1-\beta} = \frac{\sqrt{n}\delta}{2\sigma} - z_{1-\alpha}. \tag{6}$$

4.3.1 Determination of p_{acc} and p_{rej}

The two boundaries are derived as follow:

Rejection boundary:
Substitute $E(\mathrm{D}_2)$ for (4) into (3).

$$Pr\{\text{reject} \quad \mathrm{H}_0|\mathrm{D}_1\ E(\mathrm{D}_2 = f\mathrm{D}_1\} = \Phi\left(\frac{n_1\mathrm{D}_1 + n_2 f\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right).$$

Next, apply $n_2 = n - fn$ and $n_1 = fn$:

$$\begin{aligned} Pr\{\text{reject} \quad \mathrm{H}_0|\mathrm{D}_1\ E(\mathrm{D}_2 = f\mathrm{D}_1\} &= \Phi\left(\frac{fn\mathrm{D}_1 + (n-fn)f\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{fn\mathrm{D}_1 + (fn\mathrm{D}_1 - f^2 n\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{n(2f-f^2)\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{(2-f)nf\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{(2-f)n_1\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \end{aligned} \tag{7}$$

If the calculated value for the probability is greater than p_{rej} then stop the trial with the conclusion that μ_A is greater than μ_B. Thus, by setting this probability equal to p_{rej} we obtain the expression for the rejection boundary:

$$\begin{aligned} \text{Rejection} \quad \text{boundary} = z_{p_{rej}} &\geq \frac{(2-f)n_1\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma} \\ z_{p_{rej}} &\geq \frac{(2-f)n_1\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma} \\ 2\sqrt{n_2}\sigma z_{p_{rej}} &\geq (2-f)n_1\mathrm{D}_1 - 2\sqrt{n}\sigma z_{1-\alpha} \\ (2-f)n_1\mathrm{D}_1 &\geq 2\sqrt{n_2}\sigma z_{p_{rej}} + 2\sqrt{n}\sigma z_{1-\alpha} \\ \mathrm{D}_1 &\geq \frac{2\sqrt{n_2}\sigma z_{p_{rej}} + 2\sqrt{n}\sigma z_{1-\alpha}}{(2-f)n_1} \\ \mathrm{D}_1 &\geq \frac{2\sigma(\sqrt{n_2} z_{p_{rej}} + \sqrt{n} z_{1-\alpha})}{(2-f)n_1} \\ \text{Rejection} \quad \text{boundary} &= \frac{2\sigma(\sqrt{n_2} z_{p_{rej}} + \sqrt{n} z_{1-\alpha})}{(2-f)n_1} \end{aligned} \tag{8}$$

According to this formula, the trial stops early with the null hypothesis rejected if the observed value of D_1 after a fraction, f, of the total sample size has been observed is greater then the rejection boundary [7].

Acceptance boundary:
Substitute $E(\mathrm{D}_2)$ for (5) into (3).

$$Pr\{\text{reject} \quad \mathrm{H}_0|\mathrm{D}_1\ E(\mathrm{D}_2 = f\mathrm{D}_1\} = \Phi\left(\frac{n_1\mathrm{D}_1 + n_2(f\mathrm{D}_1 + (1-f)\delta) - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right).$$

Next, apply $n_2 = n - fn$ and $n_1 = fn$:

$$\begin{aligned} Pr\{\text{reject} \quad \mathrm{H}_0|\mathrm{D}_1\ E(\mathrm{D}_2 = f\mathrm{D}_1 + (1-f)\delta\} &= \Phi\left(\frac{fn\mathrm{D}_1 + (n-fn)f\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{fn\mathrm{D}_1 + fn\mathrm{D}_1 - f^2n\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{n(2f-f^2)\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{(2-f)nf\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \\ &= \Phi\left(\frac{(2-f)n_1\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma}\right) \end{aligned} \tag{9}$$

