Principles of Interim Analysis

NJ Gogtay, UM Thatte

Introduction

A clinical trial is a prospectively conducted study that evaluates the effect of two or more interventions [drugs, devices, vaccines, diagnostic tests or even surgical procedures] in human participants under a set of pre-defined conditions. The Randomized Controlled Trials [RCTs] among all trials form the cornerstone for generating evidence and are considered the “gold standard”. The number of participants required for an RCT is calculated before beginning the study and the study continues until the estimated sample size is reached. There is only one final analysis that is planned and carried out when the requisite sample size is reached. Inferences are subsequently drawn.

There are several situations, however, where one may need to review and analyze the data before the study ends. In other words, one may need to perform an Interim Analysis, i.e., analysis of data while the study is ongoing. While these analyses can be performed for any kind of study, they are usually conducted when the condition under study is life threatening, has potentially serious outcomes or the duration of therapy is long. Interim analysis (or multiple “looks” at the data while the study is ongoing) is decided at the protocol planning stage itself. Let us understand the concept of interim analysis with two well-known examples from literature where pre-planned interim analyses led to study discontinuation, but for diverse, disparate reasons.

The AIDS Clinical Trials Group [ACTG]-076 trial was designed to evaluate the use of Zidovudine [AZT] relative to placebo for the prevention of mother to child transmission of HIV from infected women to their babies. The study was planned to be run over 5 years with a target sample size of n = 748 mothers and the outcome of interest was the number of infant infections in the two groups. The study had a total of three planned interim analyses and one final analysis. The first interim analysis was decided to be conducted at the point where a third of the projected infant infections were reached. The idea behind this was to stop the study in the event that AZT was shown to be effective. The study was indeed stopped by the Data Safety and Monitoring Committee [DSMC, see later] at the first interim analysis as AZT was clearly superior [A total of n = 477 women were enrolled at this point and neonatal infection rates were 8.3% with AZT and 25.5% with placebo at 18 months of follow up]. The DSMC also recommended that all mothers in the control group be given AZT so as to derive benefit.

The Cardiac Arrhythmia Suppression Trials [CAST I and II studies] tested the hypothesis as to whether suppressing ventricular arrhythmias in patients with a recent Myocardial Infarction [MI] reduced sudden death and mortality. The DSMC permitted the study to continue at the first interim analysis. However, it was seen in the second interim analysis that in patients treated with the anti-arrhythmic agents encainide and flecainide, there was a 3.6-fold excessive risk of death [relative to patients treated with placebo] and the studies were stopped for safety concerns.

Thus, interim analyses are an important component of clinical research in general and drug development in particular. This article discusses the historical evolution of interim analyses, the definition/s, approaches to one, available methodologies to conduct one [with their merits and demerits], the concept of the “alpha spending function”, pitfalls of doing an interim analysis and finally the role of Data Safety and Monitoring Committees [DSMC] in its conduct.

Interim Analysis – Historical Perspective and Definitions

History- As clinical trials slowly but surely became the benchmark for Evidence Based Medicine, the National Institutes of Health [NIH] in the 1960s, set up a committee chaired by Dr Bernard Greenberg, a statistician from the University of North Carolina to address the challenges associated with the design, conduct, monitoring and analysis of trials and in particular complex multi-institutional studies. The idea behind setting up the committee was three-fold-detect unexpected or unacceptable toxicity early, detect benefit early and understand the trade-offs between benefits and risks early
so that research participants could be adequately protected. The committee released its report in 1967 and it came to be known as the Greenberg report, which laid the foundations of interim analysis.\textsuperscript{6} It also put forth the fundamental principle that clinical trials should not be continued for longer than necessary and should definitely not cause harm to participants. It also advocated the setting up of independent committees to monitor data as they accrue [the Data Safety and Monitoring Committees or DSMCs- see later]

Definitions - In its simplest of definitions, an interim analysis constitutes analysis of data while the study is still in progress. It can also be defined as one or more planned analyses of data before the final planned analysis that permits investigators and/or funders of the study to evaluate the probability of the study’s success [or failure] while controlling for statistical errors [see below for statistical errors].\textsuperscript{7}

Why an Interim Analysis is Needed and Questions that it Attempts to Answer

There are three broad reasons for doing an interim analysis

\textbf{Ethics}

It is an ethical imperative to stop a study for two reasons – 1) when there is adequate evidence to show that one intervention is clearly superior to the other [benefit] and 2) when there is adequate evidence to stop a study for safety [or lack thereof] and prevent further participants from being exposed to undue risk. We have already seen one example of each – the AZT versus placebo study that was stopped for benefit and the CAST studies that were stopped for safety concerns. Let us now see two more examples.

