Abstract
Determining the appropriate sample size for a study, whatever be its type, is a fundamental aspect of biomedical research. An adequate sample ensures that the study will yield reliable information, regardless of whether the data ultimately suggests a clinically important difference between the interventions or elements being studied. The probability of Type 1 and Type 2 errors, the expected variance in the sample and the effect size are the essential determinants of sample size in interventional studies. Any method for deriving a conclusion from experimental data carries with it some risk of drawing a false conclusion. Two types of false conclusion may occur, called Type 1 and Type 2 errors, whose probabilities are denoted by the symbols $\sigma$ and $\beta$. A Type 1 error occurs when one concludes that a difference exists between the groups being compared when, in reality, it does not. This is akin to a false positive result. A Type 2 error occurs when one concludes that difference does not exist when, in reality, a difference does exist, and it is equal to or larger than the effect size defined by the alternative to the null hypothesis. This may be viewed as a false negative result. When considering the risk of Type 2 error, it is more intuitive to think in terms of power of the study or $(1 - \beta)$. Power denotes the probability of detecting a difference when a difference does exist between the groups being compared. Smaller $\alpha$ or larger power will increase sample size. Conventional acceptable values for power and $\alpha$ are 80% or above and 5% or below, respectively, when calculating sample size. Increasing variance in the sample tends to increase the sample size required to achieve a given power level. The effect size is the smallest clinically important difference that is sought to be detected and, rather than statistical convention, is a matter of past experience and clinical judgment. Larger samples are required if smaller differences are to be detected. Although the principles are long known, historically, sample size determination has been difficult, because of relatively complex mathematical considerations and numerous different formulas. However, of late, there has been remarkable improvement in the availability, capability, and user-friendliness of power and sample size determination software. Many can execute routines for determination of sample size and power for a wide variety of research designs and statistical tests. With the drudgery of mathematical calculation gone, researchers must now concentrate on determining appropriate sample size and achieving these targets, so that study conclusions can be accepted as meaningful.

Key Words: Effect size, power, sample size, Type 1 error, Type 2 error

Introduction
In most areas in life, it is very difficult to work with populations. During general elections, for instance, a newspaper is likely interview a few thousand people at the most and predict results based on their responses. Similarly, in a factory that manufactures light bulbs, a few bulbs are chosen at random to assess the quality of the entire batch of light bulbs. The inherent difficulty in working with populations makes researchers chose samples to work with. These difficulties include cost, time considerations, logistics, and also ethics as it is usually unethical to study an entire population when...
a subset of the population could answer the research question.

Studies that cover entire populations go by the generic term “census.” In India, we have a decennial census that is held once in every 10 years. This is only a demographic and socioeconomic census that aims to capture data on a limited range of demographic, social, and economic indicators. However, each and every Indian citizen has to be covered. This makes it a huge exercise and necessitates that the Government of India maintain an elaborate machinery called the Office of the Registrar General and Census Commissioner under the Ministry of Home Affairs (popularly called the Census Bureau). This census, though it aims to capture only limited data, is such an elaborate affair, that by the time all the data have been collected, collated, processed and analyzed, it is almost time for the next census. Hence, what will the Indian Government do if it requires some quick answers, such as the immunization coverage in a particular district or the malnutrition prevalence in a particular region? For this, the government has to maintain another machinery called the National Sample Survey Organization (NSSO), now known as National Sample Survey Office, under the Ministry of Statistics. The NSSO conducts periodic surveys, not of the entire country’s population but of a sample of randomly selected “NSSO blocks,” to provide answers that are generally available in a 2–3 year timeframe. It routinely collects data on several socioeconomic, health, industrial, agricultural, and price indicators.