If the calculated value for the probability is less than p_{acc} then stop the trial with the conclusion that μ_A is not greater than μ_B. Thus, by setting this probability equal to p_{acc} we obtain the expression for the acceptance boundary:
(*Apply here also equation (6) for* δ)

$$\begin{aligned} \text{Acceptance} \quad \text{boundary} = z_{p_{acc}} &< \frac{(2-f)n_1\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha}}{2\sqrt{n_2}\sigma} \\ 2\sqrt{n_2}\sigma z_{p_{acc}} &< (2-f)n_1\mathrm{D}_1 + (1-f)n_2\delta - 2\sqrt{n}\sigma z_{1-\alpha} \\ (2-f)n_1\mathrm{D}_1 &< 2\sqrt{n_2}\sigma z_{p_{acc}} - (1-f)n_2\delta + 2\sqrt{n}\sigma z_{1-\alpha} \\ \mathrm{D}_1 &< \frac{2\sqrt{n_2}\sigma z_{p_{acc}} - (1-f)n_2\delta + 2\sqrt{n}\sigma z_{1-\alpha}}{(2-f)n_1} \\ \mathrm{D}_1 &< \frac{2\sigma(\sqrt{n_2}z_{p_{acc}} - \sqrt{n}z_{1-\alpha} + \frac{(1-f)n_2\delta}{2\sigma}}{(2-f)n_1} \\ \text{Acceptance} \quad \text{boundary} &= \frac{2\sigma(\sqrt{n_2}z_{p_{acc}} + \sqrt{n}f(2-f)z_{1-\alpha} - (1-f)^2\sqrt{n}z_{1-\beta})}{n_1(2-f)} \end{aligned} \tag{10}$$

According to this formula, the trial stops early with the null hypothesis rejected if the observed value of D_1 after a fraction, f, of the total sample size has been observed is less then the acceptance boundary [7].

Next, the purpose is to calculate the probabilities of false early rejection and false early acceptance under the null hypothesis for the conditional probability rule as functions of p_{rej} and p_{acc}, and then to find values of p_{rej} and p_{acc} which make these two approximately equal [7]. To calculate the two boundaries is a bit complex.

Snapinn [7] presents two integrals to determine the right values of p_{acc} and p_{rej} based on the value of α and β. He uses the equations (3),(8) and (10) to find the boundaries for an individual interim analysis. The integrals are based on the following information or assumptions:

- At the beginning of the trial is known that a interim analysis will be done after n_1 observations.

- Distribution of D_1 under H_0 is N(0, $\frac{2\sigma^2}{\frac{n}{2}}$).
- The true probability under H_0 of the fixed-sample analysis will eventually reject the H_0, as a function of D_1 (equation (3)).
- The range of values of D_1 leads the conditional probability procedure to reject early (8) and accept early (10).

The probability of early rejection or acceptance at the analysis with n_1 observations can be calculated with the following integrals:

$$\text{Rejection} = \int_{\frac{2\sigma(\sqrt{n_2}z_{p_{rej}}+\sqrt{n}z_{1-\alpha})}{(2-f)n_1}}^{\infty} \left(1 - \Phi\left[\frac{n_1\mu}{2\sigma\sqrt{n_2}} - \frac{z_{1-\alpha}}{\sqrt{1-f}}\right]\right) \frac{e^{\frac{-n_1\mu^2}{8\sigma^2}}}{\sigma\sqrt{(\frac{8\pi}{n_1})}} d\mu.$$

The integral above express the probability of early rejection but not the final rejection under H_0. Similarly, for the probability of early acceptance.

$$\text{Acceptance} = \int_{-\infty}^{\frac{2\sigma(\sqrt{n_2}z_{p_{acc}}+\sqrt{n}f(2-f)z_{1-\alpha}-(1-f)^2\sqrt{n}z_{1-\beta})}{n_1(2-f)}} \left(\Phi\left[\frac{n_1\mu}{2\sigma\sqrt{n_2}} - \frac{z_{1-\alpha}}{\sqrt{1-f}}\right]\right) \frac{e^{\frac{-n_1\mu^2}{8\sigma^2}}}{\sigma\sqrt{(\frac{8\pi}{n_1})}} d\mu.$$

This study will not elaborate on the integrals but will only present the results. The results are needed for the calculation of the critical p-values. The following table gives the appropriate values of p_{acc} corresponding to the different values of α, β and p_{rej}.