\textit{Study stopped for benefit}

The Anglo-Scandinavian Cardiac Outcomes Trial- Blood Pressure Lowering Arm [ASCOT-BPLA] is an example of a study stopped for benefit. It was designed to evaluate if amlodipine – based antihypertensive regimens were superior to atenolol-based antihypertensive regimes. The end points of interest were non-fatal myocardial infarction [primary end point] and fatal coronary heart disease. A total of 19,257 patients were followed up for a median of 5.5 years and the trial was stopped as stroke [327 vs. 422, p = 0.0003], cardiovascular death [263 vs. 342, p =0.001] and all cause death [738 vs. 820, p = 0.02] were all significantly lower in the amlodipine group. The trial was stopped for benefit.\textsuperscript{8}

\textit{Study stopped for safety concerns}

The Blood Conservation Using Antifibrinolytics [BART study] was a blinded RCT that compared aprotinin, tranexamic acid, and aminocaproic acid in patients undergoing high-risk cardiac surgery. The outcomes of interest were massive postoperative bleeding and 30-day mortality. A significantly higher 30-day mortality was seen with aprotinin relative to the other two drugs [relative risk 1.53, 95% CI 1.06-2.22] leading to the study’s early termination for safety concerns.\textsuperscript{9}

\textbf{Lack of difference}

The second reason why an interim analysis is needed is that during the course of the study, the data may point to the fact that the two treatments do not differ significantly. In such a situation, it would be desirable, for reasons of both cost and ethics to discontinue the study as it would be \textit{futile} to continue. The PRESENT study [trial number [NCT01479244] is one such example. This study was a multicentre randomized, double-blind controlled study that evaluated n=758 women with early-stage node positive breast cancer, who had low to intermediate HER-2 neu expression. The women were randomized either to receive a novel vaccine called Nelipepimut [the vaccine would stimulate cytotoxic T lymphocytes to destroy HER-2 neu expressing cancer cells and thus produce its effect] along with standard of care or the vaccine adjuvant along with standard of care. The study was stopped by the DSMC for futility as it was felt that continuing the study further would not have a shown a difference between the two treatment arms.\textsuperscript{10}

\textbf{Slow accrual}

When there is very slow participant enrolment for any reason, a decision can be taken to either continue with the study if the research question is truly important or stop the study for futility.

How an Interim Analysis Helps

Interim analysis helps to reduce the ‘time to market’ for an intervention by permitting key decisions to be made earlier as data accrues. It broadly helps in two ways- a) Avoid drawing a wrong conclusion b) Avoid taking too long to draw the right conclusion.\textsuperscript{11} These analyses, however, need to be properly designed and appropriately executed, so that the integrity of data is maintained and decision-making does not compromise either the process of drug development or that of patient safety. This decision-making process is fairly complex and an interplay of several factors is involved.

The Concept of “Distributing the Alpha Across the Interim Analyses” and Understanding the Impact of the Statistical Errors on the Analysis

When we calculate the total number of patients for a study [the sample size], we use the following four elements- alpha error [usually set at 5% and corresponds to 95% Confidence Intervals], beta error
In interim analysis, however, there is considerable importance to give to the alpha error. This is likely to stem from the fact that a false positive finding will have more serious implications. Thus, most interim analysis strategies center around it [see below].

Table 1: Comparison of the alpha values of three Frequentist approaches for interim analysis - O’Brien Fleming, Haybittle-Peto and Pocock approach [16]

<table>
<thead>
<tr>
<th>Number of analyses</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>O’Brien Fleming</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0054</td>
<td>0.0492</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0006</td>
<td>0.0151</td>
<td>0.0471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.00005</td>
<td>0.0039</td>
<td>0.0184</td>
<td>0.0412</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.000005</td>
<td>0.0013</td>
<td>0.0085</td>
<td>0.0228</td>
<td>0.0417</td>
</tr>
<tr>
<td><strong>Haybittle – Peto</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Pocock</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0294</td>
<td>0.0294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0221</td>
<td>0.0221</td>
<td>0.0221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.0158</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.0158</td>
</tr>
</tbody>
</table>

[Adapted from Piantadosi S, 1997]

There are three points to remember here – 1) If we were to “use up” the entire alpha of 5% each time that we carry out an interim analysis, we would end up grossly inflating the alpha and thereby risk a false positive error. For example, “5 looks” at the data would increase the alpha error to 15% and 10 looks to almost 20%. 13 2) While adjusting the alpha at every “look”, it is also important to remember that the sample size at each look is much smaller than the final sample size (The beta error of 10 or 20% is considered for this final sample size calculation). Thus, the power at each “look”, will be lower than that considered at the point of initial sample size calculation. In other words, both errors are impacted with every “look” at the data 3) There is no reason to believe that one error is worse than the other. Rather, the choice of the interim analysis strategy [see below] will affect each of these errors.