Most biomedical researchers will never have the luxury of conducting a census but will have to depend on a population subset, called the sample, to seek reasonably valid answers to their research questions. Look at Figure 1 where the ellipse represents a population. Suppose, a researcher draws a sample represented by the first circle (Sample 1) to answer a research question. Another researcher may be trying to answer the same research question using another sample (Sample 2) represented by the second circle. Obviously, the two circles vary in their size (radius) and location (center). The situation of the first researcher is akin to a group studying the situation in a particular region of India, while the second may be that of a group with better manpower and funding who are trying to cover all the regions of the country. It stands to logic that we will accept the results of the second group as representative of the entire country. We might also accept the results of the first group as representative of the entire country if there is not much reason to suspect there could be large regional differences. However, it is unlikely that we will accept the results of a third group, who have sampled only a tiny circle, say of a single town (Sample 3) as representing the entire country. Thus, whenever we work with samples rather than populations, it is important to ensure that the sample is optimally sized and adequately representative of the entire population. It does not matter what type of study we are doing; a clinical trial, laboratory experiment, field survey, or quality control – everywhere these two issues – that of sample size and sampling – are of paramount importance. If we do not get these right, our sample results are not generalizable to the population we have intended to study, and all our efforts will go in vain.

“How much is enough?” is often a question that plagues researchers and clinicians alike. While sample size calculations immediately bring to mind complex formulas, the aim of this module is not to present a pantheon of fearsome formulas, but rather to familiarize the reader with the principles though a few representative formulas with solved examples. Estimation of the minimum sample size required for any study can have technical variations, but the concepts underlying most methods are similar. These concepts are important as they enable researchers to use a minimum number of subjects to draw strong (valid and robust) conclusions with a limited number of research participants. It is also important to remember that whatever the formulas used, small differences in selected parameters (described below) can lead to large differences in the calculated sample size. Thus, any sample size calculation, however carefully done, will always remain approximate. In most studies, there is a primary question that the researcher wishes to answer. Sample size calculations are based on this primary objective. Finally, before locking the sample size to work with, one must take into account available funding, manpower, logistics, and, most important of all in clinical studies, the ethics of subjecting human participants to potentially harmful interventions.

**Elements in Sample Size Calculation for Randomized controlled trials - An Understanding of Key Concepts**

Determining the sample size required to answer the research question is one of the first steps in designing a study. The main aim of sample size calculation for a clinical trial is to determine the number of participants.
needed to detect a clinically relevant treatment effect. As a general rule, the greater the variability in the outcome variable, the larger the sample size needed to assess whether the observed effect (that seen when the study is completed) is a true effect.

Here, we will discuss principles of sample size calculation for two group randomized controlled trials (RCTs). The calculation of sample size for RCTs requires that an investigator specify certain factors outlined below. Broadly, as we have seen earlier in this series, data can be categorized as numerical (quantitative) and categorical (qualitative) data. For the former, information on the mean responses in the two group, \( \mu_1 \) and \( \mu_2 \), are required as also the standard deviations (SDs) or a common (pooled) SD for the two groups. For categorical data, information on proportions of successes in the two groups, \( p_1 \) and \( p_2 \), is needed. Such information is usually obtained either from published literature, a pilot study or at times “guesstimated.” The other two key components are the Type 1 (alpha) and Type 2 (beta) error probabilities. Apart from this, an understanding of whether the data is normally distributed (following the Gaussian curve) or otherwise is required. Moreover, an understanding of the null hypothesis is useful at this stage. The null hypothesis states that the two groups that are being compared are not different while the alternate hypothesis would be that the two groups are actually different.

The four elements that enter into sample size calculation formulas:

- **Effect size (d or \( \delta \))**: The size of the effect that is clinically worthwhile to detect, that is the smallest difference that is clinically meaningful. For numerical data, this is the difference between \( \mu_1 \) and \( \mu_2 \); that is the anticipated outcome means in the two groups, while for categorical data, this is represented by \( p_1 \) and \( p_2 \), that is the proportions of successful outcomes in the two groups. The effect size represents a clinically meaningful difference in the sense that it may make the physician change his practice. As stated earlier, choice of this clinically meaningful difference can be based on existing literature or a pilot study. In case of numerical data, the ratio of effect size to SD has been called “standardized difference” or the “standardized effect”.

- **Type 1 error**: The probability of falsely rejecting a null hypothesis when it is true (\( \alpha \)). This represents the false positive error and can be regarded as the probability of finding a difference between two groups where none exists. It is also called the regulator’s error as regulatory decision making takes place based on results of these comparisons. Note that the \( \alpha \) error is akin to the significance level of a study. Conventionally, the value of \( \alpha \) is set at 5% or lower. The lower the value, the larger would be the sample size.