Table 1: Appropriate values of p_{acc} for different values of α, β and p_{rej}

		p_{rej}			
α	β	0,80	0,85	0,90	0,95
0,005	0,20	0,459	0,325	0.200	0,089
	0,10	0,489	0,350	0,218	0,098
	0,05	0,517	0,375	0,237	0,108
0,01	0,20	0,430	0,306	0,190	0,086
	0,10	0,462	0,333	0,209	0,095
	0,05	0,493	0,360	0,229	0,106
0,025	0,20	0,394	0.283	0,177	0,081
	0,10	0,429	0.312	0,199	0,092
	0,05	0,465	0.345	0,224	0,106
0,05	0,20	0,367	0.266	0,168	0,078
	0,10	0,408	0.300	0,194	0,092
	0,05	0,453	0.341	0,226	0,111

4.3.2 Calculation of the p-value boundaries

Equations (8) and (10) give the rejection and acceptance boundaries used by the conditional probability procedure for values of D_1. These boundaries can be turned into p-values boundaries since the distribution of D_1 under the null hypothesis is known. According to Snapinn

[7] these p-values, which determine the conditional probability rule's decision, can be compared to the p-values from a standard, fixed sample analysis done after n_1 observations. The p-values are:
(*Calculate these using (8) and (10)*)

$$\begin{aligned}
\text{Rejection boundary} &= 1-\Phi\left(\frac{\frac{2\sigma(\sqrt{n_2}z_{p_{rej}}+\sqrt{n}z_{1-\alpha})}{(2-f)n_1}}{sqrt\frac{2\sigma^2}{\frac{n_1}{2}}}\right)\\
&= 1-\Phi\left(\frac{2\sigma(\sqrt{n_2}z_{p_{rej}}+\sqrt{n}z_{1-\alpha})}{(2-f)n_1}\cdot\frac{\sqrt{n_1}}{2\sigma}\right)\\
&= 1-\Phi\left(\frac{\sqrt{n-fn}z_{p_{rej}}+\sqrt{n}z_{1-\alpha}}{(2-f)fn}\cdot\sqrt{fn}\right)\\
&= 1-\Phi\left(\frac{\sqrt{n}\sqrt{1-f}z_{p_{rej}}+\sqrt{n}z_{1-\alpha}}{(2-f)fn}\cdot\sqrt{f}\sqrt{n}\right)\\
&= 1-\Phi\left(\frac{\sqrt{1-f}z_{p_{rej}}+z_{1-\alpha}}{(2-f)\sqrt{f}}\right) \qquad (11)
\end{aligned}$$

$$\begin{aligned}
\text{Acceptance boundary} &= 1-\Phi\left(\frac{\frac{2\sigma(\sqrt{n_2}z_{p_{acc}}+\sqrt{n}f(2-f)z_{1-\alpha}-(1-f)^2\sqrt{n}z_{1-\beta})}{n_1(2-f)}}{\sqrt{\frac{2\sigma^2}{\frac{n_1}{2}}}}\right)\\
&= 1-\Phi\left(\frac{2\sigma(\sqrt{n_2}z_{p_{acc}}+\sqrt{n}f(2-f)z_{1-\alpha}-(1-f)^2\sqrt{n}z_{1-\beta})}{n_1(2-f)}\cdot\frac{\sqrt{n_1}}{2\sigma}\right)\\
&= 1-\Phi\left(\frac{\sqrt{n}\sqrt{1-f}z_{p_{acc}}+\sqrt{n}f(2-f)z_{1-\alpha}-(1-f)^2\sqrt{n}z_{1-\beta}}{fn(2-f)}\cdot\sqrt{fn}\right)\\
&= 1-\Phi\left(\frac{\sqrt{1-f}z_{p_{acc}}+f(2-f)z_{1-\alpha}-(1-f)^2z_{1-\beta}}{\sqrt{f}(2-f)}\right)\\
&= 1-\Phi(\frac{f(2-f)z_{1-\alpha}+\sqrt{1-f}z_{p_{acc}}-(1-f)^2z_{1-\beta}}{\sqrt{f}(2-f)}) \qquad (12)
\end{aligned}$$

5 Computational Results

This section presents some figures that reflect the change in p-values boundaries based on the fraction. These results are based on the equations (11) and (12).

