

### Sample size in research. When can you break the rule?

Adequate sample size is important when conducting research, especially quantitative studies. It has to be calculated properly before data collection begins. The size should be adequate to achieve the objective of the study. Not smaller or bigger. Having said that, sample size is still an estimate based on prior knowledge and expectation. If adequate sample size is impossible, it is better not to proceed with the research.<sup>1,2</sup> But there are circumstances that allow small sample size study to be acceptable. However one should be very well informed of the rule before one attempts to break it.

#### Why sample size is important?

When a population-based study or survey is conducted, the results are usually inferred back to the bigger target population. The issue here is whether the sample selected for the study represent the population adequately. To ensure the 'representativeness', the sampling method should be random and the sample size should be adequate. Both sampling method and sample size are important in such study. In experimental research, the study population is very specific. For example, in a study to measure efficacy of a new lipid lowering agent comparing with the established ones, sample should be selected from patients with dyslipidaemia. They do not have to be from a certain institution or district or state or nation. The sample should represent patients with abnormal lipid problems. While the sampling method does not have to be random, adequate sample size is still necessary to ensure it has enough *statistical power* to prove the hypothesis that the new drug is better than the current ones.<sup>3</sup>

Statistical power (or  $1-\beta$ ) is the probability of not making Type II errors ( $\beta$ ). Type II error in hypothesis testing is accepting the result when it is wrong or in a more confusing statistical statement, not rejecting the null hypothesis when it is false. Normally, we put a limit to the size of error allowed in testing any hypothesis. For statistical power, we want it to be 0.8 or 80%. This means the error allowed is 20%, or the probability to falsely accept a wrong result is 20%. In a more sensitive clinical study when erroneous results are less tolerable, the power can be set to 90%. However, when more statistical power required, more samples needed for that study.

Sample size is also affected by how accurate you expect the difference (or the treatment effect) will be. If the expected difference between new drug and

the established one is very small, huge sample size is required for the study.<sup>4,5</sup> For example, if you expect the patients given the new lipid lowering drug will have lower average total cholesterol by 0.1 mmol/L compared to those receiving standard drug at the end of the trial (with an assumption of 0.05 mmol/L variations between subjects, i.e. the standard deviation), the sample size will be around 5 per group (or total of only 10 subjects in the two arms trial). When the expected difference is 0.01 mmol/L, the sample required is 393 per group, when calculated for independent sample t-test.<sup>6</sup> So if such study was planned with 10 samples, it will most likely fail to detect significant difference of anything smaller than 0.1 mmol/L.

#### What is effect size?

P-value has achieved a divinely status for some researchers. They 'worship' P-value to the extent that, when  $P < 0.05$ , the finding is of absolute truth and when  $P > 0.05$ , nothing further can be done. P-value is affected by sample size. The bigger the sample size the most likely the P-value to be small. P-value does not measure the magnitude of difference, or it does not measure the actual treatment effect. Effect size, on the other hand, measures the actual treatment effect. There is no one single formula for effect size, and it is not widely used. Combined use of effect size and confidence interval can be utilised to measure relationship compared to P-value, regardless the sample size.<sup>7</sup> The use of effect size is not yet easily available in many statistical software package, but this may change very soon.

#### When you can break the rule?

One can start a research with small sample size when testing a new hypothesis. When something is very new, there is no expectation, or estimation, hence difficult to calculate precise required sample size.<sup>8</sup> For such research, it is acceptable to proceed as a discovery research and descriptive analyses of the findings is adequate. The study is still weak in its statistical significance but may yield important clinical findings. In a different scenario, one may have already calculated the sample size very well but some unfavourable circumstances may lead to poor response, hence small sample size at the end of scheduled data collection. Does this mean the whole study is of no use? In such cases, there are few measures that can be taken to salvage the study. First and foremost, the actual sample size must be clearly mentioned, and this shall

be the limitation of the study. The P-value should be reported together with the effect size. Post-hoc power can also be calculated to support the need to repeat the study with larger sample size if relevant. Statistical analysis can still be performed. It depends on the level of measurement of the variables, whether they are numerical, ordinal or dichotomous. Association between two numerical variables requires smaller sample size compared to the test between proportions. In worst-case scenario, one can use non-parametric statistical tests when the sample size is not adequate to assume normal distribution.

## CONCLUSION

Sample size calculation is a very important component in research. It is a general rule to avoid conducting research with inadequate statistical power. However, when embarking on research for a true novel idea, then please proceed with the research. If you ended up with very small sample size due to your failure to anticipate non-response or attrition, you can still proceed with some statistical analyses but please avoid highlighting its P-value. The analysis can be complemented by measuring the effect size. When in doubt, the most appropriate action is to consult your statistician.

## REFERENCES

1. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on pattern analysis and machine intelligence* 1991;13:252-64.
2. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365-76.
3. Freiman JA, Chalmers TC, Smith Jr H, Kuebler RR. The importance of beta, the type II error and

sample size in the design and interpretation of the randomized control trial. *Survey of 71" negative" trials. New Engl J Med* 1978;299:690-4.

4. Lenth RV. Some practical guidelines for effective sample size determination. *The American Statistician* 2001;55:187-93.
5. Eng J. Sample size estimation: how many individuals should be studied? *Radiology* 2003;227:309-13.
6. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. *Control Clin Trials* 1990;11:116-28.
7. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* 2007;82:591-605.
8. Bacchetti P, Deeks SG, McCune JM. Breaking free of sample size dogma to perform innovative translational research. *Sci Transl Med* 2011;3:87ps24.

Assoc. Prof. Dr. Jamalludin Ab Rahman  
*B.Med.Sc., MD, MPH*  
 Department of Community Medicine,  
 Kulliyah of Medicine,  
 International Islamic University Malaysia