Sample size determination in clinical research: 1

Laofe Ogundipe

In this two-part series, the main designs in clinical research are considered and, using examples, sample size calculation is explained. The information required in order to calculate the sample size for clinical research is highlighted to enable researchers to spend their time most efficiently with statisticians.

INTRODUCTION

One of the questions that is most commonly asked in planning a research study, particularly by research ethics committee, is how many subjects will be needed? In clinical research, size matters. It enables the researcher to avoid a type 2 error, the error of incorrectly stating that there is no difference between two groups when in reality a difference exists.

A small sample is more liable to cause a type 2 error. On the other hand, too large a sample size may be unnecessary and may be wasteful of resources. In a clinical trial, too large a sample means that more subjects will be unnecessarily exposed to the less effective treatment. Therefore what we need is a sample that is large enough to provide a specified degree of precision.

Clinicians with some knowledge of sample size calculation will get a better and faster result from their statistician and will be better able to justify the choice of sample size. Readers on the other hand will understand the assumptions made in calculating the sample size and will be better able to assess the internal validity of the study. This two-part series provide a step by step guide to sample size calculation for the most common clinical research designs.

Sample size calculation, like many statistical techniques, is based on certain assumptions and in some cases on simple formulae.

Dr Laofe Ogundipe is Specialist Registrar in Psychiatry, Lyme Brook Mental Health Centre, Newcastle

Correspondence to: Dr L Ogundipe, Deramore, Ercall Heath, Nr Tibberson, Shropshire TF10 8NQ

CLINICAL RESEARCH DESIGNS

The main designs in clinical research are shown in *Table 1*.

Sample size in case series

There is no requirement for sample size calculation in a case series. A single subject can be described as a case report. This design is used to alert the scientific world to certain events or possibilities. It is not a powerful research design but the crucial role of the case report has been emphasized (Farmer, 1999).

A single subject (N-of-1) double blind randomized controlled trial is a special form of case report that is very powerful yet neglected in clinical research. A single patient is subjected to periods of active and placebo treatments, the patient serving as his/her own control. The subject and the assessor are blind to the period that the placebo or the active agent is being used. A diary of the outcome is recorded and comparison is made between the periods when the two agents are used. Statistical techniques are available to determine any statistically significant difference between the two periods. The main disadvantage of N-of-1 trial is its lack of generalizability. Its main advantage is that it improves the certainty of a treatment decision in an individual patient rather than the average effect on an experimental population. This research design is illustrated fully by Sackett et al (1991) and its limitations are described by Lewis (1991).

For N-of-l trial to be a suitable design, the condition must have reversible manifestations, be stable over a period of time and be a chronic condition that interferes with patients' lives. *Table 2* shows some of the psychiatric conditions that may be amenable to this design. The special application of this design to investigating dissocial personality disorder has been suggested (Ogundipe and Lovett, 1999).

Sample size for prevalence studies

In planning a service, it is essential to estimate the magnitude of the problem

TABLE 1.
Clinical research designs

Case series	Single case report		
	Report on a group of individuals with a particular condition		
	N-of-1 randomized controlled trial		
Prevalence studies or surveys	To estimate the prevalence of a categorical variable		
	To estimate the mean value of a continuous variable		
Comparative studies	Comparing two proportions in binary outcome measure		
	Comparing values of continuous variables	Case control study	
		Cohort study	
		Randomized controlled trial	Parallel design
			Cross-over design

TABLE 2.
Psychiatric conditions that can be investigated by N-of-1 randomized controlled trial design

Dissocial personality disorder
Attention deficit hyperactivity disorder
Tardive dyskinesia
Enuresis
Tisc/Tourette syndrome
Conduct disorder
Learning disability, e.g. self-injurious episodes

for which the service is directed and surveys or prevalence studies are important clinical studies.

Prevalence is the proportion of the population which manifests the variable or disease of interest at a particular time point. Individuals in the population can be classified as having or not having the condition, therefore prevalence is a categorical variable. Calculating sample size for prevalence studies is now illustrated in a worked example.

Example 1: We may want to determine the prevalence of schizophrenia in our catchment population of 480 000 people. How many people should we interview?

Step 1: What is the rough estimate of schizophrenia in our area? This may sound like a chicken-and-egg problem — how can we have a rough estimate of the parameter that we want to measure? We simply look at the literature for past surveys, e.g. the national figure. The national prevalence of schizophrenia has been reported in the literature to be 12 per 1000 people or 1.2% (Prevalence (P) =0.012). Assuming that the prevalence of schizophrenia in our area is similar, our prevalence is 0.012.