5.1 The p-values table example

The following table gives an example of the p-values boundaries with $\alpha = 0,025$, $\beta = 0,05$, $p_{rej} = 0,80$ and $p_{acc} = 0,465$.

Table 2: Critical p-value boundaries for early rejection and early acceptance based on $\alpha = 0,025$, $\beta = 0,05$, $p_{rej} = 0,80$ and $p_{acc} = 0,465$

f	Critical value for early Rejection	Acceptance
0,05	< 0,0001	0,9992
0,10	< 0,0001	0,9588
0,15	0,0001	0,8444
0,20	0,0004	0,7015
0,25	0,0011	0,5653
0,30	0,0021	0,4487
0,35	0,0034	0,3538
0,40	0,0049	0,2785
0,45	0,0065	0,2196
0,50	0,0080	0,1737
0,55	0,0094	0,1381
0,60	0,0108	0,1104
0,65	0,0120	0,0889
0,70	0,0130	0,0722
0,75	0,0139	0,0592
0,80	0,0147	0,0489
0,85	0,0155	0,0409
0,90	0,0165	0,0346
0,95	0,0179	0,0296
1,00	0,0250	0,0250

Assume that a treatment research (without any adjustment) is conducted with multiple testing and a point of analysis is reached. If for example 50% of the patients have completed the trial, then a p-value of 0,0080 or less is required to reject the null hypothesis and a p-value of 0,1737 or more is required to accept the null hypothesis. The trial continues otherwise.

5.2 The change in p-values

The following two figures give the change in p-values boundaries based on the fraction. These p-values are produced based on the equations (11) and (12) with $\alpha = 0,025$ and $\beta = 0,05$.