- **Type 2 error**: The probability of failing to reject a false null hypothesis (\( \beta \)). This represents the false negative error and is the probability of not finding a difference between the two groups studied when one actually exists. It is also called the investigator’s error. The value of \( \beta \) is conventionally set at 20% or lower. The lower the value, the larger would be the sample size. Since \( \beta \) error is the inability to detect a difference, it follows the \( (1 - \beta) \) is the ability to detect a difference should one actually exist and is referred to as the “power” of a study. A study must have at least 80% power to detect a difference. A \( \beta \) value of 10% will confer 90% power to the study.

Note again that the relationship between \( \alpha \) and \( \beta \) are reciprocal. If one tries to lower \( \alpha \), the value of \( \beta \) will go up, unless one expands the sample size. The only way to achieve zero \( \alpha \) and zero \( \beta \) errors is to work with an infinite sample size, which is not possible.

Selection of \( \alpha \) and \( \beta \) error values, will in turn lead to selection of the standard normal deviates, \( Z_\alpha \) and \( Z_\beta \) values, that are actually entered into the formulas (Table 1). The formulas incorporate a factor \( (Z_\alpha + Z_\beta)^2 \), which has been referred to as the power index.

- **Standard deviation (SD)** of the outcome measure of interest in the underlying population (SD or \( \sigma \)). This is the variability or spread associated with quantitative data. The larger the variability, the larger would be the sample size required to attain a given power at the chosen level of significance. In many cases, although the SD is not exactly known, one has a rough estimate and the sample size may be calculated based on the maximum variance that is likely. If the variance in the observed data is smaller, the study will attain higher power. If the SD is completely unknown, the solution is to conduct a pilot study to obtain an estimate of the variance of the outcome measure.

Although it is often assumed that study groups would have the same variance, this is not always the case. If the variance of the outcome measure in question varies widely between the different groups, a “pooled” SD value has been used. This can be calculated for two groups as:

| Table 1: Commonly used standard normal deviate values used in sample size calculations |
|-----------------------------------------------|
| **Direction of testing** | **\( \alpha \) or \( \beta \)** | **Value** |
| \( Z_\alpha \) | Two-sided | \( \alpha=0.05 \) | \( Z_\alpha=1.960 \) |
| \( Z_\alpha \) | Two-sided | \( \alpha=0.025 \) | \( Z_\alpha=2.326 \) |
| \( Z_\alpha \) | Two-sided | \( \alpha=0.01 \) | \( Z_\alpha=2.576 \) |
| \( Z_\alpha \) | One-sided | \( \alpha=0.05 \) | \( Z_\alpha=1.645 \) |
| \( Z_\alpha \) | One-sided | \( \alpha=0.025 \) | \( Z_\alpha=1.960 \) |
| \( Z_\alpha \) | One-sided | \( \alpha=0.01 \) | \( Z_\alpha=2.326 \) |
| \( Z_\beta \) | \( \beta=0.20 \) | \( Z_\beta=0.840 \) |
| \( Z_\beta \) | \( \beta=0.10 \) | \( Z_\beta=1.282 \) |
\[ SD_{\text{pooled}} = \sqrt{\left( \frac{SD_1^2 + SD_2^2}{2} \right)} \]

Note that it is also important to decide whether one should do two-tailed or one-tailed testing. In most situations, two-tailed testing is the norm, although it requires a larger sample size. One-tailed testing is acceptable only if one can be sure that change or difference can only be in one direction and not in either direction.

**A Few Solved Examples Based on Formulas**

Although sample size tables can provide useful guides for determining sample size, they are based on commonly used values for different parameters, and one may need to calculate the necessary sample size for different combinations of Type 1 error probability, power, and variability. Formulas are then needed to obtain exact sample size estimations for such situations. There are umpteen formulas to suit various study designs. We will consider few common examples here.

