Does Sample Size Calculation Begin and End in the Parameters Assignment in a Formula?

Diklah Geva, IntegriStat Ltd.

Diklah@IntegriStat.Com

November 2021

You have decided to engage in a study and defined the research objectives. So how much does it cost and how many study participants are required? In an initial conversation with the statistician, you were asked to bring all kinds of background data. She asked multiple questions that their relevance to determining the required size of the sample was ambiguous. Why can you not just plug some numbers in the formula and have an answer?

This paper surveys the factors affecting the determination concerning the sample size. Yes, There are quantifiable probability parameters. Other factors are not quantifiable yet will affect the sample size. In meeting with the researcher, the statistician clarifies the design factors that influence the selection of the formula for calculating the sample size. Afterward, she attempts to obtain from you information concerning the parameters to be assigned in the selected formula. Then, an initial numbers will be calculated. This number still is requiring various subsequent corrections. Now it is possible to determine the required sample size.

The statistician verifies the correlation between all the study variables and that it is feasible to test the research question in the proposed sampling scheme. The researcher should understand the considerations for determining the sample size, how the results will be presented if the research hypothesis is confirmed, and in case it is rejected.

At the end of this process, the statistician will provide a document to reason the sample size determination. A clinical trial protocol requires documentation of the considerations for determining the sample size, supported by demonstrating the sensitivity between the statistical power, minimal effect-size, and sample size, all while considering the context in which the research is conducted, as will be detailed further on.

Selecting the Formula for Calculating the Sample Size

Planning the Research is intended to determine the most efficient way to examine the research question. What is the measured variable (is it continuous or discrete, is it normal?) Which research design is the most efficient? Should there be comparisons between sub-groups? Or demonstrate a correlational relationship as a dose-response? Will there be repetitions in the study or follow-up over time? Are there co-factors? All these will affect the selection of the **statistical tools** for the data analysis of the results, and therefore, should fit the formula for calculating sample size.

For simple **statistical analyses**, there are simple formulas. However, any deviation from a simple experimental design, requires a more complex formula by applying advanced statistical methods. Often, there is no established " off the shelf" formula for example, if the expected distribution is abnormal, and in such cases, it is customary to use simulations. In other studies, an approximation to and establish formula is used while noting the deviations and estimating its direction.

The Sampling scheme can be critical. Undoubtedly, it is the most important point in statistical counseling. How the participants are recruited and how groups and repetitions are defined can determine if the data collected can address the research question. Statistical counseling is required in order to tailor a sampling scheme that allows the research question to be answered most efficiently and unambiguously.

A late referral to statistical counseling will not repair a flawed design. Studies with a poor sampling scheme generate data that does not enable answering the research question. The saddest thing in an erroneous sampling program is that even continued recruitment of participants will fail to correct for the inherent bias and rectify the situation. Applying statistical methods, as complex as they may be, will find it difficult to save the study.

The Parameters in a Sample Size Calculation Formula

The parameters taken into consideration in the formula are probabilistic: significance, power, and dispersion, and their size can be estimated only statistically. Concerning each parameter, it is customary to test their range and gain an impression from the sensitivity of the calculated sample size to a change in the parameter. A test like this is often called power analysis. It is used to reveal the range in which the sample size is very sensitive to these parameters, and generally graphically displayed to assist in determining the sample size. It is recommended to avoid a situation where a minute change in the parameters brings a large change in the sample size estimate. One should exhibit this sensitivity analysis as part of the document recording the considerations for sample size determination.

Statistical Significance (α) - is the probability of confirming the research question, although "it truthfully" would not require such confirmation. For example, we found a significant difference between a treatment group and a control group. However, when we repeated the experiment several more times, we discovered that the result did not repeat itself, and the initial result was an anomaly and not representative. It is customary to set this type of error (also known as a Type I error) at a probability of 5% or $\alpha = 0.05$. It means that statistically speaking, of every 100 experiments that we conducted, five (5) will result in a false positive.

Statistical power (1-β) - is a probability of confirming the research question when the research question should be confirmed. In effect, β (Type II error) is the probability *not* to confirm the research hypothesis when it should be "truly" confirmed. This error indicates a false negative between the treatment group and its control. It is customary to demand a statistical power of 80% - 95% (or, in other words, a Type II error β of 5%-20%). Intuitively, too small of sampling will have low power, *i.e.*, a low ability to confirm a research hypothesis, where it is "the correct truth". The same hypothesis will be confirmed if tested in a study with a greater number of subjects, *i.e.*, having a higher power.

Clinical effect-size (δ) - is the difference between the groups or the dose-response correlation that is be expected to be seen in the study. Sample size depends on the clinical effect size; identifying a large clinical effect demands a small sample. In contrast, identification of a small effect demands a larger sample which enables a greater sensitivity. For example, when we want to demonstrate that a diet plan assists in weight reduction, we will clarify what will be considered a positive result? Is it an average decrease of 3 kg for three months or an average decrease of 30 grams during the same period? It is customary to discuss this with the researcher and clinician and define in advance which effect-size is expected to be seen and whether it has clinical significance.

The degree of dispersion of the outcome variable (σ^2) - is the measurement for homogeneity of the study participants and the repetition of the outcome variable within each group. If there is a high dispersion, it is difficult to identify a significant difference between the groups, and therefore, a larger sample is required. The dispersion can stem from many factors: Precision of the measuring tools, associated factors influencing the measurement, and the degree of heterogeneity of the test subject group.

