Available online at www.abhipublications.org
Research Article
International Journal of Pharmacy and Engineering (IJPE)
ISSN 2320-849X

Interim analysis in clinical Trials- A simulation based approach

Aditya Chakraborty*

Department of Mathematical Sciences, University of West Florida, Pensacola, FL- 32514, United States

Abstract

In clinical trials and many other scientific studies, an interim analysis is an important aspect of an experiment. The experiment of interim analysis is done before collection of all data. If a treatment is useful or unsafe relative to an existing placebo group, researchers are morally bound to evaluate the perceived difference using the evidence at hand and make careful judgment to terminate the experiment earlier. In interim analysis, whenever a new drug shows an adverse effect on human being while testing the e effectiveness of several drugs, researchers immediately stop the trial by realizing the severity and aftermath of continuing the experiment and so forth. The purpose here is also taking into account the fact that adequate number of subjects receive most effective treatment at the earliest stage till the drug shows an adverse effect. Interim analysis often shortens average sample sizes and shorten required duration of the experiment. One would be reluctant to use additional resources if he/she already have enough evidence about the treatment effect.

In this paper, we studied interim analysis in clinical trials using simulation study. We simulate a sample of N=1913 patients and conduct interim

analysis using two looks, three looks and four looks. Two looks mean sample was divided into two groups, similarly for three and four looks. Stopping of the experiment depends on the critical values at each individual steps. In our simulation process, in four looks no adverse effect has been found.

Keywords: Interim Analysis, Stopping rule, Alpha Spending function, Sequential Probability Ratio Test (SPRT).

1. Introduction

1.1 Clinical Trials: The Overview

Clinical trials can be viewed as bio medical research investigations of human beings to test the safety of new treatments. Clinical trials are done in phases. The study follows strict scientific standards (safeguard). These standards or safeguards help to produce reliable and dependable results. Sometimes experiments is done in lab on animals. Trials may work well on animals but not work well on human subjects. So trials on human subjects are essential. For safety purposes experiment begin with small group of human subjects. In later stages, investigators gain additional information about the benefit as well as the risks from the new drug. There are three possible outcomes generally can take place whenever a treatment is applied.

It offers improvement. It offers no improvement or it causes unpredictable damage.

1.2 Why clinical trials

Clinical trials [1] give us an indication about the treatments which functions effectively or ineffectively on human subjects. Generally a clinical research answers the following two major questions:

Whether the new treatment functions effectively on human subject? If it works effectively, the next question then focuses on how efficiently it works. Does it function better and more advanced way than the current treatment being used to treat a specific disease? Needless to say that, a new treatment that does not show something advanced and newfangled undoubtedly is not worth studying.

Is the new treatment safe to use? To answer this question one must realize that, no treatment or methodologies even which is already in use is completely free of risk. However, emphasize should be given on the fact whether the benefits of the new treatment balance the possible risks.

1.3 Interim analysis in clinical trials

An interim analysis [2] is any assessment of data done during the patient enrollment time or follow-up stages of a trial with the objective of assessing performance, the quality of the data collected, or treatment effects. Interim analysis is also called data-dependent stopping or early stopping. Interim analyses are most frequently used to find convincing enough evidence to say that there is a significant treatment difference, and that the difference is convincing enough to stop the trial at a point earlier than planned at least. Ethical and Economic details are also taken into account to stop the trial early. The ethical reason is the most important reason to stop the trial. We want to make sure that the maximum number of patients receives the most effective treatment at the earliest stage. Since clinical trials are expensive, there are also economic

reasons to include as few patients as possible. Interim analysis is also performed to possibly diminish the expected number of patients and to shorten the follow-up time needed to make a conclusion.

1.4 Relevance of Interim analysis

Clinical trials are systematic investigations on human subjects. Often ethical committee is formed to come up with the protocol of the trials. The researchers must follow ethical committee's recommended protocol for all phases of the trial. Ethical aspect focuses on:

i) Minimum sample size to achieve primary objective of experiment.

ii) High variability of the efficacy of the treatment.

iii) Doubtful statistical assumptions.

iv) Clause monitoring of the trials.