**Sample size for one mean, normal distribution**

\[ N = \frac{(Z_{\alpha} + Z_{\beta})^2 \times \sigma^2}{\delta^2} \]

Problem 1: A physician wants to know if the mean heart rate after use of a new drug differs from the healthy population rate of 72 beats/min. He considers a mean difference of 6 beats/min to be clinically meaningful. He also chooses 9.1 beats/min as the variation expected based on a previously published study. How many patients will be needed to carry out the study with 5% probability of Type 1 error and 80% power?

In this example, we have:

- Z value (two-sided) related to the probability of falsely rejecting a true null hypothesis \((\alpha) = 0.05\) \(Z_{\alpha} = 1.96\)
- Z value related to the probability of failing to reject a false null hypothesis \((\beta) = 0.80 Z_{\beta} = 0.84\)
- SD of the outcome being studied \((SD \ or \ \sigma) = 9.1\) beats/min
- Size of the effect that is clinically worthwhile to detect \(\delta = 6\) beats/min.

Substituting in the formula, we get:

\[ N = \frac{(1.96 + 0.84)^2 \times 9.1^2}{6^2} = 18.03 \]

Thus, 18 patients are needed to be studied with the new drug.

**Sample size for two means, quantitative data**

\[ N = \frac{(Z_{\alpha} + Z_{\beta})^2 \times \sigma^2}{\delta^2} \]

Where \(\delta = (\mu_1 - \mu_2)/2\).

Problem 2: Channick et al. studied the effects of the dual endothelin-receptor antagonist bosentan in patients with pulmonary hypertension in a randomised placebo-controlled trial. (Channick RN, Simonneau G, Sitbon O, Robbins IM, Frost A, Tapson VF, et al. Effects of the dual endothelin-receptor antagonist bosentan in patients with pulmonary hypertension: A randomised placebo-controlled study. Lancet 2001;358:1119-23). They proposed to detect a mean difference of 50 m in the 6-min walk test given a common SD of 50 m between the two groups at 80% power and 5% significance using one-sided test. The anticipated dropout rate was 25%. How many patients did they require.

In this example, we have:

- Z value (two-sided) related to probability of falsely rejecting a true null hypothesis \((\alpha) = 0.05 Z_{\alpha} = 1.65\) for a one-sided test
- Z value related to probability of failing to reject a false null hypothesis \((\beta) = 0.80 Z_{\beta} = 0.84\)
- SD of the outcome being studied \((SD \ or \ \sigma) = 50\ m\)
- Size of the effect that is clinically worthwhile to detect \(\delta = 50\ m\).

Substituting in the formula, we get:

\[ N = \frac{(1.645 + 0.84 \times 50)^2}{25^2} = 24.7 \]

Thus, 24 patients needed to be randomized to the two study groups. To allow for the anticipated dropout rate, the recruitment target was 32. In this example, the authors have used a one-sided test as the outcome of interest is the 6-min walk test and one does not expect the placebo to worsen the baseline distance the patient can walk. However, one must be careful though in the choice of the one-sided test. It is to be used only when the outcome of interest truly can be in one direction only.

Note that an alternative version of the formula given above is:

\[ N = 2 \times (Z_{\alpha} + Z_{\beta})^2 \times \frac{\sigma_1^2 + \sigma_2^2}{\delta^2} \]

In this version, we can accommodate two separate SD values for the two groups and \(\delta\) remains simply as \((\mu_1 - \mu_2)\).

**Sample size for two proportions, categorical data**

\[ N = \frac{4 \times (Z_{\alpha} + Z_{\beta})^2}{\delta^2} \]

Where \(\delta = (p_1 - p_2)/\sqrt{(p_1(1-p_1)+p_2(1-p_2))/2}\).

Problem 3: In the bypass versus angioplasty in severe ischemia of the leg study on bypass versus angioplasty for leg ischemia (Adam DJ, Beard JD, Cleveland T, Bell J, Bradbury AW, Forbes JF, et al. Bypass versus angioplasty in severe ischaemia of the leg [BASIL]: Multicentre, randomised controlled trial. Lancet 2005;366:1925-34.). The statistical calculations were based on the 3-year
survival value of 50% in the angioplasty and 65% in the bypass group. At 5% significance and 90% power, how many patients would be needed to detect a difference between the two groups?

