



## EDITORIAL

# Small studies: strengths and limitations

A. Hackshaw

**A** large number of clinical research studies are conducted, including audits of patient data, observational studies, clinical trials and those based on laboratory analyses. While small studies can be published over a short time-frame, there needs to be a balance between those that can be performed quickly and those that should be based on more subjects and hence may take several years to complete. The present article provides an overview of the main considerations associated with small studies.

### HOW SMALL IS “SMALL”?

The definition of “small” depends on the main study objective. When simply describing the characteristics of a single group of subjects, for example the prevalence of smoking, the larger the study the more reliable the results. The main results should have 95% confidence intervals (CI), and the width of these depend directly on the sample size: large studies produce narrow intervals and, therefore, more precise results. A study of 20 subjects, for example, is likely to be too small for most investigations. For example, imagine that the proportion of smokers among a particular group of 20 individuals is 25%. The associated 95% CI is 9–49. This means that the true prevalence in these subjects generally is anywhere between a low or high value, which is not a useful result.

When comparing characteristics between two or more groups of subjects (*e.g.* examining risk factors or treatments for disease), the size of the study depends on the magnitude of the expected effect size, which is usually quantified by a relative risk, odds ratio, absolute risk difference, hazard ratio, or difference between two means or medians. The smaller the true-effect size, the larger the study needs to be [1, 2]. This is because it is more difficult to distinguish between a real effect and random variation. Consider mortality as the end-point in a trial comparing drug A and a placebo with 100 subjects per group. If the 1-yr death rate is 15% for drug A and 20% for the placebo, the risk difference is 5%, but this represents only five fewer deaths associated with drug A. It is not easy to determine whether this difference is due to the action of the new drug or simply chance. There could just happen to be five fewer deaths in one group. However, if the death rates were 5 *versus* 40%, this represents 35 fewer deaths among 100 subjects

receiving drug A, which are unlikely to all be due to chance. Therefore, a trial of 100 patients per arm is too small if the expected difference is 5%, but large enough if the expected difference is 35%. Figure 1 illustrates how study size influences the conclusions that can be made.

### STRENGTHS

Studies with a small number of subjects can be quick to conduct with regard to enrolling patients, reviewing patient records, performing biochemical analyses or asking subjects to complete study questionnaires. Therefore, an obvious strength is that the research question can be addressed in a relatively short space of time. Furthermore, small studies often only need to be conducted over a few centres. Obtaining ethical and institutional approval is easier in small studies compared with large multicentre studies. This is particularly true for international studies.

It is often better to test a new research hypothesis in a small number of subjects first. This avoids spending too many resources, *e.g.* subjects, time and financial costs, on finding an association between a factor and a disorder when there really is no effect. However, if an association is found it is important to make it clear in the conclusions that it was from a hypothesis-generating study and a larger confirmatory study is needed.

Small studies can also make use of surrogate markers when examining associations, *i.e.* a factor that can be used instead of a true outcome measure, but it may not have an obvious impact that subjects are able to identify. For example, in lung cancer, the true end-point in a clinical trial of a new intervention is overall survival: time until death from any cause. “Death” is clearly clinically meaningful to patients and clinicians, thus if the intervention increases survival time this should provide sufficient justification to change practice. A surrogate marker is tumour response, *i.e.* complete or partial remission of the cancer. Surrogate end-points are often associated with more events, which are observed relatively soon after the intervention is administered; therefore, subjects may not require a long follow-up period. Both of these characteristics allow a smaller study to be conducted in a short space of time. Observing no change in the surrogate marker usually indicates there is unlikely to be an effect on the true end-point, thus avoiding an unnecessary large study.

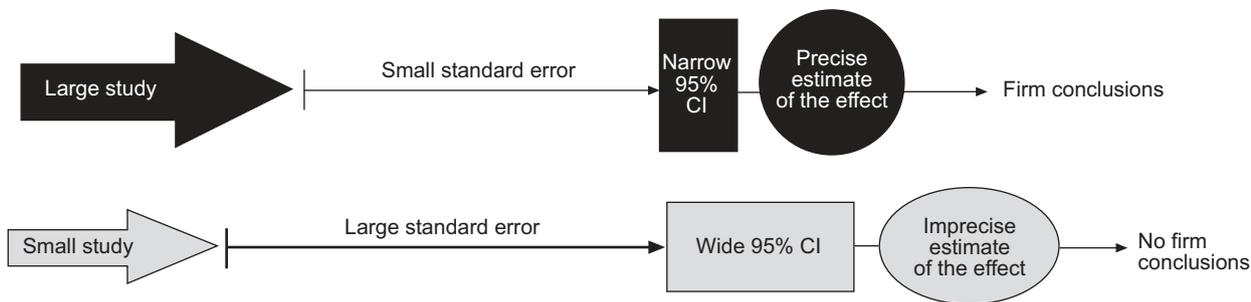
### LIMITATIONS

The main problem with small studies is interpretation of results, in particular confidence intervals and p-values (fig. 1). When conducting a research study, the data is used to estimate the true effect using the observed estimate and 95% confidence

University College London, Cancer Research UK & UCL Cancer Trials Centre, University College London, London, UK.

STATEMENT OF INTEREST: None declared.