These answers can improve outcomes significantly. If there is a doubt about ethical aspect of the trials, then an independent data monitoring committee(IDMC) can be set up to oversee the entire process.

Following are some reasons regarding why we stop a trial earlier than planned. [3]

- If side effect or injuriousness is so fatal that it is almost impossible to carry out the experiment, relative to the benefits.
- If the data is substandard as far as the quality is concerned. If treatments are convincingly dissimilar in nature.
- If accrual of patients is so sluggish that is very difficult to complete the experiment in right time.
- If the research questions have no more importance owing to the development from other research studies.

- If some conclusive information is already there from an external study, which makes the study useless to carry on.

- If the study integrity is sabotaged and resources at hand for continuation of the study are misplaced due to embezzlement or impropriety.

1.5 Type I and Type II Errors

Generally, there are two types of errors [4] occur in statistical testing problem. Which are called as type I error and type II errors. Optimality of tests are measured using these two types of errors. We denote, P [type I error] =α= P [reject H_0 when H_0 is true] and β = P[type II error] = P [Accept H_0 when H_0 is false]. An upper bound for is a significance level of the test procedure. Thus, if we choose α =.05 and find H_0rejected, then we should say Ho is rejected at 5% level of significance. Which implies if we repeat the test 100 times, in at most 5 cases true Ho will be rejected. Power of the test is defined as the probability of correctly rejecting the null hypothesis when the null hypothesis is false, Power of the test I defined as the probability of correctly rejecting the null hypothesis when the null hypothesis is false. i.e. Power = 1-β = P [reject H_0 when H_0 is false].1-α is called the specificity of the test and 1-β = power is called the sensitivity of the test. Here H_0 denotes the null hypothesis.

1.6 The problem with Type 1 error

In interim analysis, we divide the whole sample size into some equal or unequal sub samples and then analysis is done based on those sub samples.

We can't fix our alpha level as 0:05 throughout the study while performing an interim analysis.

We cannot just conclude that one treatment performs better than other if we see p-value (p) ≤ 0.05

Each time we look at the data, we have the likelihood of a committing type I error. If we see at the data multiple times, and we consider alpha as 0.05 as our standard significance level, then we have a 5% chance of stopping every time. Under the true null hypothesis and two looks at the data, we can approximate the error rates as:

Probability of stopping trial at first stage: 0.05

Probability of stopping trial at second stage: (0.95)*(0.05) = 0.0475

Hence, total probability of stopping becomes (.05 +.0475) = 0.0975

2. Methods:

2.1 Three basic approaches

There are three different statistical schools of thought. [5]

1. Likelihood Theory
2. Bayesian Theory
3. Frequentist Theory

All vary in their approaches and all treat interim analysis differently.

1. **Likelihood Method** - *Sequential probability ratio tests*: Even though commonly unrealistic specifically for the trials containing outcomes with prolonged follow-up times, it's conceivable to evaluate treatment e effects as soon as each patient is accrued, treated, and assessed. A likelihood function can be formed from a probability model for a series of random

variables which will agree with the outcome measurements on the experimental units.

2. Bayesian Method [6]

The Bayesian method to statistical design and inference is very dissimilar from the classical (frequentist) approach. The rudimentary considerations for Bayesian analysis are as follows.

- Before beginning of a trial, a Bayesian encapsulates the existing knowledge or belief about the treatment effect, in the form of a probability distribution. This is known as the prior distribution for unknown parameter θ.
- The data from the trial are then observed, X, and the likelihood function of X given is constructed.
- In essence of the prior distribution of θ, the distribution of theta given x based on the data X is the computed, known as posterior distribution of theta(θ).

Here θ is viewed as a random variable, about which probability statements are to be made.

This is the appealing aspect of the Bayesian approach. The precision of the estimates can be improved by incorporating prior information in Bayesian Analysis. In some situations, though Bayesian Analysis may be advantageous in comparison to Frequentist approach, different choices of prior can bring about different decisions.