In this example, we have:

- Z value (two-sided) related to the probability of falsely rejecting a true null hypothesis ($\alpha$) = 0.05 $Z_\alpha = 1.96$
- Z value related to the probability of failing to reject a false null hypothesis ($\beta$) = 0.90 $Z_\beta = 1.282$
- Proportion of success in one group ($p_1$) = 0.65
- Proportion of success in the other group ($p_2$) = 0.50.

Therefore, $\delta$ will be $(0.65 - 0.50)/\sqrt{(0.575 \times 0.425)} = 0.15/0.49434 = 0.30343$.

Substituting in the formula, we get:

$$N = \frac{4 \times (1.96 + 1.282)^2}{(0.30343)^2} = 456.63$$

Thus, 456 patients overall or 228 per group are needed (assuming groups to be of equal size) to conduct the study.

Note that an alternative version of the formula given above is:

$$N = 2 \times (Z_\alpha + Z_\beta)^2 \times \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2}$$

In this version, we can directly accommodate the two proportions.

**Alternatives to General Formulas**

Beyond working with formulas by hand, sample size estimations can also be done using tables based on these general formulas, quick formulas, graphical methods (nomograms), and software.

**Quick formulas**

Various simplified versions of the general formulas have been proposed to enable rapid sample size calculations for standard situations. For instance, Lehr’s formula can be used for quick calculation of sample size for comparison of two equal-sized groups for power of 80% and a two-sided significance level of 0.05.

The required size of each equal sized group is $16/($standardized difference)$^2$.

For power of 90%, numerator becomes 21.

However, if the standardized difference is small, this formula tends to overestimate sample size. In comparison to the general formula presented above.

Recall that standardized difference is the ratio of effect size to SD. For unpaired t-tests, the standardized difference is calculated simply as $\delta/\sigma$. Jacob Cohen in his 1988 book (Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed., Hillsdale, NJ: Lawrence Earlbaum Associates; 1988) dealt elaborately with effect sizes and standardized differences and proposed thumb rules for these. They have been widely used in psychology, but many feel that predetermined standardized differences are too restrictive.

**Nomograms**

Nomograms offer a graphical alternative method of sample size calculation. They are cleverly designed on the basis of general formulas. The Altman's nomogram, devised by Doug Altman and published in his 1991 book (Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991) is a popular graphical method that enables sample size estimations for paired and unpaired t-tests and the Chi-square test. You can easily download a copy from the internet and try it. To use the nomogram, we first need to translate our required difference into a standardized difference.

Various other nomograms have been devised. For studies of diagnostic tests, Malhotra and Indrayan have proposed a convenient nomogram for sample size based on anticipated sensitivity/specificity, and estimated prevalence (Malhotra RK, Indrayan A. A simple nomogram for sample size for estimating sensitivity and specificity of medical tests. Indian J Ophthalmol 2010;58:519-52). The Schoenfeld and Richter nomograms give sample size for detecting difference in median survival between two treatment groups in survival analysis (Schoenfeld DA, Richter JR. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. Biometrics 1982;38:163-70). There are other nomograms for other study designs.

**Software**

An understanding of the principles helps in inputting the desired information into software and quickly getting the calculations done. Many softwares are available that provide sample size calculation routines. Some come as part of larger statistical packages while some are standalone power and sample size software. Power and Sample Size Calculation (PS) is a free software (current version 3.1.2; 2014) developed by the Department of Biostatistics, Vanderbilt University, USA, that provides sample size routines for both interventional and observational studies. It can be downloaded (http://biostat.mc.vanderbilt.edu/wiki/main/powersamplesize), installed and used without restrictions. Power Analysis and Sample Size Software (PASS) is a comprehensive commercial package developed by NCSS, a company based in Kayesville, Utah, USA. It is regarded as an industry standard and provides sample size calculations for over 650 statistical tests and confidence interval (CI) scenarios (http://www.ncss.com/software/pass/). nMaster (current version 2.0; 2011), developed by the Department of Biostatistics, Christian Medical College, Vellore, India, is a very affordable...
package that also provides a large number of sample size routines required for most academic studies. It is important to remember, however, that the sample size calculations are prone to errors and, even with software, it is usually a good practice to get these calculations verified by someone experienced in the field before embarking upon the study.

