**Priya Ranganathan,
C. S. Pramesh[1],
Marc Buyse[2]**

*Department of Anaesthesiology,
Tata Memorial Centre, [1]Department
of Surgical Oncology, Division of
Thoracic Surgery, Tata Memorial
Centre, Mumbai, Maharashtra, India,
[2]Department of Biostatistics, Hasselt
University, Belgium*

Address for correspondence:
Dr. Priya Ranganathan,
Department of Anaesthesiology,
Tata Memorial Centre, Ernest Borges
Road, Parel, Mumbai - 400 012,
Maharashtra, India.
E-mail: drpriyaranganathan@gmail.
com

Common pitfalls in statistical analysis: "No evidence of effect" versus "evidence of no effect"

Abstract

This article is the first in a series exploring common pitfalls in statistical analysis in biomedical research. The power of a clinical trial is the ability to find a difference between treatments, where such a difference exists. At the end of the study, the lack of difference between treatments does not mean that the treatments can be considered equivalent. The distinction between "no evidence of effect" and "evidence of no effect" needs to be understood.

Key words: Biostatistics, bias, statistical

It is not uncommon in published literature to find authors making claims of equivalence of two treatments. However, these conclusions may sometimes be incorrect and need to be interpreted cautiously. Superiority trials compare treatments to prove that one is more effective than the other. While interpreting the results of such trials, two possibilities exist – a Type I error (finding a difference between treatments where a difference does not actually exist) and a Type II error (not finding a difference between treatments where a difference does exist). The power of the study is defined as the ability to find a treatment effect where such an effect exists.[1] Power is calculated as (1 – Type II error) and is conventionally set at 80–90%. This means that if a treatment effect does exist, the study will detect it 80–90% of the time.

However, this also means that there is a 10–20% chance that the true treatment effect may not be picked up by the study.[1]

Superiority trials may fail to show differences between treatment groups ("negative" studies) for three reasons: (a) There is genuinely no difference between the two treatments, (b) the treatment effect is smaller than accounted for in the sample size calculations or (c) the sample size is smaller than what would be required to detect a clinically important benefit. The sample size for a trial is calculated based on power, Type I error and the expected treatment effect.[1] Estimates of treatment effect are usually obtained by reviewing literature on the same topic, by doing pilot studies or as a last resort, by "guesstimates" of either the expected treatment effect or what is considered by experts in the field as a clinically relevant benefit. Since the sample size is inversely proportional to the square of the treatment effect, many researchers inflate the expected treatment effect in order to reduce the sample size and keep recruitment targets realistic. In other cases, despite having a formal sample size calculation (or equally often, without a formal calculation), investigators may choose to recruit fewer patients for logistic reasons.

The fall-out of either of the above is a failure of the study to detect a treatment effect – "no evidence of the effect" - when a true treatment effect does exist. However, this is incorrectly interpreted by many authors and readers to be the same as "evidence of no effect." For example, Sung *et al.* conducted a study to compare the efficacy of emergency sclerotherapy with octreotide infusion for variceal hemorrhage.[2] The calculated sample size was 1800 patients; an arbitrary sample size of 100 patients was settled for, while acknowledging the risk of a Type II error. Expectedly, the study failed to show any difference in outcome between the groups; however, the authors (erroneous) conclusion was "we have shown octreotide to be a safe and effective treatment for acute variceal haemorrhage and recommend its use…" To the uninitiated reader, this paper could be misinterpreted that either of the two treatments was appropriate for variceal hemorrhage – an extremely dangerous conclusion to draw from the available data. A *post-hoc* analysis showed that the study had only 5% power to detect the postulated difference.[3] In a clinical situation like acute variceal hemorrhage (which has a very high mortality without effective treatment), adoption of this recommendation could potentially cost many lives.

Lack of efficacy of a treatment (or "equivalence" of two treatments) cannot be casually derived from the negative results of a superiority trial – a trial with an "equivalence" design and a predefined equivalence margin is needed to arrive at this conclusion. "Absence of evidence of the effect" is not "Evidence of absence of effect."

REFERENCES

1. Altman DG, editor. Principles of statistical analysis. In: Practical Statistics for Medical Research. 1st ed. London: Chapman and Hall; 1991. p. 169.
2. Sung JJ, Chung SC, Lai CW, Chan FK, Leung JW, Yung MY, *et al.* Octreotide infusion or emergency sclerotherapy for variceal haemorrhage. Lancet 1993;342:637-41.
3. Altman DG. Octreotide infusion versus injection sclerotherapy. Lancet 1993;342:1486.