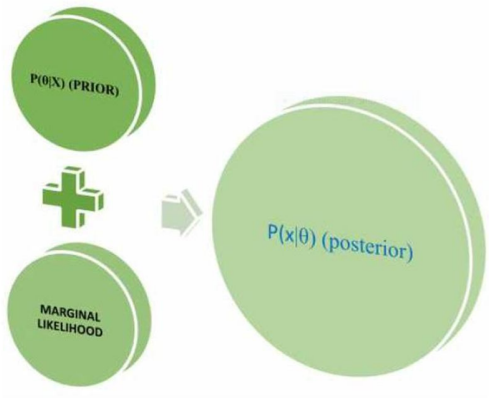

Figure - I

3 Frequentist Method [6]

In contrast to the Bayesian methodology, the frequentist approach considers as a fixed but unknown quantity (called a parameter) which can be estimated from the data at hand.

.From a frequentist point of view, repeated hypothesis testing of accruing data can increase the type I error rate of a clinical trial. Therefore, the frequentist approach to interim monitoring of clinical trials deals with focusing on how to control Type 1 error rate. Most of the multi-center clinical trial interim analyses are conducted once or twice per year.

2.3 Some approaches:

2.3.1 *Haybittle-Peto's approach*: This boundary value deals with decision rule [7] prompting about when to stop trial before the time when it was supposed to be ended. It states that whenever we find the probability (p value) less than or equal to .001 while performing an interim analysis, we reject the null hypothesis (no significance difference between the

treatments). The final analysis though is performed at the usual level of significance (.05). One major advantage of using Heybittle-Peto approach is that it uses the fixed critical value(.001) in each individual interim steps and since the final analysis is performed at natural significance level(5%) it become easier for researchers to interpret the result based on the outcome. According to some researchers, one obvious criticism of the approach is that the stopping a trial becomes too difficult resulting Heybittle-Peto approach (based on intuitive reasoning) a conservative one.

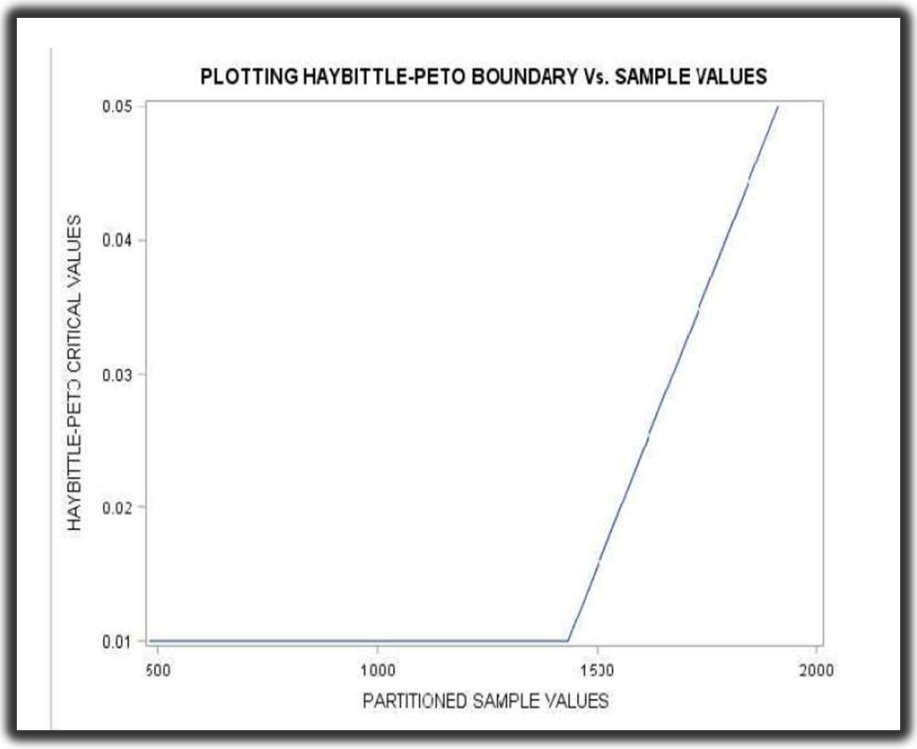

Figure- II

2.3.2 *O'brien Fleming's approach*: This approach [7] is probably the most popularly used group sequential approach since the overall significance level here approximately equals to the desired level of significance (.05) which is achieved by summing the significance levels of the previous interim steps. Though O'Brien Fleming's approach (based on statistical reasoning) considers

the fact of partitioning the significance level in a better and more rational way it too doesn't allow to attain the statistical significance at an early stage. It is not as conservative a Heybittle-Peto's approach though.