**Adjustments to Calculated Sample Size**

Once a “raw” sample size is calculated, various adjustments may be needed to accommodate variations in study objectives and to keep an adequate safety margin for potential attritions. Thus, adjustments may be needed for multiple outcomes, unequal group sizes, dropouts, planned subgroup analysis, cluster sampling, and so on. For instance, it is important that we calculate sample size on the basis of a single primary outcome. If we have multiple primary outcomes, there is no other way out than to calculate sample sizes separately for each of these outcomes and then work with the largest so that the rest are covered as well. Other common adjustments are discussed below.

**Adjusting for Unequal Sized Groups**

The methods described above assume that the comparison is to be made across two equal sized groups. However, this may not be the case in practice, for example in an observational study or in an RCT with unequal randomization. In this case, it is possible to adjust the numbers to reflect this inequality. The first step is to calculate the total (across both groups) sample size assuming that the groups are equal sized. This total sample size \( N \) can then be adjusted according to the actual ratio of the two groups \( k \) with the revised total sample size \( N' \) being:

\[
N' = \frac{N (1 + k)^2}{4k}
\]

For instance, consider that a placebo-controlled trial requires a total sample size of 120 if the two groups are of equal size. If it is decided that twice as many subjects would be randomized to the active treatment group than to the placebo group, the new sample size \( N' \) would be:

\[
N' = \frac{120 \times (1 + 2)^2}{4 \times 2} = \frac{120 \times 9}{8} = 135
\]

Thus, 135/3 = 45 patients need to be allocated to placebo treatment and 135 × 2/3 = 90 to active treatment groups.

**Adjusting for Consent Refusals, Withdrawals, and Missing Data**

Any sample size calculation is based on the total number of participants who are needed in the final analysis. In practice, eligible subjects will not always be willing or continue to take part, and it will be necessary to approach more subjects than are needed in the first instance. In addition, even in the very best designed and conducted studies, it is unusual to finish with a data set in which complete data are available for every subject. Subjects may fail to turn up, refuse to be examined or their samples may be lost. In studies involving long follow-up, there will always be a substantial degree of attrition. It is therefore often necessary to estimate the number of subjects that need to be approached to achieve the final desired sample size.

The adjustment may be done as follows. If a total of \( N \) subjects is required in the final study, but a proportion \( q \) are expected to refuse to participate or to drop out before the study ends, then the total number of subjects \( N' \) who would have to be approached at the outset to ensure that the final sample size is achieved:

\[
N' = \frac{N}{1 - q}
\]

Thus, if 135 is the estimated total sample size required and maximum 20% of recruited subjects are expected to drop out before study ends, the recruitment target would be:

\[
N' = \frac{135}{1 - 0.2} = \frac{135}{0.8} = 168.75
\]

Thus, 169 eligible subjects would need to be recruited in order that at least 135 subjects complete the study.

Using the formula \( \frac{100}{100 - x} \), where \( x \) represents the estimated maximum drop out fraction in percentage, one can easily derive the following set of correction factors [Table 2] that may be applied to the total sample size to derive the screening or recruitment target sample size.

As with other aspects of sample size calculations, the proportion of eligible subjects who will refuse to participate or who drop out will be unknown at the onset of the study. However, good estimates will often be possible using information from similar studies in comparable populations or from an appropriate pilot study. Note that it is particularly important to account for dropouts in the budgeting of studies in which initial recruitment costs or treatment costs are likely to be high.

| Estimated maximum dropout fraction (%) | Total sample size to be multiplied by |
|----------------------------------------|--------------------------------------|
| 5                                      | 1.05                                 |
| 10                                     | 1.11                                 |
| 15                                     | 1.18                                 |
| 20                                     | 1.25                                 |
| 25                                     | 1.33                                 |
| 30                                     | 1.43                                 |
| 35                                     | 1.54                                 |
| 40                                     | 1.67                                 |
Other Considerations in Sample Size Determination

There are some additional issues in sample size determination that are not often discussed, but nevertheless may be relevant in certain situations and therefore merit attention.