CORRESPONDENCE: A. Hackshaw, University College London, Cancer Research UK & UCL Cancer Trials Centre, University College London, 90 Tottenham Court Road, London W1T 4TJ, UK. Fax: 44 2076799899. E-mail: ah@ctc.ucl.ac.uk



**FIGURE 1.** Schematic diagram showing how study size can influence conclusions. CI: confidence interval.

interval. Consider hypothetical clinical trials evaluating four new diets for reducing body weight (table 1). The results for diet A are clear: they are clinically important (the weight loss is large) and highly statistically significant (the p-value is very small, indicating that the observed weight loss of 7 kg is unlikely to be due to chance). The true mean weight loss associated with the new diet is estimated to be 7 kg, but there is 95% certainty that the true value lies somewhere between 6.4 and 7.6 kg. Ideally all intervals should be as narrow as this, but usually only large studies can produce such precise results. In diets B and D, the confidence intervals are also narrow, but all around a small and clinically unimportant effect so one can be fairly confident that these diets are not worthwhile. The statistically significant result for diet B is simply due to performing a very large study, but it would not justify using the new diet.

The most difficult results to interpret are those for diet C. Although the confidence interval includes zero, most of the

range is below zero and the p-value is just above the conventional cut-off value of 0.05. This is likely to be due to the study not being large enough. The data must be interpreted carefully. The lack of statistical significance does not mean there is no effect [3], because the true mean weight loss could be 3 kg, or even as large as 6.3 kg. It is better to say “there is some evidence of an effect, but the result has just missed statistical significance”, or “there is a suggestion of an effect”. There needs to be a careful balance between not dismissing outright what could be a real effect and also not making undue claims about the effect.

Another major limitation of small studies is that they can produce false-positive results, or they over-estimate the magnitude of an association. Table 2 illustrates this limitation using trials that have evaluated thalidomide in treating lung cancer [4, 5]. After the smaller studies were reported, there was much hope for thalidomide, particularly because it is administered orally. However, the large trial did not show any benefit.

Statistical significance	Clinical significance	
	Yes	No
<b>Yes</b>	Diet A: n=1,000 Mean difference -7.0 kg 95% CI -7.6 to -6.4 kg p-value <0.0001 Conclusions: Large study Large effect	Diet B: n=2,000 Mean difference -0.5 kg 95% CI -0.9 to -0.1 kg p-value=0.025 Conclusions: Large study Small effect
<b>No</b>	Diet C: n=36 Mean difference -3.0 kg 95% CI -6.3 to +0.3 kg p-value=0.07 Conclusions: Study not large enough Probably a real and moderate effect, but insufficient results to draw a reliable conclusion	Diet D: n=400 Mean difference -0.2 kg 95% CI -1.2 to +0.8 kg p-value=0.69 Conclusions: Study probably large enough Probably a small effect

Effect size is the mean difference in weight (new diet minus control). CI: confidence interval. N represents the total number of subjects in a two-arm trial.

**TABLE 2** Example of comparative evidence from phase II and III trials: thalidomide and advanced small-cell lung cancer

Two small single-arm phase II trials and a small randomised placebo-controlled trial reported consistent evidence to suggest that thalidomide could greatly increase overall survival when used with standard chemotherapy. Patients lived noticeably longer than expected.

The 1-yr survival rate in these three studies were 46 (n=25), 52 (n=30) and 49% (n=49); all higher than the expected value of 20–30%.

In the small randomised trial (based on administering thalidomide to patients who had already responded to standard chemotherapy), the median survival was 11.7 (n=49) and 8.7 (n=43) months in the thalidomide and placebo arms, respectively; which was a substantial difference.

However, a large double-blind placebo-controlled phase III trial (n=724) of thalidomide *versus* placebo was conducted. The results showed no evidence of an effect. The median survival was 10.1 and 10.5 months in the thalidomide and placebo arms, respectively. The 1-yr survival rates were 37 and 41%, respectively.

There are also limitations associated with the statistical analysis. When examining risk factors or other association, it is often necessary to allow for the effect of important prognostic factors (confounders). This is done using methods such as multivariate linear or logistic regression and Cox's regression (for survival data). However, when the number of observations is small and researchers attempt to adjust for several factors, these methods can fail to produce sensible results or they produce unreliable results.

### CONCLUSION

There is nothing precise about a sample size estimate when designing studies. It provides an approximate size of the study. It does not matter if one set of assumptions yields 100 subjects but another gives 110 because this represents only an extra five subjects per group. What is more important is whether 100 or 200 subjects are needed. There is always some guesswork involved in specifying the assumptions for sample size, particularly when determining the effect size, which is often quite different from what is observed at the end of the study.

There is nothing wrong with conducting well-designed small studies; they just need to be interpreted carefully. While small studies can provide results quickly, they do not normally yield reliable or precise estimates. Therefore, it is important not to make strong conclusions about a risk factor or trial intervention, whether the results are positive or not. Instead, data

from such studies should be used to design larger confirmatory studies. If the aim is to provide reliable evidence on a risk factor or new intervention, the study should be large enough to do so. The editorial board of the *European Respiratory Journal* often review very interesting studies but based on small sample sizes. While the board encourages the best use of such data, editors must take into account that small studies have their limitations.

### REFERENCES

- 1 Pocock SJ, ed. *Clinical Trials: A Practical Approach*. New York, John Wiley Sons, 1983.
- 2 Machin D, Campbell MJ, Fayers PM, Pinol APY, eds. *Sample Size Tables for Clinical Studies*. 2nd Edn. Oxford, Blackwell Science, 1997.
- 3 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.
- 4 Lee SM, James L, Buchler T, Snee M, Ellis P, Hackshaw A. Phase II trial of thalidomide with chemotherapy and as a maintenance therapy for patients with poor prognosis small-cell lung cancer. *Lung Cancer* 2008; 59: 364–368.
- 5 Lee SM, Rudd RM, Woll PJ, *et al*. Two randomised phase III, double blind, placebo controlled trials of thalidomide in patients with advanced non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). *J Clin Oncol* 2008; 26: 8045.