

Sample Size Estimation in Clinical Trial

Tushar Vijay Sakpal

Abstract

Every clinical trial should be planned. This plan should include the objective of trial, primary and secondary end-point, method of collecting data, sample to be included, sample size with scientific justification, method of handling data, statistical methods and assumptions. This plan is termed as clinical trial protocol. One of the key aspects of this protocol is sample size estimation. The aim of this article is to discuss how important sample size estimation is for a clinical trial, and also to understand the effects of sample size over-estimation or under-estimation on outcome of a trial. Also an attempt is made to understand importance of minimum sample to detect a clinically important difference. This article is also an attempt to provide inputs on different parameters that impact sample size and basic rules for these parameters with the help of some simple examples.

Key Words: Sample size, statistically significant, clinically significant, Type I error, Type II error, power,

Introduction

The article will cover

- 1) Why sample size estimation is important in a clinical trial (CT)
- 2) Inputs and their importance for sample size estimation in CT
- 3) Examples on sample size estimation
- 4) Discuss some basic rules and guidelines

Importance of Sample Size in CT

Let us see an example. Dr. ABC had developed a drug NEW which was effective in reducing pain. The drug NEW is clinically better in terms of efficacy and safety as compared to test drug. Hence, the drug NEW truly deserves to reach market so that patients suffering from pain can be at ease.

Dr. ABC decided to conduct a CT in order to get approval from regulators. How will the sample size affect Dr. ABC's trial on patients suffering from pain?

Scenario 1 – Sample size under-estimation (Sample size selected was less than what was required)

At the end of trial when data was analyzed, statistical significance was not achieved (statistical jargon, p-value was ≤ 0.05).

Underestimation of sample size may result in drug turning out to be statistically non-significant even though clinical significance exists.

Scenario 2 – Sample size over-estimation (Sample size selected was much more than what was required)

Dr ABC conducted CT with a large number of subjects. When the data was analyzed, strong statistical significance was achieved as p-value turned out to be 0.000001. However Dr ABC has to consider the following issues:

- 1) The above p-value indicates that a smaller sample size was needed to prove statistical significance. This raises ethical issues as more subjects were exposed to test drug, which could have deserved getting NEW drug.
- 2) As sample selected was very large even a small difference between NEW and test drug will turn out statistically significant even if that difference is not clinically meaningful.

Hence sample size is an important factor in approval / rejection of CT results irrespective of how clinically effective / ineffective, the test drug may be.

What does it take to estimate a statistically appropriate sample size?

Basic Requirement for Estimation of Sample Size in CT

The estimation of sample size along with other study related parameters depends on Type I error, Type II error and Power.

Type I Error:

Consider we are testing two brands of paracetamol to evaluate if Brand 1 is better in curing subject's suffering from fever as compared to Brand 2. As both brands contain paracetamol, it is expected that the effect of both brands is similar. Let us try to build a statistical hypothesis around this

Null Hypothesis (H_0): Brand 1 is equal to Brand 2

Alternate Hypothesis (H_1): Brand 1 is better than Brand 2

Let us try to evaluate error that can occur.

Error 1: Based on analysis it is concluded that Brand 1 is better than Brand 2, basically we reject H_0 . Knowing that Brand 1 is equal to Brand 2 (H_0), we are making an error here by rejecting H_0 . This is called as Type I error. Statistically it is defined as

Tushar Vijay Sakpal

Principal Statistician, Biostatistics, PharmaNet Clinical Services Pvt. Ltd.

Sample Size Estimation

Type I error = (Reject H_0 / H_0 is true).

Probability of Type I error is called as level of significance and is denoted as α .

Type II Error:

Consider we are testing paracetamol against placebo to evaluate if paracetamol is better in curing subject's suffering from fever as compared to placebo. (It is expected that the effect of paracetamol is better than placebo). Let us try to build a statistical hypothesis around this

Null Hypothesis (H_0): paracetamol is equal to placebo

Alternate Hypothesis (H_1): paracetamol is better than placebo

Let us try to evaluate the error that can happen.