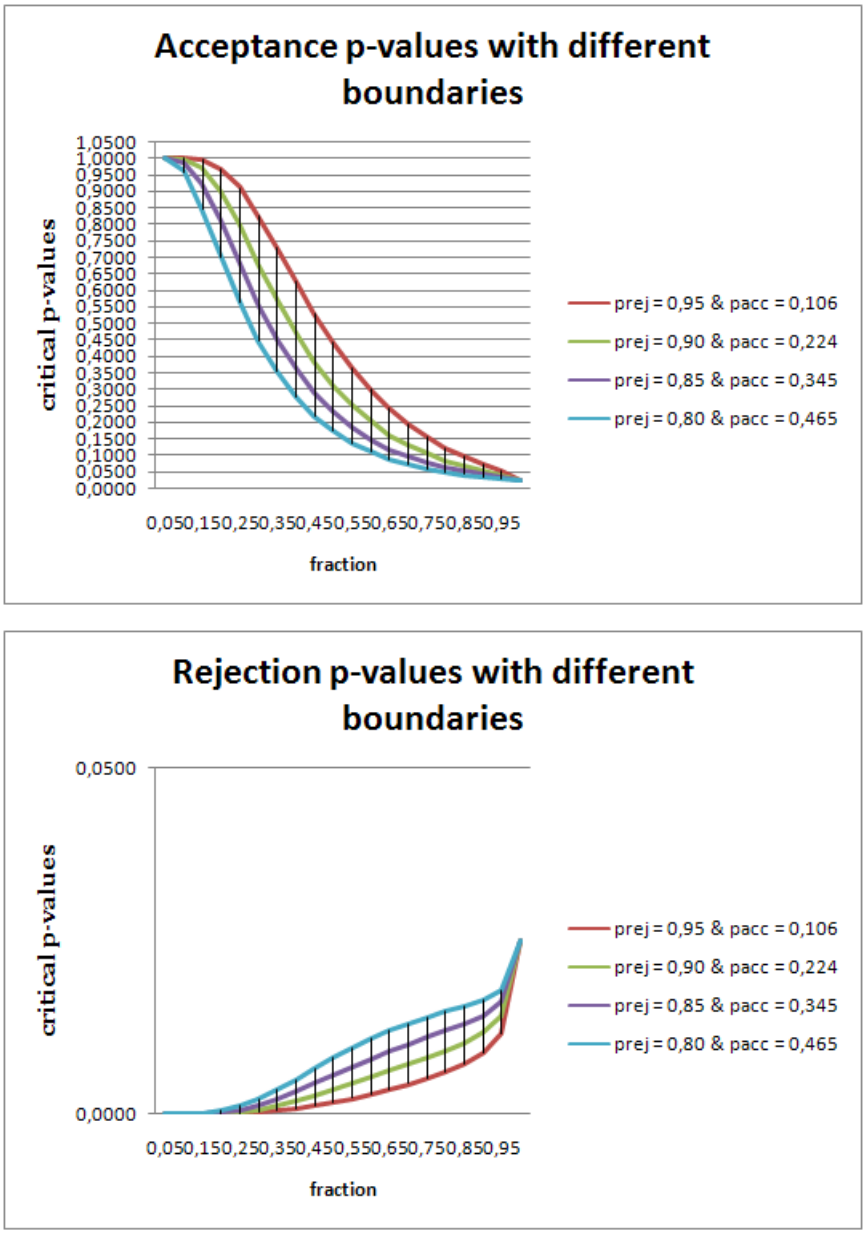

Figure 3: Graphs of p-values with different probability boundaries

It is to be noticed that the β and p_{acc} depends on the choice of p_{rej} and α. The p_{acc} differences in respond to α and β (based on Table 1) are larger when $p_{rej} = 0.80$ instead of $p_{rej} = 0.95$. The figures above reveal that, the smaller the p_{rej} and p_{acc}, the larger the reduction in expected sample size. (See figures 3). The appendix contains some details of the calculation performed.

6 Discussion

According to Snapinn's [7] research, critical values for early rejection given a three stage conditional probability rule with $p_{rej} = 0,90$ resembles the O'Brien-Fleming [1] boundaries. Snapinn [7] also concluded based on a simulation study he has done that his procedure (method) can achieve a nice decrease in expected sample size compared to a fixed sample size design. This is done with very little effect on the significance level or the power of the trial. The conditional probability procedure enables one to stop a trial the moment a significant level is reached. Based on the advantages described in this study one may conclude that Snappin's [7] procedure is an ethical-dilemma proof method.

On the other side, this study brought two disadvantages and a question on the Snapinn [7] procedure. The first disadvantage seem to be the fact that a trial is never stopped in the first stage of the method, for the differences in p-values are too small [12]. Secondly, Schouten [12] also comments that the sensitivity of the statistical analysis in the first part of the experiment is too limited in order to proof a realistic difference. Another downside of this method is based on a comment made by Gill [13]. It concerns the sensitiveness of the model for small binary experiments.

These might be some of the reasons why researchers choose the Pocock [2] or the O'Brien-Fleming [1] procedures. More research must be performed on the Snapinn's [7] method, in order to make a more concrete judgment of its practical application.

7 Glossary

Conditional Probability: the probability that an event will occur under the condition that another event occurs first: equal to the probability that both will occur divided by the probability that the first will occur [20].

One-tail test: When the alternative hypothesis H_1 is one-sided like $\theta > \theta_0$ or $\theta < \theta_0$, then the rejection region is taken only on one side of the sampling distribution. It is called one-tailed test or one-sided test. When H_0 is one-sided to the right like $\theta > \theta_0$, the entire rejection region equal to α is taken in the right end of the sampling distribution. The test is called one-sided to the right. The hypothesis H_0 is rejected if the calculated value of a statistic, say Z falls in the rejection region. The critical value is Z_α which has the area equal to α to its right [19].

p-value: The probability that a variate would assume a value greater than or equal to the observed value strictly by chance: $P(z >= z_{observed})$ [18].

Power of a test: Power is broadly defined as the probability that a statistical significance test will reject the null hypothesis for a specified value of an alternative hypothesis. Another way to define it is the ability of a test to detect an effect, given that the effect actually exists. [21].

