

Sample Size and its Importance in Research

Chittaranjan Andrade

ABSTRACT

The sample size for a study needs to be estimated at the time the study is proposed; too large a sample is unnecessary and unethical, and too small a sample is unscientific and also unethical. The necessary sample size can be calculated, using statistical software, based on certain assumptions. If no assumptions can be made, then an arbitrary sample size is set for a pilot study. This article discusses sample size and how it relates to matters such as ethics, statistical power, the primary and secondary hypotheses in a study, and findings from larger vs. smaller samples.

Key words: *Ethics, primary hypothesis, research methodology, sample size, secondary hypothesis, statistical power*


Studies are conducted on samples because it is usually impossible to study the entire population. Conclusions drawn from samples are intended to be generalized to the population, and sometimes to the future as well. The sample must therefore be representative of the population. This is best ensured by the use of proper methods of sampling. The sample must also be adequate in size – in fact, no more and no less.

SAMPLE SIZE AND ETHICS

A sample that is larger than necessary will be better representative of the population and will hence provide more accurate results. However, beyond a certain point, the increase in accuracy will be small and hence not worth the effort and expense involved in recruiting the extra patients. Furthermore, an overly large sample

would inconvenience more patients than might be necessary for the study objectives; this is unethical. In contrast, a sample that is smaller than necessary would have insufficient statistical power to answer the primary research question, and a statistically nonsignificant result could merely be because of inadequate sample size (Type 2 or false negative error). Thus, a small sample could result in the patients in the study being inconvenienced with no benefit to future patients or to science. This is also unethical.

In this regard, inconvenience to patients refers to the time that they spend in clinical assessments and to the psychological and physical discomfort that they experience in assessments such as interviews, blood sampling, and other procedures.

Access this article online	
Website: www.ijpm.info	Quick Response Code
DOI: 10.4103/IJPSYM.IJPSYM_504_19	

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Andrade C. Sample size and its importance in research. Indian J Psychol Med 2020;42:102-3.

Clinical Psychopharmacology Unit, Department of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India

Address for correspondence: Dr. Chittaranjan Andrade

Clinical Psychopharmacology Unit, Department of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bengaluru - 560 029, Karnataka, India. E-mail: andrdec@gmail.com

Received: 09th December, 2019, **Accepted:** 09th December, 2019, **Publication:** 06th January, 2020.

ESTIMATING SAMPLE SIZE

So how large should a sample be? In hypothesis testing studies, this is mathematically calculated, conventionally, as the sample size necessary to be 80% certain of identifying a statistically significant outcome should the hypothesis be true for the population, with P for statistical significance set at 0.05. Some investigators power their studies for 90% instead of 80%, and some set the threshold for significance at 0.01 rather than 0.05. Both choices are uncommon because the necessary sample size becomes large, and the study becomes more expensive and more difficult to conduct. Many investigators increase the sample size by 10%, or by whatever proportion they can justify, to compensate for expected dropout, incomplete records, biological specimens that do not meet laboratory requirements for testing, and other study-related problems.

Sample size calculations require assumptions about expected means and standard deviations, or event risks, in different groups; or, upon expected effect sizes. For example, a study may be powered to detect an effect size of 0.5; or a response rate of 60% with drug vs. 40% with placebo.^[1] When no guesstimates or expectations are possible, pilot studies are conducted on a sample that is arbitrary in size but what might be considered reasonable for the field.

The sample size may need to be larger in multicenter studies because of statistical noise (due to variations in patient characteristics, nonspecific treatment characteristics, rating practices, environments, etc. between study centers).^[2] Sample size calculations can be performed manually or using statistical software; online calculators that provide free service can easily be identified by search engines. G*Power is an example of a free, downloadable program for sample size estimation. The manual and tutorial for G*Power can also be downloaded.

PRIMARY AND SECONDARY ANALYSES

The sample size is calculated for the primary hypothesis of the study. What is the difference between the primary hypothesis, primary outcome and primary outcome measure? As an example, the primary outcome may be a reduction in the severity of depression, the primary outcome measure may be the Montgomery-Asberg Depression Rating Scale (MADRS) and the primary hypothesis may be that reduction in MADRS scores is greater with the drug than with placebo. The primary hypothesis is tested in the primary analysis.

Studies almost always have many hypotheses; for example, that the study drug will outperform placebo on

measures of depression, suicidality, anxiety, disability and quality of life. The sample size necessary for adequate statistical power to test each of these hypotheses will be different. Because a study can have only one sample size, it can be powered for only one outcome, the primary outcome. Therefore, the study would be either overpowered or underpowered for the other outcomes. These outcomes are therefore called secondary outcomes, and are associated with secondary hypotheses, and are tested in secondary analyses. Secondary analyses are generally considered exploratory because when many hypotheses in a study are each tested at a $P < 0.05$ level for significance, some may emerge statistically significant by chance (Type I or false positive errors).^[3]

INTERPRETING RESULTS

Here is an interesting question. A test of the primary hypothesis yielded a P value of 0.07. Might we conclude that our sample was underpowered for the study and that, had our sample been larger, we would have identified a significant result? No! The reason is that larger samples will more accurately represent the population value, whereas smaller samples could be off the mark *in either direction* – towards or away from the population value. In this context, readers should also note that no matter how small the P value for an estimate is, the population value of that estimate remains the same.^[4]

On a parting note, it is unlikely that population values will be null. That is, for example, that the response rate to the drug will be exactly the same as that to placebo, or that the correlation between height and age at onset of schizophrenia will be zero. If the sample size is large enough, even such small differences between groups, or trivial correlations, would be detected as being statistically significant. This does not mean that the findings are clinically significant.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Norman G, Monteiro S, Salama S. Sample size calculations: Should the emperor's clothes be off the peg or made to measure? *BMJ* 2012;345:e5278.
2. Andrade C. Signal-to-noise ratio, variability, and their relevance in clinical trials. *J Clin Psychiatry* 2013;74:479-81.
3. Andrade C. Multiple testing and protection against a type 1 (false positive) error using the Bonferroni and Hochberg corrections. *Indian J Psychol Med* 2019;41:99-100.
4. Kraemer HC. Is it time to ban the P value? *JAMA Psychiatry* 2019;76:1219-20.