Error 2: If analysis concludes that paracetamol is equal to placebo, we accept H_0 . Knowing that paracetamol is better than placebo (H_1) we are making an error here by accepting H_0 . This is called as Type II error. Statistically it is defined as

Type II error = (Accept H_0 / H_1 is true).

Probability of type II error is denoted as β .

In above case if analysis concludes that paracetamol is better than placebo, we reject H_0 which would be correct decision. Probability of such a decision taking place is called as "Power".

Power = Probability (Reject H_0 / H_1 is true) which is actually $1 - \beta$.

We can tabulate type I and type II error as

Decision Taken \ Actual Fact	H_0 is True	H_1 is True
Reject H_0	Type I error	No error
Accept H_0	No error	Type II error

Following are the basic considerations for estimating sample size for a CT

1) Study Design

In CT, different statistical study designs are available to achieve objectives. Typical designs that may be employed are parallel group design, crossover design etc.

For sample size estimation study design should be explicitly defined in the objective of the trial. A cross over design will have different approach and formula for sample size estimation as compared to parallel group design.

2) Hypothesis Testing

This is another critical parameter needed for sample size estimation, which describes aim of a CT. The aim can be equality, non-inferiority, superiority or equivalence.

Equality and equivalence trials are two-sided trials where as non-inferiority and superiority trials are one-sided trials.

Superiority or non-inferiority trials can be conducted only if there is prior information available about the test drug on a specific end point.

3) Primary Study End Point

The sample size calculation depends on primary end point

of a CT. The description of primary study end point should cover whether it is discrete or continuous or time-to-event. Sample size is estimated differently for each of these end points. Sample size is adjusted if primary end point involves multiple comparison.

4) Expected Response Test vs. Control

The information about expected response is usually obtained from previous trials done on the test drug. If this information is not available, it could be obtained from previous published literature. The important information required is:

- a) response expected with test drug (mean expected score / proportion of subjects achieving success.)
- b) response expected with control drug (mean expected score / proportion of subjects achieving success.)
- c) variability (standard deviation.)

5) Clinical Important / Meaningful Difference / Margin

This is one of most critical and one of most challenging parameters. The challenge here is to define a difference between test and reference which can be considered clinically meaningful. To put it in plain English, this is the difference which will make your family doctor to give you new medication over and above the existing gold standard which he is prescribing for last 10 – 20 years. This threshold figure is at times not easily available and should be decided based on clinical judgment.

6) Level of Significance

This is typically assumed as 5% or lesser.^{1,2} Type I error is inversely proportional to sample size.

7) Power

As per guideline.^{1,2} power should not be less than 80%. Type II error is directly proportional to sample size.

8) Drop-out rate

The sample size estimation formula will provide number of evaluable subjects required for achieving desired statistical significance for a given hypothesis. However in practice we may need to enroll more subjects to account for potential dropouts.

If n is the sample size required as per formula and if d is the dropout rate then adjusted sample size N_1 is obtained as $N_1 = n / (1 - d)$.

9) Unequal Treatment Allocation

For some of clinical trials it is ethically desirable to have more subjects in one arm compared to the other arm. For example, for placebo-controlled trials with very ill subjects it is unethical to assign equal subjects to each arm. In such cases sample size is adjusted as below.

If n is sample size required as per formula and n_1 is the sample size for test and n_2 is sample size for placebo and $n_1 / n_2 = k$, then

Sample Size Estimation

$n1 = (0.5) \times n \times (1 + k)$
 $n2 = (0.5) \times n \times (1 + (1/k))$
Suppose $n = 100$ and ratio between test ($n1$) and placebo ($n2$) is 2:1 ($k=2$) then
 $n1 = (0.5) \times (100) \times (1+2) = 150$
 $n2 = (0.5) \times (100) \times (1+[1/2]) = 75$
As a statistician one should have clear and clinically meaningful inputs on the above points prior to working on sample size estimation.

Calculation of Sample Size

Example 1: Comparing two proportions

A placebo-controlled randomized trial proposes to assess the effectiveness of Drug A in curing infants suffering from sepsis. A previous study showed that proportion of subjects cured by Drug A is 50% and a clinically important difference of 16% as compared to placebo is acceptable.