Significant Level of a test (α): In hypothesis testing, you seek to decide whether observed results are consistent with chance variation under the null hypothesis, or, alternatively, whether they are so different that chance variability can be ruled out as an explanation for the observed sample. The range of variation of samples that are consistent with the null hypothesis is examined, and if the observed sample is "too far out", the null hypothesis is rejected. The line you choose to divide "too far out" from "not too far out" is the level of significance. [22].

References

[1] *A Multiple Testing Procedure for Clinical Trials*, Peter C. O'Brien and Thomas R. Fleming, *Biometrics*, vol. 35, No. 3 (September 1979),pp. 549-556.

[2] *Group Sequential Methods in the Design and Analysis of Clinical Trials*, Stuart J. Pocock, *Biometrika*, vol. 64, No. 2 (August 1997),pp. 191-199.

[3] *Group Sequential Methods for Clinical Trials with a one-sided hypothesis*, David L. Demets and James H. Ware,*Biometrika*, vol. 67, No. 3 (December 1980),pp. 651-660.

[4] *The Assessment of Subjective opinion and its use in Relation to stopping Rules for Clinical Trials*, *The Statistician*, vol. 32, No. 1/2, March-June 1983, pp. 153-160.

[5] *Stopping Rules for Clinical Trials Incorporating Clinical Opinion*,L. S. Freedman, D. Lowe, P. Macaskill, *Biometrics*, vol. 40, No. 3 (September 1984),pp. 575-586.

[6] *Two-Stage Clinical Trial Stopping Rules*, Janet D. Elashoff and Terry J. Reedy, *Biometrics*, vol. 40, No. 3 (September 1984),pp. 791-795.

[7] *Monitoring Clinical Trials with a Conditional Probability Stopping Rule*,Steven M. Snapinn, *Statistics in Medicine*, vol. 11, 1992, pp. 659-672.

[8] *Overrunning and Underrunning in Sequential Clinical Trials*, John Whitehead,PhD, *Controlled Clinical Trials*, vol. 13, 1992, pp. 106-121.

[9] *When to stop a Clinical Trial*, Stuart J Pocock, *British Medical Journal (BMJ)*, vol. 305, 25 July 1992, pp. 235-240.

[10] *Textbook of Clinical Trials*, David Machin, Simon Day and Sylvan Green, Wiley 2004

[11] *Evaluating Clinical Research*, second edition, Bengt D. Furberg & Curt D. Furberg, ©2007 Springer Science+Business Media,LLC.

[12] *Clinical Statistics*, H.J.A Schouten, 1999.

[13] *Statistics, ethics and probiotica*, Richard D. Gill, Mathematical Institute Leiden University, April 16, 2008, Discussion Paper.

[14] *NRC Handelsbland*, January 23, 2008; online retrieval date: August 4, 2009.

[15] *http://acronyms.thefreedictionary.com/Clinical+Trial*, References in periodical archives, October 10, 2009.

[16] *http://dictionary.reference.com/browse/meta-analysis*, October 11, 2009.

[17] *http://medical.webends.com/kw/Clinical Trials Data Monitoring Committees*, October 11, 2009.

[18] *http://mathworld.wolfram.com/P-Value.html*, October 17, 2009.

[19] *http://www.emathzone.com/tutorials/basic-statistics/one-tailed-test.html*, October 17, 2009.

[20] *http://dictionary.reference.com/browse/conditional+probability?o=100074*, October 17, 2009.

[21] *http://cc.uoregon.edu/cnews/summer2000/statpower.html*, October 28, 2009.

[22] *http://www.statistics.com/resources/glossary/l/levelsig.php*, October 28, 2009.