First, the above approaches to sample size have assumed that a simple random sample is the sampling design. More complex designs, like stratified random samples or cluster samples, must take into account the variances of subpopulations, strata, or clusters before an estimate of the variability in the population as a whole can be made.

A second consideration is the extent of the analysis that is planned to be performed. If descriptive summaries and simple inferential statistics are planned than almost any sample size is good enough. On the other hand, larger samples are required if multivariate analysis such as multiple regression, analysis of covariance, logistic regression, log-linear analysis, and Cox's proportional hazard analysis are planned for more rigorous assessment of the combined effect of multiple variables. Methods are still evolving to estimate optimum sample size for such multivariate techniques. In studies with “adaptive design,” the sample size may keep changing as the study progresses, based on results obtained. These calculations are quite complex.

Finally, an adjustment in the sample size may be needed to accommodate a comparative analysis of subgroups. There are various suggestions but no hard and fast recommendations toward this. Even if total sample is relatively large, skewed distribution of variables in question can result in erroneous conclusions on subgroup analysis. The safest approach is to plan all subgroup analyses beforehand and ensure that the likely subgroup sample size would have adequate power to demonstrate a clinically important effect on comparison of subgroups.

The sample size for surveys requires different considerations. This is highlighted in Box 1.

Precision versus Effect Size

Frequently, studies are designed to yield an estimate of some parameter of interest, for example, the mean of a continuous variable, proportion, or a difference in proportions. A difference in proportions can be measured as a raw difference, a relative risk, an odds ratio, and in other ways. In each case, however, the precision of the final estimate can be measured by the width of the 95% CI. The larger the sample size used to calculate the final CI, the narrower the CI will be. Therefore, the desired final CI width can be used, instead of the desired effect size, to determine the appropriate sample size for a study. Since the predicted precision or CI width depends mostly on the variance of the data and much less on the effect size, a study can be planned to yield a given precision or CI width without choosing a likely effect size.

When designing a study to yield a CI with a given width, the general technique is to choose a sample size so that the ‘average, resulting CI will have the given width. It is important to realize that, given such a sample size, there is a 50-50 chance that the final width will be narrower or wider than the desired width, even if the estimate used for the variance of the outcome data is accurate. Alternatively, one can choose a sample size so that there is a defined probability (e.g., 0.80 or 0.95) that the final CI width is below a given value. This calculation is more analogous to a typical power calculation, but is more complex and not generally part of currently-available sample size software.

Post hoc Power Analysis

A power calculation needs to be before a study is initiated to determine the appropriate sample size. A power calculation can always be done once the study data have been generated but it is not good, and rather unfair, statistical practice to tweak parameters of the a priori power analysis on the basis of post hoc power analysis.

How can one interpret data from a negative study in which no power calculation was initially performed? Although tempting, performing a post hoc power analysis to estimate the effect size that could have been found with the actual sample size and with a given power, is inappropriate and should not be done. The correct approach to such data is to calculate the 95% CI for the outcome of interest, based on the final data, and use this interval to guide interpretation of the study results. Incidentally, a CI width-based power and sample size calculation can also be done before initiating the study. This is routinely done in cohort, case-control and other epidemiological studies.

What if There is No Choice About Sample Size

Finally, before we close this chapter, let us address this common dilemma. Often, a study has a limited budget, and this curtails the possibility of a “comfortably” large sample. It is hard to argue with a budget. It is equally unwelcome to give up (the aptly called “terminator” approach) and say that the study cannot be done. The practical alternative is to realize that by tweaking aspects of sample size calculation it may still be possible to execute the study within the constraints of a restricted budget and therefore the imposed sample size. It may be possible to raise the effect
Determining the sample size for a simple survey requires consideration of four elements:

- **ME**: This denotes the extent of deviation from the true result that you are willing to tolerate. Thus if the true prevalence figure for a disease in a population is say 30%, and we are happy if our result is in the 25%-35% range then we are tolerating a±5% margin of error. Lower margin of error requires a larger sample size. In fact the sample size can change dramatically with change in the margin of error. By convention, however, this figure should not exceed 5%.