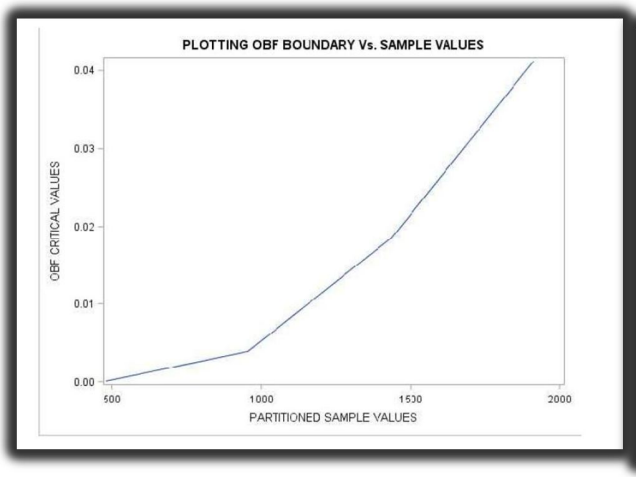

Figure-III

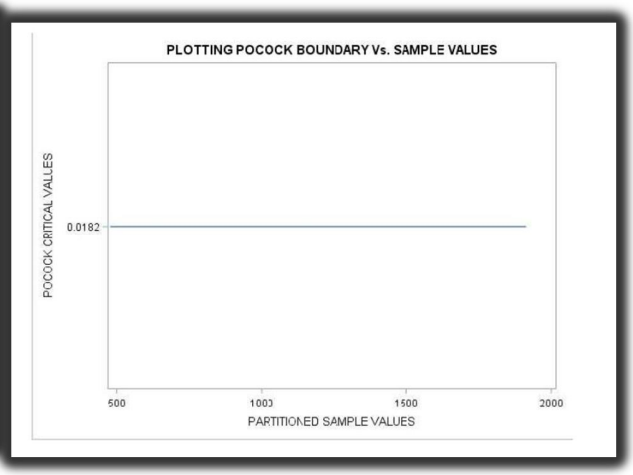

Figure-IV

2.3.3. *Pocock's approach*: This approach [7] was came into being by the medical scientist Stuart Pocock in 1977.It is simple compared to the above stated two approaches which is manifested from the fact that in each interim analysis it uses the same threshold p value. Among the three procedure described here, Pocock''s approach allows the best chance of early trial termination. Some researchers dislike this approach because of its certain disadvantages.

- The number of interim analysis has to be decided prior to the starting of any analysis and as soon as the analysis is started it is not possible to add any extra analysis to it.
- Researchers and investigators sometimes get confused about how to interpret the p value. As for example, suppose four analysis are planned and statistical significance has not reached in any of these. Suppose that the p value at the final analysis is 0:0364(> 0:0182 from table). Under this

scenario, if interim analysis had not been scheduled, then p value would be considered to project a statistical significant result (as 0.0364 < .05).

3. Statistical analysis and experiment:

3.1 Procedure:

(A) Theoretically, we would not fix level [and power] and determine the sample size for what is known as a "Fixed Sample Procedure". In Clinical Trials, patients are not expected to come all at a time; it takes a prolonged time span to attain the required sample size. Therefore, the possibility of interim use of the data and interim statistical analysis need to be explored.

(B) In contrast with the fixed sample [8] procedure, one may recommend a fixed number of Interim Tests as one goes along with the data collection process. Then we determine the Total Sample Size to attain a specified level once we specify the decision rule and the size of data for each interim test. Usually, the total sample size is divided into k equal parts and interim tests are based on the accumulated data. The rule or continuation to the next stage until k^{th} stage is reached when we end up with a terminal decision.

- Prominent contributors in this field are Armitage, Haybittle-Peto, Pocock, O'Brien and Fleming.