In interim analysis, however, there is considerable importance to give to the alpha error. This is likely to stem from the fact that a false positive finding will have more serious implications. Thus, most interim analysis strategies center around it [see below].

**Using the Alpha Judiciously – the O’Brien-Fleming, Pocock, Haybittle-Peto Methods**

There are several ways in which the alpha can be adjusted or used judiciously over the range of interim analyses planned. These are described below along with their merits and demerits and also given in Table 1. 16

**O’Brien and Fleming approach**

This popular method uses a very small amount of the alpha in the initial stages and reserves a large part of the alpha for the final analyses. For example, when one interim and one final analysis are planned, 0.0054 of the alpha is expended first and 0.0492 reserved for the final analysis. This method ensures that it is difficult to reject the null hypothesis in the early stages of the study, but relatively easy later on.

**Pocock approach**

This method divides the alpha error equally amongst the total number of analyses planned. For example, if there is one interim and one final analysis, the p value expended at each analysis is the same i.e., 0.029 [Table 1].

**Haybittle Peto or Peto approach**

This approach uses a miniscule amount of the alpha in the initial “looks” [much lower than the O’Brien and Fleming approach], but the final analysis is always performed using the entire 5% alpha. [Table 1] This method thus makes it very easy for investigators and readers to apply as 5% at the end of the study is what they are comfortable with. The criticism of this approach is that that the extremely low alpha values are going to make it almost impossible to stop the study at the early stages.

All three approaches are called Frequentist approaches [see later]. They are all inflexible in the way that the interim analysis is planned and executed. This led statisticians to devise yet another approach; one of the Alpha Spending Function.

**Lan and de Mets- the alpha spending function**

This is a very flexible procedure that can accommodate unequal timing, additional looks at data beyond what was originally
planned including even extending the trial. Researchers can choose to “spend” their alpha (i.e., conduct interim analyses) any way they want and the method still ensures that the total alpha “spent” is no more than 0.05 (or whatever was specified at the beginning of the study).17

**Approaches to Interim Analysis – the Frequentist, Bayesian and Mixed [Bayesian – Frequentist approach]**

The approaches to an interim analysis can be broadly divided into two – Frequentist and Bayesian. Let us first understand the philosophy behind the two approaches with an example from day to day life which will serve as a metaphor for understanding these approaches in clinical research and drug development.

All of us use cell phones and are prone to misplacing them. Let us say that Mr. X generally misplaces his phone and more so at home. There are the usual three places at home that he is likely to have left the phone – the bedroom, the television room, and the dining area. Thus, when he realizes that his cell phone is “missing”, he “knows” that it likely that it will be found in one of the three places with equal probability [33% at each of the three places]. This is “prior” knowledge or information; one that available before beginning the study. Next, he needs a “way” to locate his cell phone [the study methods]. Finding the cell phone will “answer” his search or the research question [the study results]. He calls from the landline at home [the method stated in the protocol] and traces the sound. Based on where the sound leads him, he will “find” the cell phone.

Let us assume that Mr. X misplaces his phone at home on three consecutive days - Day 1, Day 2 and Day 3. On all three days, he finds it in the dining area [rather than the other two areas]. Despite finding it in the dining area on three days in succession, his initial “model” that was developed based on prior information [which stated the probability of finding it in the three places would be equal], does not change. This is the “Frequentist” approach.

On the other hand, when the cell phone is found in the dining area on three consecutive days, the Bayesian approach would “change the model” to now state that the probability of finding the phone in the dining room is higher than the other two places. In other words, the results have changed the model [Bayesian thinking].

The Frequentist and Bayesian approaches thus represent two philosophies about quantifying this uncertainty that exists in the clinical research and drug development process. Let us now understand that actual processes followed in the two approaches.

**Frequentist approach -** Here, information from existing/available data also called prior information [previous studies, literature] is used at the protocol development stage. At the point of data analysis, prior information is considered as a complement to, but not part of, the formal analysis18 and the initial model remains unchanged. This approach presents results using p values, standard error of mean and the 95% confidence intervals and is also called the “traditional” or “classical” approach. Interim analyses in literature have been largely dominated by the Frequentist approach. An example of the use of the Frequentist philosophy can be found in the study by Schwartz [2001]19 who evaluated the effect of atorvastatin vs placebo on early recurrent ischemic events in acute coronary syndromes. The study protocol pre-specified 3 interim analyses that would each use an alpha value of 0.001 and the final analysis had an alpha value of 0.049 allocated to it. [O’Brien Fleming approach]. The study showed the utility of 80mg/d of atorvastatin in reducing ischemic events over a 16-week treatment period at the final analysis. The results of the study did not change the initial model the investigators began with.