- **CL**: This is the extent of uncertainty we can tolerate. In the above example, if we are saying that we are happy if our result is in the 25%-35% range, then how confident are we that the true figure will actually lie in this range. Higher confidence level requires a larger sample size but 95% is generally the minimum confidence level we should work with. With this choice, when you survey a sample of the population, you don’t know that you’ve found the correct answer, but you do know that there’s a 95% chance that you’re within the margin of error of the correct answer.

- **Population size**: We need some idea of the size of the population we are to sample from. However, the sample size doesn’t change much for populations larger than 20,000. This is a great thing because if the population size is unknown, we can simply take an arbitrarily large figure of 20,000 or bigger.

- **RD**: This may also be thought of as the crude prevalence. If you test a random sample of 10 people for a disease, and 9 of them test “positive” then the prediction that you make about the general population is different than say if 5 test positive. Therefore this factor enters into the calculation. Setting the response distribution to 50% is the most conservative choice as it assumes a most heterogenous population and will return the largest sample size. If we have a crude idea of the response distribution or prevalence, we can use this in the calculation. Any figure moving closer to 0 or 100, than 50%, will return a smaller sample size, the other variables remaining constant.

The formula that we use is:

\[ \text{Sample size} = \frac{[RD] \times (1-RD)}{([ME/CL score]^2)} \]

If we are working with a finite population, we apply a finite population correction:

\[ \text{Corrected sample size} = \frac{(\text{Sample size} \times \text{population})}{(\text{sample size} + \text{population} - 1)} \]

Let us work out an example. Suppose we want to find out what proportion of dermatology residents, from a population of 1000 residents, hate biostatistics. We will assume 5% margin of error and 95% confidence level. We have no prior idea of the response distribution and so will use 50%.

Applying the first formula, we get

\[ \text{Sample size} = \frac{[0.5] \times (1-0.5)}{([0.05/1.96]^2)} = \frac{0.25}{0.00065077} = 384.16 \]

Since our population is relatively small, we apply a finite population correction

\[ \text{Corrected sample size} = \frac{(384.16 \times 1000)}{1383.16} = 277.74 \]

Thus we need to survey about 278 respondents to find out the answer to our question. Note that if we reduce our error margin to 3%, the required sample size jumps to 517.

In real life surveys, we need to adjust the calculated sample size for anticipated nonresponse rate and also anticipated attrition rate, if the survey is to be administered more than once. If we adopt stratified or cluster sampling strategy while selecting respondents, adjustments also need to be made on these counts.

A very easy to use survey sample size calculator, that is free to use, can be found online at [http://www.raosoft.com/samplesize.html](http://www.raosoft.com/samplesize.html)

**Conclusion**

Determining the appropriate sample size for a study is a fundamental aspect of ethical research. Performing a valid sample size calculation requires estimates of the permissible Type I error rate (\(\alpha\)) and power, variability in the data, the effect size sought, as well as a planned method of analysis. Although the concepts underlying power analysis and sample size determination are relatively simple, the large number of different study designs and analysis methods results in a bewildering number of different sample size formulas. Therefore, the use of power and sample size software is helpful and is gaining widespread usage.
**Financial support and sponsorship**
Nil.

**Conflicts of interest**
There are no conflicts of interest.

**Further Reading**
1. Julios SA. Sample sizes for clinical trials with normal data. Stats Med 2004;23:1921-86.
2. Devane D, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. J Adv Nursing 2004;47:297-302.
3. Karlsson J, Engebretsen L, Dainty K, ISAKOS scientific committee. Considerations on sample size and power calculations in randomized clinical trials. Arthroscopy 2003;19:997-9.
4. Statistical power and sample size. In: Cleophas TJ, Zwinderman AH, Cleophas TF, Cleophas EP. Statistics applied to clinical trials. 4th ed. Amsterdam: Springer; 2009. p. 81-97.
5. Calculation of required sample size. In: Kirkwood BR, Sterne JA. Essential medical statistics. 2nd ed. Oxford: Blackwell Science; 2003. p. 413-28.