The classic formulas for calculating sample size relate solely to these four parameters. Generally, α and (1-β) are customarily determined according to α = 5% and (1-β) = 85%. While the degree of

dispersion σ and the effect-size δ are estimated from previous studies. Some express the effect size in terms of σ units and thus require sensitivity to a single unkown parameter. This cancel the need for a separate definitions of the effect-size and the outcomes' dispersion.

Corrections in the calculation

Now that the sample size, based on the appropriate formula and probability parameters, was calculated. However, there are still several practical issues that may be liable to force us to correct the initial number. For example, is it possible to recruit an equal number of test subjects for the study's two groups? Are there multiple comparisons? What is the dilution rate? And the like.

Multiple comparisons: If multiple comparisons are conducted in a study, and we are interested in examining more than one hypothesis, we must correct the significance level according to the number of comparisons. Let us assume that a certain study tests the treatment effect compared to control according to a decrease in weight and improved quality of life. A decrease in weight is measured in kilograms, reducing the Body Mass Index (BMI) and fat percentage. In addition, the quality of life is measured according to a depression index and participation in social activity. In this example, we have numbered approximately five (5) outcome measurements by which we will examine the treatment's effect. In this case, if every test is conducted at a significance of 5%, then basically, intuitively, the general significance level has risen to 25%, because the probability of a false positive when repeating the study 100 times is 5% for each measurement and a total of 25% for five independent outcomes.

In the case of multiple comparisons, we will activate a correction according to a rule of thumb stating to divide the overall significance - 5% by the number of comparisons. If five (5) comparisons are conducted, each comparison will be conducted at a significance level of 1%, so that the overall significance level of the entire study is preserved at 5%. Of course, this causes a significant increase in the sample size. Therefore, allocating the total significance to a single primary endpoint is often practiced, and the remaining outcome measurements in the study are set

as the secondary endpoints. Regulatory authorities are careful to examine that the endpoint is noted a-priori in a protocol, and that the statistical methods match the study's design.

Attrition due to dropouts - In many studies, not all the participants will conclude their participation in the study, for various reasons. In such a case, there is a concern that the number will become smaller and lose its required statistical power. This concern may be addressed by revising the sample size for expected dropouts' rate. It is generally accepted to assume 10%-25% dropout, but the dropout rate differs greatly between the various disciplines, and it is necessary to estimate the compatible attrition rate. Any reference to the subject of attrition during follow-up adds reliability to a study. The longer the follow-up the higher is the expected attrition rate.

Statistical analysis should also refer to attrition: a comparison will be made between those dropuout of the study and those who concluded the study, and it will be tested to determine whether attrition rate is identical in both the treatment and control groups. For example, a case where the treatment itself causes attrition and only a small group completed the treatment-protocol until its conclusion. This may also lead to a bias due to the unbalanced treatments between the groups. The response here is primarily at the data-analysis level: Attention must also be given to attrition in the analysis were the final results are verified not only on the complete participants but also in the intent to treat participants (i.e., all cases recruited to the study). Therefore, during data collection it is required to document the drop-outs in the same manner in which the completers are documented.

The Research Context:

The context in which the present study is being conducted and previous studies in the specific field will significantly determine the sample size.

The Objective of the Research: We distinguish here between Exploratory or Confirmatory research. Is the research intended to confirm results and assumptions that have already been tested in previous studies, or is it exploratory research to test feasibility? In exploratory research, we do not know, for the most part, the precise measure of the dispersion of the outcome variable, and we do not correct for multiple comparisons, understanding that we want to receive more results and therefore compromise reliability. In confirmatory research, we must verify all the preliminary details and maintain the entire significance level on the primary research question.

The Research Agenda: In the context of clinical studies, we will examine what phase the study is at (Phase I, II, III, IV). It is possible to suffice with a smaller sample in the earlier phases. It is worthwhile checking at the same time whether there are similar studies that can be combined. We will clarify whether the goal is academic publication or a research requirement for approval by health and regulatory authorities. All these factors are considered in the research design, the sampling scheme, and in turn, are determining the sample size.

What is the common practice in the field: Ultimately, the medical field in which the research is conducted will determine the accepted framework. For the most part, studies engaged in complex surgical operations will see smaller samples of 10-15 in each group.

In studies of rare diseases, too, large samples are not possible. We will see larger samples in studies where it is easier to gather extensive information without invasive intervention. Research areas with many intervening variables and heterogeneous groups require the largest samples, sometimes around 100-200 group subjects. The researcher must be familiar with the field in which he works and examine the sample sizes in similar studies published in the professional literature. Large scale studies with samples of ten of thousand participants, usually focus on testing a treatment under heterogeneous conditions. They hypothesize for moderate effect size while being able to report on even rare side effects.

The Budget: Above all, the budget framework and time frame determine the research design and size. In a limited budget, we must limit the scope of the research question. For example, it is possible to control dispersion by concentrating a single homogeneous subgroup and by selecting and accurate tools to evaluate the study endpoint.

Discussion

We have reviewed the process for determining the sample size and the factors involved here.

The statistician checks the optimal research design with the researcher when all the factors mentioned above are taken under consideration. This review should coordinate expectations between the research context, budget, and possible results. First and foremost, the statistician verifies that it is possible to answer the research question using statistical tools. In certain cases, the conclusion might be that the experiment should not start at all. In most cases, the researcher will better understand what he expects from the experiment, what scientific weight the results of his experiment will bear, and what are the study's limitations. When planning a new study, statistical counseling serves as a sanity check and it is meant to produce the sample size considerations document.

 The author may be contacted at diklah@integristat.com