How many will I need for this study?

Gissane, C.

School of Sport, Health and Applied Science, St Mary's University, Twickenham, Middlesex, TW1 4SX, UK.

conor.gissane@stmarys.ac.uk

Good research studies are planned,[1] and at some point in the process investigators ask the question, "How many participants do we need?" Researchers should consider this in relation to their specific question. Eng,[2] stated that it should be settled before beginning the study. The decision can be made on several considerations, such as resources, practical issues on conducting the research and time constraints.[3] The sample size can also be estimated formally, a process that allows the researcher to achieve both a statistically significant and clinically important result.[4] Sample size estimations are often maligned and misunderstood. They are common in health care disciplines, but not other disciplines.[3] They have their supporters,[4, 5] their critics[6, 7] and alternatives have been offered.[8, 9] However, their use is unlikely to disappear,[3] and if conducting a randomised controlled trial they are essential.[10] This paper will attempt to illustrate how sample size estimates can be used in certain situations and offer the calculations in their simplest form.

So what does a researcher have to do? The determination of sample size has been defined as, "The mathematical process of deciding, before a study begins, how many subjects should be studied."[11] It is the minimum number of participants needed to identify a significant difference, provided it exists.[4] Sample size estimation shares its origins with hypothesis testing.[12] To conduct a sample size estimate, the researcher must decide some parameters by using their expertise of the subject area. These are: the significance level (α), the power of the study (β), the difference to be detected and the

variability of the measure under observation. In addition they have to decide if they are doing a one sided or a two sided test. In this paper all tests are two sided.

Two parameters are often set by convention, significance and power. The significance level (α) of the study is mostly set at 0.05. It corresponds to the probability of making a type I error, rejecting the null hypothesis when it is true. Controlling for type I errors prevents the rise of ineffective treatments.[13] The power of a study is usually set at 0.8. The power of a study refers to the chance of accepting the null hypothesis when it is false (Type II error).[14] Biau et al.,[13] referred to type I errors as false positives and type II errors as false negatives. They also emphasised the importance of sample size estimates as the sample size determines the risk of false negative results. The researcher also needs to estimate the minimum expected difference between the two groups being investigated. This is a subjective parameter and it will be based on clinical judgements and the expertise of the researchers.[2] Lastly, the variability of the outcome measure under investigation must be determined. For a continuous variable (interval/ratio), a standard deviation is required. It is unlikely to be available prior to the study[13], so it is usually determined from preliminary data or a review of the literature can provide estimates.[2]

The actual calculations for sample sizes can be mystifying and frightening when first encountered (equation 1).[13] However, investigators have produced other methods that are friendlier, which are presented as equations, tables[15] or nomograms.[16, 17] Perhaps Lehr[18] offers the easiest way estimate the number of subjects per group at p<0.05 and power =80% (equation 2). Where *s* is the standard deviation and *d* is the difference to be detected. The author suggested the formula was easy to commit to memory. Also, if another level of power or significance was required it is simply a matter of substituting different number into the equation (table 1).

$$n = \frac{2 \times \left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{d^2} + 0.25 \times Z_{1-\alpha/2}^2 \qquad (1)$$

Where $Z_{1-\alpha/2}$ =1.96 at P = 0.05 and $Z_{1-\beta}$ = 0.84 at power of 0.8 and d^2 is the targeted standardised difference between means (expected mean difference/standard deviation)

$$n = 16 \times \frac{s^2}{d^2} \tag{2}$$

Where s is the standard deviation and d is the difference to be detected.