3.2 Some Literature Review

The prominent contributors like Pocock, Haybittle-Peto,Pocock and O'BrienFleming, Armitage worked extensively with the theory and development of this topic and found some interesting results[11][12][13]. Some of them are illustrated below. Here is an attempt towards Literature Review on some available Interim Tests. [Here Z *(.)' s are the percentile

values in each stage of a normal distribution with mean 0 and variance unity]

- If we set Z *(I) = Z *(I,II) =…….= Z*(I,II,….. k) = 1:96, then, for k = 1, level attained = 0:05. It can be shown that for k= 2,3,4,5 respectively levels attained are given by .08,.14,.20,.35.

Table showing different cut-off points for different approaches.

R	Number of interim Looks	O'Brian-Flemming		Haybittle-Peto		Pocock	
		Critical Value(Z)	Alpha	Critical Value(Z)	Alpha	Critical Value(Z)	Alpha
2	1	2.782	.0054	3.291	.002	2.178	.0294
	2	1.967	.0492	1.96	.05	2.178	.0294
3	1	3.438	.0006	3.291	.001	2.289	.0221
	2	2.431	.0151	3.291	.001	2.289	.0221
	3	1.985	.0471	1.96	.05	2.289	.0221
4	1	4.084	.00005	3.291	.001	2.361	.0182
	2	2.888	.0039	3.291	.001	2.361	.0182
	3	2.358	.0184	3.291	.001	2.361	.0182
	4	2.042	.0412	1.96	.05	2.361	.0182
5	1	4.555	.000005	3.291	.001	2.413	.0158
	2	3.221	.0013	3.291	.001	2.413	.0158
	3	2.630	.0085	3.291	.001	2.413	.0158
	4	2.277	.0228	3.291	.001	2.413	.0158
	5	2.037	.0417	1.96	,05	2.413	.0158

- Suggested Ad Hoc Rule: For Z (I) = Z (I,II) =…….= Z (I,II,……k) = 2.6 and for large k this yield approximate level 0:05.
- Haybittle-Peto Procedure: If we use a common value 3.291 for all Z* above except for final stage when 1.96 is used. For k = 5 it can be shown that this choice attains = 0.05
- Pococks Procedure: Use common value for Z* to attain exact levels .For k = 5 and = 0:05, common value is Z = 2.413

3.3 Cure rate [9]

The proportion of individuals having a particular illness that are cured by a given treatment (drug), called the cure fraction or cure rate. Inherent in the conception of a cure is the everlasting termination to the particular instance of the disease. When a person having common cold recovers from it, the person is said to be cured, even though the person might someday catch another cold. Contrariwise, a person who has successfully managed a disease, such as diabetes, so that it yields no unwanted symptoms for that time being, but without actually permanently terminating it, is not cured.

3.4 A problem and analysis

Suppose certain drug producing company wants to introduce a new drug in the market. Let it be our test drug A(test treatment).Here, a testing problem involving the % cure rate for patients under a 'test treatment'(Drug A) against a 'standard treatment'(Drug B) with both-sided alternatives has been developed to work out on explicit expressions for

the sample size and the cut-off point considering = .05 and power [against specified alternatives with a difference of 5%] = 90%. An interim analysis in four looks has been performed involving the two drugs each for one group.

- **Obtaining the data**: Data has been simulated using SAS and VBA. Cure rate has been assumed .35 for this problem as chances of being cure is very low for life threatening diseases [10].

Hypothesis of interest: $H_0: P_T = P_S$ Vs. $H_1: P_T \neq P_S$……(*)

Where P_T: % cure in test treatment (Drug A) and P_S: % Cure in standard treatment (Drug B) Given,

α= 0:05, Power = 90% δ= 0:05 (clinical meaningful difference)

3.5 Full Analysis:

Determining the sample size: Here $Z_{\alpha/2} = 1.96$, $\beta = .10$, $Z_{\beta}=1.282$ and δ= .05.

Given the above information, we computed N= the total sample size to be 1913 approximately. That is a total of 1913 subjects are required for full experiment.