A Calculated values that belong to figure (3)

alpha	0,025
beta	0,05
prej	0,95
pacc	0,106

Boundaries		
f	Rejection	Acceptance
0,05	0,0000	1,0000
0,1	0,0000	0,9999
0,15	0,0000	0,9946
0,2	0,0000	0,9682
0,25	0,0001	0,9108
0,3	0,0002	0,8273
0,35	0,0004	0,7294
0,4	0,0007	0,6280
0,45	0,0011	0,5310
0,5	0,0016	0,4428
0,55	0,0022	0,3653
0,6	0,0028	0,2987
0,65	0,0035	0,2424
0,7	0,0043	0,1953
0,75	0,0051	0,1561
0,8	0,0060	0,1235
0,85	0,0072	0,0961
0,9	0,0087	0,0728
0,95	0,0115	0,0519
1	0,0250	0,0250

alpha	0,025
beta	0,05
prej	0,9
pacc	0,224

Boundaries		
f	Rejection	Acceptance
0,05	0,0000	1,0000
0,1	0,0000	0,9974
0,15	0,0000	0,9697
0,2	0,0001	0,8987
0,25	0,0002	0,7963
0,3	0,0006	0,6822
0,35	0,0011	0,5711
0,4	0,0018	0,4707
0,45	0,0026	0,3840
0,5	0,0034	0,3112
0,55	0,0044	0,2513
0,6	0,0053	0,2024
0,65	0,0063	0,1629
0,7	0,0072	0,1310
0,75	0,0081	0,1053
0,8	0,0091	0,0845
0,85	0,0103	0,0675
0,9	0,0117	0,0533
0,95	0,0141	0,0409
1	0,0250	0,0250

alpha	0,025
beta	0,05
prej	0,85
pacc	0,345

Boundaries		
f	Rejection	Acceptance
0,05	0,0000	0,9999
0,1	0,0000	0,9870
0,15	0,0000	0,9211
0,2	0,0002	0,8090
0,25	0,0005	0,6816
0,3	0,0012	0,5598
0,35	0,0021	0,4530
0,4	0,0032	0,3635
0,45	0,0043	0,2906
0,5	0,0056	0,2320
0,55	0,0068	0,1854
0,6	0,0079	0,1485
0,65	0,0090	0,1193
0,7	0,0101	0,0963
0,75	0,0110	0,0781
0,8	0,0120	0,0636
0,85	0,0130	0,0519
0,9	0,0142	0,0425
0,95	0,0161	0,0344
1	0,0250	0,0250

alpha	0,025
beta	0,05
prej	0,8
pacc	0,465

Boundaries		
f	Rejection	Acceptance
0,05	0,0000	0,9992
0,1	0,0000	0,9588
0,15	0,0001	0,8444
0,2	0,0004	0,7015
0,25	0,0011	0,5653
0,3	0,0021	0,4487
0,35	0,0034	0,3538
0,4	0,0049	0,2785
0,45	0,0065	0,2196
0,5	0,0080	0,1737
0,55	0,0094	0,1381
0,6	0,0108	0,1104
0,65	0,0120	0,0889
0,7	0,0130	0,0722
0,75	0,0139	0,0592
0,8	0,0147	0,0489
0,85	0,0155	0,0409
0,9	0,0165	0,0346
0,95	0,0179	0,0296
1	0,0250	0,0250

B The p-values of Snapinn reproduced

The values of n_1 and n_2 does not matter in the critical p-values calculation.

			Z
alpha	0,025	0,975	1,959964
beta	0,05	0,95	1,644854
mu	0		
stdev	1		
n1	4500		
n2	4500		
p_acc	0,1		-1,28155
p_rej	0,95		1,644854

Rejection Boundary		
f	Verdeling	Rejection Boundary
0,1	1	0,0000
0,2	0,99999	0,0000
0,3	0,99983	0,0002
0,4	0,999303	0,0007
0,5	0,998382	0,0016
0,6	0,997168	0,0028
0,7	0,995735	0,0043
0,8	0,993988	0,0060
0,9	0,991264	0,0087
1	0,975	0,0250

Acceptance Boundary		
f	Verdeling	Acceptance Boundary
0,1	0,00014663	0,9999
0,2	0,03178679	0,9682
0,3	0,17268515	0,8273
0,4	0,37200018	0,6280
0,5	0,55718675	0,4428
0,6	0,70127815	0,2987
0,7	0,80465151	0,1953
0,8	0,87652006	0,1235
0,9	0,9272033	0,0728
1	0,975	0,0250

Figure 4: These are reproduced values of snapinn