Sample size in health research

Ari Samaranayaka, Claire Cameron, Robin M. Turner

What is sample size and why does it matter?
How big should my sample be? Will I have enough statistical power? How much precision will my estimates have? These are important questions asked by researchers while designing a study. When the study is over, the researcher may ask: why was my result significant, or why was it not significant? Could it be related to the sample size achieved in the study? This leads to the critical questions; why is sample size important, and when should it be considered in the research process?

Loosely speaking, an a-priori sample size estimate is the optimum number of items (usually individuals) that should be included in the study in order to answer the research question.

Sample size determination is an integral part of the research process
There are two types of analyses where the sample size estimate is pivotal in the design of the study:
1. A study where you wish to estimate a quantity with a certain level of precision. Prevalence studies are a common example of this where you may want to estimate the proportion of people in the population with a particular disease.
2. A study where you are testing a hypothesis, i.e. when you will be using a P value (see our Statistics Primer in issue 31). One of the most common examples of this would be when comparing two groups to see if they have different effect estimates (e.g. means or proportions). We often call this difference the “effect size”.

The process of deciding on the statistical analysis plan requires the best guess at which differences/variabilities are expected in the outcome measure. This often leads to refinement of the research question (see Figure 1) as the outcome measure and study design have to be precisely defined: they directly influence the analysis plan and the sample size calculations. In practice, the initial calculations can produce a sample size that cannot be achieved within practical limitations. In such cases, changes need to be made (in consultation with content experts) to the analysis plan, and even to the research aims, so that an achievable sample size estimate can be obtained. This iterative process is the norm in the development of most studies. The determination of sample size cannot be seen in isolation; it is an integral part of the overall study design process.

The approach taken for a particular study depends entirely on whether the research question, and subsequent analysis plan, are about an estimation (Approach 1) or testing a hypothesis (Approach 2). Therefore, sample size determination cannot be separated from the other components of the design process.

These two approaches are illustrated below, using simple study designs as examples. Unlike other disciplines where the experimental environment can be tightly controlled, health-related studies often have to use observational study designs. This means that we may need to account for potential confounders in the analysis. Other complexities can sometimes arise, such as needing repeated measurements from each individual, meaning that the measurements are no longer independent (which is one of the assumptions underpinning many standard statistical methods). The analysis plan will need to be tailored correspondingly to account for situations like these. In some instances, these adaptations can lead to complex analyses, which require customised approaches to sample size determination. These customised approaches may require simulations or other complex methods. We have chosen simple study designs below to illustrate the application of these two approaches.

Approach 1
To illustrate the precision-based approach, consider a survey aiming to estimate the prevalence of a chronic health condition, such as di-

Figure 1: Place of sample size calculation within the research process

abates. For simplicity, assume that we have a simple random sample, meaning that every person in the population has an equal chance of being selected for the survey. Suppose that, based on our knowledge of the literature, we expect the prevalence of this disease to be approximately 0.2 (20%). We want the 95% confidence interval around our estimated prevalence to be no wider than 0.16 to 0.24, i.e. no wider than 0.20 ± 0.04. We call this ‘0.04’ the half width of the 95% confidence interval. This is also known as the ‘margin of error’. Under this study design, the confidence interval can be estimated as

\[ P \pm Z \sqrt{\frac{P(1-P)}{N}} \]

In this equation, \( P \) is the prevalence estimate (0.20 in this case), \( Z \) is the value from the standard normal distribution corresponding to 95% (which is 1.96), and \( N \) is the sample study size. The precision of our estimate is determined by the half-width of the 95% confidence interval (the quantity on the right-hand side of ± sign). We need to find a sample size \( N \) that makes this quantity = 0.04 or lower. Putting this information together gives the following expression:

\[ 0.04 \geq Z \sqrt{\frac{P(1-P)}{N}} \]

This can be rearranged to determine the sample size:

\[ N \geq \frac{Z^2 \cdot P(1-P)}{0.04^2} \]

This becomes \( N \geq 385 \) for our example where \( Z=1.96 \) and \( P=0.20 \). Note that we always round up to the next integer when estimating sample size (for both approaches). This is because if we round down, our sample size will be slightly too small, and our confidence interval will be slightly wider than required.

The estimation of the confidence interval is not only dependent on the sample size, but also on our initial guess of the prevalence (\( P \)). However, given the aim of this study was to estimate the prevalence (\( P \)) in the first place, we might not know what value of \( P \) to use in order to include it in the sample size estimation. What can we do if we don’t have an approximate value? Prevalence must lie between zero (no one has the disease) and one (everyone has the disease). It is helpful to know that the quantity \( P(1-P) \) in our equation takes on a maximum value when \( P=0.5 \), and a minimum when \( P=0 \) or \( P=1 \) (pull out your calculator and check this for varying values of \( P \)). Therefore, the sample size corresponding to \( P=0.5 \) is the largest and the most conservative sample size (i.e. our sample will always be large enough regardless of the prevalence value). However, if the required precision changes (i.e. the half-width is bigger or smaller), then the sample size will also change.