Table 1. Coefficients to substitute for 16 for different significance and power specifications.[18]

Power	Alpha level		
	0.01	0.05	0.10
0.80	23.5	16	12.5
0.90	30	21	17.5
0.95	36	26	22

Example:

Campbell et al.,[15] offered an example where clinicians were looking for 5 mmHg change in blood pressure using a standard deviation of 17 mmHg. This calculation had set power to 0.80 and significance to 0.05. Equation 1 gives a value of 183 participants per group, after rounding up (equation 3). The Lehr[18] equation (equation 4) gives a value of 185 participants per group after rounding up. It is always a good idea to round up a sample size estimate, as rounding down will technically leave you short of participants. The estimates produced by the two equations are similar.

$$n = \frac{2 \times (1.96 + 0.84)^2}{(5/17)^2} + 0.25 \times 1.96^2 = \quad 183 \tag{3}$$

$$n = 16 \times \frac{17^2}{5^2} = \quad 185 \tag{4}$$

Not all studies seek to compare two groups or two treatments. Sometimes the objective is to describe a given characteristic in a specific population. In this type of design estimating the sample size is important because it determines how precise the estimate can be. If a researcher is prepared to declare

3

an acceptable margin of error, sample size can be estimated using equation 5. This is equal to half the width of a 95% confidence interval.

$$n = \frac{4s^2}{d^2}$$

(5)

Where s is the standard deviation and d is the declared acceptable margin of error.

Again using the data from Campbell,[15] it is decided that a confidence interval width of 5 is required, and equation 6 estimates 47 participants are needed.

$$n = \frac{4(17)^2}{5^2} = 47$$

(6)

Sample size calculations have been an important development in statistics over the last 30 years.[9, 19] Their use has been beneficial, increasing median sample size in studies 100 fold.[9] More participants increases power, and reduces the chance of not detecting an effect when an effect does exist.[4] Using sample size estimates has been criticised philosophically,[6] and on the inaccurate determination of the variability measures used. If the variability is underestimated, the study will be underpowered.[13] Nevertheless, carrying out studies such randomised controlled trials demands their use, and they do focus the attention to correct planning. When they are used, the estimates should be carried out on the primary outcome measure, which is decided *a priori*.[9] Lastly, put time and effort into deciding the variability of the outcome measure, getting it wrong will drastically affect the study.[3]

References

[1]   Lenth R. Some practical guidelines for effective sample size determination. The American Statistician. 2001;55:187-93.

[2]   Eng J. Sample size estimation: how many individuals should be studied? Radiology. 2003;227:309-13.

[3]   Norman G, Monteiro S. Sample size calculations: should the emporer's clothes be of fthe peg ofr made to measure? BMJ. 2012;345:e5278.

[4]   Burmeister E, Aitken LM. Sample size: how many is enough? Australian Critical Care. 2012;25:271-4.

[5]   Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol. 2012;65:301-8.

[6]   Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Medicine. 2010;8:17.

[7]   Bacchetti P, Wolf LE, Segal MR, McColloch CE. Ethics and sample size. Am J Epidemiol. 2005;161:105-10.

[8]   Bacchetti P, McColloch CE, Segal MR. Simple, defensible sample sizes based on cost efficiency. Biometrics. 2008;64(2):577-94.

[9]   Bland JM. The tyranny of power: is there a better way to calculate sample size? BMJ. 2009;339:1133-5.

[10]   Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallell group randomized trials. BMC Medicine. 2010;8:18.

[11]   Porta M, Last J, editors. A dictionary of epidemiology. New York: OUP; 2008.

[12]   Descoteaux J. Statistical power: an historical introduction. Trends in Quantitative Methods for Psychology. 2007;3(2):28-34.

[13]   Biau D, Kerneis S, Porcher R. The importance of sample size in the planning and interpretaion of medical research. Clin Orthop Relat Res. 2008;466:2282-8.

[14]   Everitt BS. Dictionary of Statistics in Medical Sciences. Cambridge: Cambridge University Press; 1995.

[15]   Campbell MJ, Julious SA, Altman D. Estimating sample sizes for binary, ordered categorical and continuous outcomes in two group comparisons. BMJ. 1995;311:1145-8.

[16]   Altman DG. Practical statistical for medical research. London: Chapman & Hall; 1991.

[17]   Whitley E, Ball J. Staistics review 4: sample size calculations. Critical Care. 2002;6:335-41.

[18]   Lehr R. Sixteen s-sqaured over d-squared: a relation for crude sample size estimates. Statistics in Medicine. 1992;11:1099-102.

[19]   Corty EW, Corty RW. Setting the sample size to ensure narrow confidence intervals for precise estimation of population values. Nursing Research. 2011;60(2):148-53.