3.6 Interim Analysis: In interim analysis, we will now divide our total sample size, hat is 1913 subjects in two, three and four different parts and based on the results obtained in each step we'll conclude accordingly. One can divide the total sample size as many different parts as he/she wants depending on the requirement.

Two looks: Now we'll divide the whole sample into two halves. That is each arm involves 956 subjects. For each 956 subjects we perform the test given in (*) and in each step our result suggests fail to reject the null hypothesis (in *) by three standard rules [11] [12] [13]. (Pocock, Haybittle-Peto, Pocock and O'Brien Fleming)

Three Looks: Here we divided our complete data set (1913) approximately into three equal parts that is each arm consists of 638 observations and like a two look problem we similarly perform a test involving the % cure rates of the two drugs. Here also these rules suggest not to reject the null.

Four Looks: In four looks problem we chop our total number of observations approximately in 4 equal parts that is 478 subjects in each hand and in each part we generate number of cured and non-cured by the two drugs. Hence by computing the proportion of cure rate we perform the statistical tests [12] [13]. In this case also, we fail to reject our null hypothesis by the criterions of the three prominent contributors.

4. CONCLUSION AND FURTHER STUDY

4.1 Conclusion:

Since we assumed the average cure rate (35%) to be very small for this problem, statistically as well as intuitively it seems that cure rate for both the drugs are same. Patient enrollment would have stopped immediately if at any stage our computed Z value (percentile value of a normal distribution having mean 0 and variance unity) exceeded the cut-off points

(percentile values) of the same distribution as suggested by Haybittle-Peto, O''Brien Fleming and Pocock.

4.2 Ideas for further study

1. Like alpha-spending function discussed previously for controlling type I error margin, is there any rule to get maximum power between and within looks?

2. Among the three rules discussed earlier which rule is the most efficient with respect to economic as well as statistical point of view? Different estimators using the rules can be defined and their precision can be compared.

3. Is there any other simple sampling algorithm for obtaining the boundary values of Interim Analysis?

4. In the above problem what should be appropriate way to do if we come across a rejection at a certain stage but there is a long way to go as far as the sample size is concerned?

5. Acknowledgement

The author is truly grateful to Dr. Bikas k Sinha for giving the idea of research.

6. References

[1] Michael R. Chernick The Essentials of Biostatistics for Physicians, Nurses, and Clinicians. August 4, 2011,John Wiley & Sons

[2] Sophie D. Foss and Eva Skovlund Journal Of Clinical Oncology,clinical Oncology

[3] Guosheng Yin Clinical Trial Design: Bayesian and Frequentist Adaptive MethodsJune 7, 2013 John Wiley & Sons

[4] David L. DeMets and Gordon Lan The alpha spending function approach to interim data analysesalpha spending function

[5] O'Brien,; P.C. Fleming, T.R. (1979) A Multiple Testing Procedure for Clinical TrialsBiometrics 35: 549556

[6] Christopher Jennison Bruce W. TurnbullGroup Sequential Methods with Applications to Clinical Tri-alsSeptember 15, 1999 CRC Press

[7] Pocock SJ (1983) Clinical Trials: A Practical Approach.New York: Wiley.

[8] Cynthia O. Sin and K.K Gordon Lan Flexible Interim Analysis For Sample Size Re-estimation and Early Stopping: A Conditional Power ApproachNew York,NY 10017-5755,New London,CT 06320

[9] Shein-Chung Chow, Hansheng Wang, Jun Shao Sample Size Calculations in Clinical Research (Chapman Hall/CRC Biostatistics Series) 2 Rev Exp Edition

[10] A.S Hedayat and Bikas K. Sinha Interim Statistical Analysis In Clinical Trials

[12]Haybittle JL (1971) Repeated assessment of results in clinical trials of cancer treatmentBr J Radiol 44:793- 797.

[13] Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials Biometrika 64:191-199.

Received: *11th November 2017, Revised: 19th November; 2017, Accepted: 22nd November; 2017*

Corresponding Author: * Aditya Chakraborty*

Department of Mathematical Sciences, University of West Florida, Pensacola, FL- 32514, United State, Email aditya40in@gmail.com