Step 2: How accurately should we conduct the survey? The more accurate we want to be, the bigger the sample size required. If we choose to be accurate to the nearest 0.5%, our standard error will be 0.25% (standard error(SE)=0.0025). This is the margin of error (at 95% confidence interval) below or above (half of the accuracy each side) the estimated figure.

Step 3: Calculate the sample size using the formula in *Figure 1*.

Sensitivity analysis: A sample size of 1897 is quite large and it may not be possible to recruit such a large number of people for the study. Is there no way we can reduce it? A sensitivity analysis can be carried out. This is a series of calculations of sample size, based on different assumptions and different requirements of precision. If we choose to be less precise in our study, we may be able to reduce the sample size, provided the precision is still acceptable.

In the example above, let us be less accurate. We will therefore reduce the precision of our estimate of the prevalence of schizophrenia by increasing the margin of error from 0.5% to the nearest 0.9%. This will give us a standard error of 0.45% (half the margin of error on either side of the estimate). Using the formula again the sample size is 587, as illustrated in *Figure 1*.

Example 2: To illustrate the principle further, let us assume that we are interested in estimating the proportion of people presenting at the accident and emergency department with deliberate self-harm who suffer from depressive disorder. If our literature search found a paper which reported that 54% of people attending the accident and emergency department in their hospital with deliberate self-harm suffer from depressive disorder, we would expect a similar proportion in our service. If we want to measure this with a precision of 10% at 95% confidence interval, the standard error will be 5%.

Figure 2 shows the calculation of the minimum sample size required, which will be 100 people. If we can not recruit 100 subjects, we may reduce the precision to the nearest 20%. This gives a sample size of 25 as shown in *Figure 2*.

Finite population correction: It is important to note that the size of the population from which the sample is drawn is not crucial to our calculation if the sampling fraction is less than 5%. Sampling fraction is the sample size divided by the total population from which the sample is drawn. If, however, the sampling fraction is more than 5%, a procedure called finite population correction (Moser and Katton, 1979) can be used to adjust (reduce) the calculated sample size. Discussion of this method is beyond the scope of this article.

Sample size for estimating the value of a continuous variable

Prevalence is a proportion or a categorical variable but sometimes we may be interested in estimating the value of a continuous variable. These are variables that can be measured in units and can take any number in a range of values. We may want to calculate the pro-

$$\text{Sample size} = N = \frac{P(1-P)}{Se^2}$$

where P=estimated prevalence from literature

Se=standard error = 0.0025

$$N = \frac{0.012(1-0.012)}{0.0025 \times 0.0025} = \frac{0.012 \times 0.988}{0.00000625} = \frac{0.011856}{0.00000625} = 1897$$

or $\text{sample size} = N = \frac{P(1-P)}{Se^2}$

where P=estimated prevalence from literature

Se=standard error = 0.0045

$$N = \frac{0.012(1-0.012)}{0.0045 \times 0.0045} = \frac{0.012 \times 0.98}{0.0000202} = \frac{0.011856}{0.0000202} = 587$$

Figure 1. Sample size calculation for Example 1 (prevalence studies).

$$N = \frac{P(1-P)}{Se^2} = \frac{0.54 \times 0.46}{0.05 \times 0.05} = \frac{0.2484}{0.0025} = 99.36$$

$$\textit{or } N = \frac{P(1-P)}{Se^2} \quad \frac{0.54 \times 0.46}{0.1 \times 0.1} = \frac{0.2484}{0.01} = 24.84$$

Figure 2. Sample size calculations for Example 2.

portion of people falling above or below a specified value of the variable, in which case we can use the procedure for a categorical variable explained above. Alternatively, we may want to calculate the mean of the variable in the population of interest. In this case, the formula is slightly different.

For example, chronic alcohol dependence is associated with depletion of thiamine in the body. We may want to estimate the proportion of people having a thiamine level below 100 nmol/l using sample size calculation for a categorical variable above. Alternatively, we may want to calculate the mean serum level of thiamine in people with chronic alcohol dependence. The procedure for calculating sample size when estimating a continuous variable is as follows:

Example 3: Calculating the mean serum level of thiamine in a population of alcoholics.

