The Importance of a Sample Size Calculation
by Stat One LLC

What is a Sample Size?

The sample size for a clinical trial is the number of subjects to be enrolled in the study. It is important for a study to include a sufficiently large sample size to obtain compelling or statistically significant results. However, a sample size that is larger than necessary wastes time and resources. For example, the study could produce a *statistically* significant result without producing a *clinically* significant result. This article provides background for the non-statistician to consider when working with a statistician to determine the appropriate sample size for a clinical trial.

Background Theory: Statistical Concepts for Sample Size Assessments

The objective of the clinical trial is to reject the null hypothesis (no difference between treatment groups) and show a statistically significant result (often represented by a p-value) demonstrating the results of the trial have a low probability of occurring due to random chance if there were no treatment effect. When drawing conclusions from a clinical trial, there are two types of errors that could potentially be made, referred to as Type I error and Type II error:

Type I error:

This error occurs if the null hypothesis is rejected when it should not be. If there is no actual difference between the treatment and control group but a difference is still found by the study, a type I error has occurred. If a type I error occurs, a company may advance the study of a particular treatment to later phase trials even though there is no treatment effect, thus wasting resources.

Type II error:

This error occurs if the null hypothesis is accepted incorrectly. If there is a true difference between the treatment and control group but the difference was not shown from the results of the study, a type II error has occurred. If a type II error occurs, a company may stop the development of a particular treatment even though that treatment has a clinically significant effect.

The chance of a Type I error is always strictly controlled in a clinical trial and is frequently referred to as the α-level of the study. The α-level or is explained by the definition of statistical significance for the primary analysis. Frequently the definition of statistical significance is based on observing a p-value of less than or equal to 0.05 on a two-sided test or a p-value less than or equal to 0.025 on a one-sided test.

Statisticians look to control the chance of committing a Type II errors through the use of sample size calculations. When performing sample size calculations, the statisticians use the term power or statistical power of the study. Statistical power is simple the chance of not committing a Type II error given a set of response assumptions (statistical power = 1 – the probability of a Type II error). A few points to understand about statistical power include:

- The FDA and other regulatory agencies, as a rule, require at least 80% power.
    - Note that 80% power means a 20% chance of failing to recognize statistical difference.
- Stat One recommends using a sample size the provides 90% power if possible.
    - 90% power means a 10% chance to fail by chance (half that of 80% power)
- For co-primary endpoints, typically each endpoint must be powered to at least 90% (0.9 x 0.9 = 0.81 so overall power in this case would be 81%).

Early phase studies are not typically powered in the manner that a pivotal study will be powered. However, study results may play into decisions about future product development. In that case, one typically considers if the sample is likely to give the correct trend if the desired treatment effect exists. Frequently, in medical device studies, there may only be a

single group of subjects getting an experimental treatment with the primary endpoint evaluated against a pre-determined performance goal.

Information Needed for a Sample Size Calculation

The information needed for a sample size calculation ideally takes into account a broad range of factors. Table 1 below summarizes some factors that should be considered when conducting a sample size evaluation.

Table 1, Input and Considerations for Sample Size Evaluation

Input	Description	Example
Study endpoint(s)	Outcomes being measured by a clinical trial	Lack of development of adverse events, increase in survival rate or time to events, shoulder pain improvement, lack of development of a rash, improved scores in a quality-of-life questionnaire, etc.
Study primary hypotheses	A conservative, tentative assumption about the population(s)	Zero difference in endpoint outcome between treatment group and control group, no more than X% lower in the treatment group than the control group, etc.
Statistical Analysis method	The chosen approach(es) to data collection and handling, particularly regarding the statistical tests that will be performed	Random controlled trial, balanced design, two treatment arms, questionnaire to determine pain scale score before and after treatment, two-sample t-test
Study α-level and Statistical Power	Error rates for a study (described above)	α-level is 0.05 on a two-sided test
Assumptions	Characteristics about the population that are expected to be true.	Treatment arm assignment is independent of other factors, pain improvement follows a normal sampling distribution, etc.
Clinically Meaningful Differences	A threshold value for which results would be considered meaningful to patients and clinicians.	It may not be meaningful to a patient if you told them that treatment A improves shoulder pain by 5% more than a placebo, but it may be meaningful for them to hear that it improves shoulder pain by 30% more than a placebo.
Other Considerations	Secondary details pertaining to logistics, feasibility, etc.	Desire for interim analyses, population size, ability to enroll subjects, number of sites, company ability to fund a study, time to complete study, patient drop-out or follow-up, importance of secondary endpoints.

An Example Study for Sample Size Calculations:

Let us consider a hypothetical study with a binary endpoint. Table 2 shows the input for consideration of sample size in our hypothetical study. Suppose we are primarily interested in whether subjects who receive a novel injection experience symptom relief more frequently than those who receive standard treatment. Binary endpoints are usually analyzed as proportions:

Table 2, Example Study Information

Input	Values
Study endpoint(s)	Symptom relief at 2 weeks following treatment
Study objective	Demonstrate the efficacy of the add-on treatment (injection) is relative to standard of care therapy.
Study primary hypotheses	The null hypothesis is that the response rate in the two arms is the same: $H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$ where p_1 and p_2 are response rates for active and placebo groups, respectively.
Study Design	Randomized controlled trial with 1:1 treatment allocation with active and placebo arms
Statistical Test	Chi-square test
Type I error rate	0.05 for a two-sided test
Desired Statistical Power	90%
Response Assumptions	The placebo group (standard of care) has a 60% response based on literature review.
Clinically Meaningful Differences	A minimum difference of at least 10% is considered meaningful by some physician experts. However, a difference of 15% may be required for insurance coverage and would lead to higher usage. Hence, the study will be powered based on a 75% response rate assumed for the active arm (60% vs 75%).
Other Considerations	No interim analysis, 5% of subjects will be lost to follow-up

Based on this information, a sample size of 428 was selected or 214 per arm. Details of the sample size calculation are as follows:

Using a 5% drop-out rate, 203 subjects per arm will be available for analysis. Assuming a 60% response rate in the control arm and a 75%, the sample size of 203 subjects per arm provides 90% power for the primary efficacy analysis.