**Approach 2**

This approach can be more challenging, due to the fact that the sample size estimates are particular to the hypothesis test that you have decided will answer your research question. There are a number of commonly used simple hypothesis tests, including the independent samples t-test and the paired t-test for investigating the difference between two means, and the binomial test for investigating the difference between two proportions.

Statistical power is a concept that arises out of hypothesis testing. With a hypothesis test, we are trying to make an inference about the population while looking at the results of a study. There are two errors that can be made when using hypothesis tests. You could, for example, incorrectly conclude that there is a difference between the means of two groups when there really is no difference in the population. Alternatively, you could incorrectly state that there is no difference in those means when, in fact, there is a real difference in the population. This second type of error is related to “statistical power”: the power of a hypothesis test is the probability of finding an effect (i.e. the difference between two groups), if this really exists. When you are running a study, you want this probability to be high, to ensure that you do not miss a real effect, if it exists. The size of your sample, along with the variability in the measures you are using, directly impacts the amount of statistical power that your study will have.

Formulae used in every statistical hypothesis test can be rearranged to estimate sample size, based on the assumptions about the power, variability, and the difference you are trying to detect. In order to estimate the sample size, you need to have a good understanding of the difference that you are trying to detect. This difference should be the minimum clinically important difference (also known as the “effect size”). This is one of the more difficult parts to guess at when initially designing the study, and has a critical impact on the estimated sample size. Guessing an appropriate value for this is not a biostatistics question. For example, a study could be designed to detect a 1 mm-Hg difference in systolic blood pressure between two groups. Is this 1 mm-Hg difference an important difference clinically? What about a 5 mm-Hg difference? If the study was powered to detect a 15 mm-Hg difference but missed a 10 mm-Hg difference, is that a problem? These are the questions to be considered by content experts when looking at the minimum clinically important difference. Other critical information needed when estimating sample size is the expected variability of the measure. If an approximate value for this cannot be found from the literature, then pilot studies can be useful not only to test out all aspects of the design, but to inform future sample size estimates for the main study.

There are formulae available to calculate the sample size for many commonly used hypothesis tests. Alternatively, there are some good online calculators that exist, although we suggest that you check which assumptions they make, and that they reference the particular formulae they are using. Using a calculator is a good way to understand how the minimum clinically important difference (or other measures of interest, such as variability or the desired level of power) may impact on sample size for an assumed significance level. The significance level (usually set at 5%) is the level below which a \( P \) value will be considered “significant”, although we do not advise significance be based on a \( P \) value alone.

If you change the following values one at a time, you will notice that the required sample size is large:

1. When the effect size of interest is small
2. If the variability is large
3. If higher statistical power is needed
4. If the selected significance level is smaller

The estimated sample size (from both approaches) then needs to be scaled up to account for things like participation rates and loss to follow-up, otherwise the final analysis will not have sufficient participants to meet the sample size requirements, and the study will be considered under-powered.

Sample size calculations are not applicable for some studies. Pilot or feasibility studies, proof of concept studies, and clinical audits do not aim to estimate a quantity with a given precision, or to test something with specific power; therefore, a formal determination of sample size is not applicable. Even though formal sample size calculations may not be required, it is still important to consider whether the sample size is big enough for the desired purpose; is the pilot study large enough to be able to pilot all features of the trial and obtain reasonable information to inform the full trial in future?

Sometimes researchers decide that it is not possible, for practical reasons, to use the sample size appropriate to answer their research question. We sometimes see statements like, “Our non-significant results could be due to our small sample size.” The problems with this statement are (1) it assumes that they would have obtained a significant result if they had had a larger sample when there is no evidence to support that claim, and (2) it attempts to justify running an underpowered study. In order to support such statements, some researchers calculate the statistical power of the already conducted study to show the insufficiency of the power (this is called a post hoc
power calculation). No post hoc power calculation is acceptable, nor can it be used as a substitute for a sample size calculation in the initial stages of designing and planning a study.

In summary, sample size estimation is an educated guess to ensure that the effect of interest is likely to be detected if it really exists, or that the precision will be of a certain magnitude. The sample size calculation is therefore an integral aspect of the research study design process, which cannot be viewed in isolation. The process of estimating sample size helps clarify the research question and the analysis plan. It requires input from content experts, and cannot be done purely by a biostatistician without a deep understanding of the design, analysis plan, and outcome measures. In reality, sample size estimates can be complicated by many factors, such as contamination of interventions, non-independence of participants, longitudinal outcome measures, design effects, and covariate adjustments, to name but a few. Obtaining professional input in the initial stages of the design process is highly recommended in these complicated situations, preferably by including a biostatistician as a member of your research team.

References

About the authors
› Dr Ari Samaranayaka, BSc, MPhil, PhD, is a Senior Research Fellow and Biostatistician in the Biostatistics Centre, Division of Health Sciences, University of Otago
› Dr Claire Cameron, BSc(Hons), DipGrad, MSc, PhD, is a Senior Research Fellow and Biostatistician in the Biostatistics Centre, Division of Health Sciences, University of Otago
› Associate Professor Robin Turner, BSc(Hons), MBiostat, PhD, is the Director of the Biostatistics Centre, Division of Health Sciences, University of Otago

Correspondence
Dr Ari Samaranayaka: arisamaranayaka@otago.ac.nz