Level of significance = 5%, Power = 80%, Type of test = two-sided

Formula of calculating sample size is

$$n = [(Z_{\alpha/2} + Z_{\beta})^2 \times \{(p1(1-p1) + (p2(1-p2)))\} / (p1 - p2)^2]$$

where

n = sample size required in each group,

$p1$ = proportion of subject cured by Drug A = 0.50,

$p2$ = proportion of subject cured by Placebo = 0.34,

$p1-p2$ = clinically significant difference = 0.16

$Z_{\alpha/2}$: This depends on level of significance, for 5% this is 1.96

Z_{β} : This depends on power, for 80% this is 0.84

Based on above formula the sample size required per group is 146. Hence total sample size required is 292.

The description sample size in the protocol will be:

A sample size of 292 infants, 142 in each arm, is sufficient to detect a clinically important difference of 16% between groups in curing sepsis using a two-tailed z-test of proportions between two groups with 80% power and a 5% level of significance. This 16% difference represents a 50% cure rate using drug A and 34% cure rate using placebo.

Example 2: Comparing two means

An active-controlled randomized trial proposes to assess the effectiveness of Drug A in reducing pain. A previous study showed that Drug A can reduce pain score by 5 points from baseline to week 24 with a standard deviation (σ) of 1.195. A clinically important difference of 0.5 as compared to active drug is considered to be acceptable. Consider a drop-out rate of 10%.

Level of significance = 5%, Power = 80%, Type of test = two-sided

Formula of calculating sample size is

$$n = [(Z_{\alpha/2} + Z_{\beta})^2 \times \{2(\sigma)^2\} / (\mu1 - \mu2)^2]$$

where

n = sample size required in each group,

$\mu1$ = mean change in pain score from baseline to week 24

in Drug A = 5,

$\mu2$ = mean change in pain score from baseline to week 24

in Active drug = 4.5,

$\mu1 - \mu2$ = clinically significant difference = 0.5

σ = standard deviation = 1.195

$Z_{\alpha/2}$: This depends on level of significance, for 5% this is 1.96

Z_{β} : This depends on power, for 80% this is 0.84

Based on the above formula, the sample size required per group is 90. Hence total sample size required is 180. Considering a drop-out rate of 10% total sample size required is 200 (100 in each arm).

The description sample size in the protocol will be: A sample size of 180 subjects, 90 in each arm, is sufficient to detect a clinically important difference of 0.5 between groups in reducing pain assuming a standard deviation of 1.195 using a two-tailed t-test of difference between means with 80% power and a 5% level of significance. Considering a dropout rate of 10% the sample size required is 200 (100 per group).

Basic Rules for Estimating Sample Size in CT

Some basic rules for on sample size estimations are:

- 1) Level of significance – It is typically taken as 5%. The sample size increases as level of significance decreases.
- 2) Power – It should be $\geq 80\%$. Sample size increases as power increases. Higher the power, lower the chance of missing a real effect.
- 3) Clinically meaningful difference - To detect a smaller difference, one needs a sample of large size and vice a versa.
- 4) Sample size required to demonstrate equivalence is highest and to demonstrate equality is lowest.

The sample size estimation is challenging for complex designs such as non-inferiority or, time to event end points. Also, the sample size estimation needs adjustment in accommodating a) unplanned interim analysis b) planned interim analysis and c) adjustment for covariates.

In conclusion, the sample size is one of the critical steps in planning a CT and any negligence in its estimation may lead to rejection of an efficacious drug and an ineffective drug may get approval. As calculation of sample size depends on statistical concepts, it is desirable to consult an experienced statistician in estimation of this vital study parameter.

Acknowledgement:

I would like to thank Dr. Suresh Bowalekar, Managing Director PharmaNet Clinical Services Pvt. Ltd., for his guidance and inputs on writing this article.

References:

- 1 Shein-Shung C., Jun S. and Hansheng W. Sample Size Calculation in Clinical Trial. New York, Marcel Dekker Inc., 2003, Chapter 1, page 11, section 1.2.3.
- 2 ICH Topic E9, Statistical Principal for Clinical Trials Step 4, Consensus guideline, 05Feb1998 Note for Guidance On Statistical Principles For Clinical Trial. Available via <http://www.ich.org>.