Step 1: What is the rough estimate of the standard deviation (the spread around the mean value) of serum levels of thiamine in chronic alcoholics? Again, we are asking a question (estimate of a parameter) to which our research is supposed to provide answers. Standard deviation is a measure of the spread of individual results from the mean value. We don't know the mean value and we don't have a single result, so how do we know how the individual results differ from the mean? Again, we search the literature. Has anyone ever reported the standard deviation of the serum levels for thiamine in alcohol dependent people? If no estimate is available in the literature, the solution is to do a pilot run of the study (Bland, 1992; Armitage and Berry, 1996), estimate thiamine levels in about 10 subjects and then calculate the standard deviation from this.

Now if in the literature the standard deviation of thiamine in chronic alcoholics has been reported to be 40 nmol/l, we will take the following steps in calculating the required sample size.

$$\text{Sample size} = N = \frac{SD^2}{Se^{2*}}$$

$$N = \frac{40 \times 40}{5 \times 5} = 64$$

The minimum sample required is 64

Figure 3. Sample size calculations for Example 3. * Variance divided by the square of the standard error (Bland, 1992).

Step 2: How accurately do we want to estimate the thiamine level? There is no law about how accurate we have to be in our rough estimate of the sample size. Standard error is a measure of the accuracy of our estimate (the margin of error on either side of the estimate). To be accurate to within 10 nmol/l, at 95% confidence interval, will give a standard error of 5 nmol/l above or below the true value. We can set the size of the standard error we want and choose the sample size to achieve this. (Bland, 1992). The standard error can also be related to the mean value if this is known but for simplicity, the formula for this has been omitted. Therefore, standard error=Se=5.

Step 3: Calculate the sample size from the simple formula in *Figure 3*.

Will it be possible to recruit 64 people to the study? If not, do not despair, statistics can help. We can carry out a sensitivity analysis, and assume a smaller or greater margin of error. We can recalculate the sample size by increasing or reducing the accuracy of our study.

For example, to estimate the mean of thiamine level to the nearest 26 nmol/l, a standard error of 13 will require a sample size of 10. Also, note that a more precise estimation, with a smaller margin of error, for example to the nearest 3.6nmol/1, giving a standard error of 1.8 will require a sample size of about 500.

KEY POINTS

- Sample size in clinical research is an important ethical issue.
- A calculated sample size for research can only be a rough estimate for guidance.
- A priori determination of sample size will help avoid publishing misleading results.
- Certain assumptions will be required but this is better than a hit-and-miss attempt at choosing sample size.
- For prevalence studies, the precision required and an estimate of the prevalence (for categorical variables) or the standard deviation (for continuous variables) are the necessary information for calculating sample size.

CONCLUSIONS

When estimating prevalence of a condition, we need three pieces of information:

1. The type of variable, categorical or continuous
2. If categorical, the rough estimate of the prevalence of the variable we want to estimate, which is obtainable from the literature. If continuous, the rough estimate of the standard deviation
3. The precision of the study (the margin of error we are prepared to accept).

AVOIDING MISLEADING RESULTS

If we can only recruit 30 subjects when 500 subjects will be required for our predetermined accuracy, we may have to abandon the study altogether. This will avoid publishing a misleading and unethical result, or a result with an unacceptable margin of error. Alternatively, we may consider a collaborative multicentre study and pool together resources from many centres. Whatever we do, we must let the readers know the precision of our estimate by stating the confidence interval, the range of values that, with a given level of probability, include the true but unknown value. **HM**

The author would like to thank Peter Jones, Professor of Mathematical Statistics, Dr Richard Hodgson, Consultant Psychiatrist, City General Hospital, Stoke-on-Trent and Dr Sayeed Haque, Medical Statistician, South Birmingham NHS Trust, for their comments on earlier drafts of this paper. Any remaining error is solely the authors responsibility.

Armitage P, Berry G (1996) *Statistical Methods in Medical Research.* Blackwell Scientific Ltd, Oxford

Bland M (1992) *An Introduction to Medical Statistics.* Oxford University Press, Oxford

Farmer A (1999) The demise of published case report - is resuscitation necessary? *Br J Psychiatry* **174:** 93–4

Lewis J (1991) Controlled trials in single subjects: limitations of use. *Br Med J* **303**: 175–6

MacRae K (1992) *Statistics in Psychiatric Research.* WB Saunders, London

Moser CA, Katton G (1979) *Survey Methods in Social Investigations.* Buttler and Tanner, London

Ogundipe L, Lovett L (1999) Childhood hyperkinetic disorder/attention deficit disorder grown up. *Int J Psychiatry Clin Pract* **3:** 217–19

Sackett DHB, Haynes RB, Tugwell P, Guyatt G (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Little Brown and Company, London