**Bayesian approach –** This approach uses and learns from evidence as it accumulates.18 Here, while we do use prior information similar to the Frequentist approach; also called prior belief and we now combine it with new information that accumulates as the study continues. The combination of old information coupled with new evidence goes beyond protocol development; can be applied to the conduct of the trial and also at the analysis stage. In this approach, prior information and the accruing results are considered seamless and inferences are drawn and updated each time that new data becomes available. The Bayesian approach thus directly address the question of how new evidence should change what we currently believe.20 These are also called as “learn as you go” approaches.

Bayesian analyses are computationally intense and thus less used compared to the Frequentist approaches. Recent advances in the past decade, in particular the development of computational algorithms, hardware and high computing speed have made it possible to use Bayesian approaches much more and these are now seen fairly frequently in literature. An example of a protocol that uses a Bayesian approach can be seen in the study planned by Carlson21 on the use of Docosahexaenoic acid (DHA) supplementation [relative to soybean + corn oil] for reducing the frequency of early pre-term births. The rationale for the study is that DHA has been associated with longer gestation, higher birth weight and less pre-term births [prior information]. A total of n = 1200 women are planned to be...
enrolled. The first interim analysis is slated once n=150 women have been enrolled in each group. The data accrued will be analyzed in a blinded manner. Subsequently, more patients will be assigned to the more promising group. In other words, the data gathered will change the conduct of the study.

Studies can also use a combination of the two approaches.

**Study Designs and Interim Analyses**

RCTs can incorporate interim analysis in one of several ways. 1) A Traditional Design RCT, one that has a fixed sample size, but no interim analysis. 2) A Traditional Design RCT with interim analysis using one of the Frequentist approaches 3) Interim analysis that actually “modifies one or more aspects of the study design, or even the hypothesis based on analysis of accumulating data”20,22 This constitutes an Adaptive Design.

**Data Safety and Monitoring Committees [DSMC]**

A clinical trial DSMC is a group of individuals [independent of the study and appointed by the sponsor] with pertinent expertise who review accruing data on a regular basis from one or more ongoing clinical trial and take decisions independent of the funder of the study.22 These committees are also called Independent Data Safety Monitoring Committees [IDSMC]. Like Interim Analyses, historically, the establishment of DSMCs can also be traced to the Greenberg report6 that recommended the establishment of independent committees based on the recognition that looking at accumulating study data was essential to ensure the ongoing safety of trial participants. Another reason stated by the Greenberg report was that those closely involved with the study design and conduct may not be objective enough to review interim data for any emerging concerns and address them.

The primary role of the DSMC is to ensure that the safety of participants in the study is protected. This is done by them in one of three ways 1) terminate the study early either for safety concerns and thereby prevent further harm or when there is overwhelming evidence of efficacy. The latter ensures that all participants then receive the better intervention 2) terminate the study when there is adequate evidence pointing towards futility in continuation 3) Permit study continuation in anticipation of benefit after a due assessment of risk – benefit based on data accrued thus far.

The committee usually has 3–5 individuals with extensive clinical experience both in the disease under study, and in the management of large complex clinical trials. A DSMC has two clearly designated positions – that of the Chair and the statistician. The decision-making meetings are “closed door” and the chair communicates the minutes to the sponsor. It is important that the DSMC consider the “totality of evidence” before any recommendation and in particular when they recommend that trials be stopped early. Apart from the statistical analyses, the amount of data accrued, the nature of the results seen up to that point, their implications, cost of gathering more data and the benefit-risk assessment of what subsequent patients will be exposed to, all must be considered in tandem for decision making.

**Conclusions**

The process of interim analysis or reviewing data before it is fully collected helps in decision making in both drug development and clinical research. Regardless of the approach used for an interim analysis, it is important to remember that statistical analysis is simply a part of the whole picture. Beyond statistical significance, any treatment difference seen needs to be clinically meaningful — that is, large enough to actually matter when the intervention is used finally in clinical practice. The decision regarding termination [for benefit, safety concerns or futility] or continuation of a study, thus, must also be driven by other aspects such as cost of the drug, ease of administration, nature of the toxicity seen, and benefit-risk assessment for the subsequent patients.23

**Acknowledgements**

The authors are grateful to Dr Jaideep Gogtay [Cipla Ltd, Mumbai], Dr L Jeyaseelan [Christian Medical College, Vellore], Dr Sudeep Gupta [Advanced Centre for Training Research and Education [ACTREC], Navi Mumbai] and Dr Manju Sengar [Tata Memorial Hospital, Mumbai] for their critical inputs that helped refine the manuscript.

**References**

6. Heart Special Project Committee. Organization, review, and administration of cooperative studies [Greenberg Report]: a report from the Heart Special Project Committee to the National Advisory Heart


