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

*Departments of Anaesthesiology
and [1]Surgical Oncology, Division of
Thoracic Surgery, Tata Memorial
Centre, Parel, Mumbai, Maharashtra,
India, [2]International Drug Development
Institute, Hasselt University, Belgium,
Europe*

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: Clinical versus statistical significance

Abstract

In clinical research, study results, which are statistically significant are often interpreted as being clinically important. While statistical significance indicates the reliability of the study results, clinical significance reflects its impact on clinical practice. The third article in this series exploring pitfalls in statistical analysis clarifies the importance of differentiating between statistical significance and clinical significance.

Key words: Biostatistics, confidence intervals, data interpretation, statistical

One of the common problems faced by readers (and authors!) of medical articles is in the interpretation of the word "significance." The term "statistical significance" is often misinterpreted as a "clinically important" result. The confusion stems from the fact that many people equate "significance" with its literal meaning of "importance," whereas in statistics, it has a far more restrictive connotation. This article explains the idea of the statistical significance and differentiates it from clinical relevance or importance, which is an entirely different concept. In the previous article, in this series, we looked at different ways of expressing statistical significance ("*P*" values versus confidence intervals).[1] Measures of statistical significance quantify the probability of a study's results being due to chance. Clinical significance, on the other hand, refers to the magnitude of the actual treatment effect (i.e., the difference between the intervention and control groups, also known as the "treatment effect size"), which will determine whether the results of the trial are likely to impact current medical practice. The "*P*" value, frequently used to measure statistical significance, is the probability that the study results are due to chance rather than to a real treatment effect. The conventional cut off for the "*P*" value to be considered statistically significant is of 0.05 (or 5%). What a $P < 0.05$ implies is that the possibility of the results in a study being due to chance is <5%.

In clinical practice, the "clinical significance" of a result is dependent on its implications on existing practice-treatment effect size being one of the most important factors that drives treatment decisions. LeFort suggests that the clinical significance should reflect "the extent of change, whether the change makes a real difference to subject lives, how long the effects last, consumer acceptability, cost-effectiveness, and ease of implementation".[2] While there are established, traditionally accepted values for statistical significance testing, this is lacking for evaluating clinical significance.[3] More often than not, it is the judgment of the clinician

(and the patient) which decides whether a result is clinically significant or not.

Statistical significance is heavily dependent on the study's sample size; with large sample sizes, even small treatment effects (which are clinically inconsequential) can appear statistically significant; therefore, the reader has to interpret carefully whether this "significance" is clinically meaningful. A study published in the Journal of Clinical Oncology compared overall survival in 569 patients with advanced pancreatic cancer who were randomised to receive erlotinib plus gemcitabine versus gemcitabine alone.[4] Median survival was found to be "significantly" prolonged in the erlotinib/gemcitabine arm (6.24 months vs. 5.91 months, $P = 0.038$). The $P = 0.038$ means that there is only a 3.8% chance that this observed difference between the groups occurred by chance (which is less than the traditional cut-off of 5%) and therefore, statistically significant. In this example, the clinical relevance of this "positive" study is the "treatment effect" or difference in median survival between 6.24 and 5.91 months – a mere 10 days, which most oncologists would agree is a clinically irrelevant "improvement" in outcomes, especially when considering the added toxicity and costs involved with the combination.

Most journals now endorse the use of the CONSORT statement for reporting of parallel-group randomized trials, which emphasizes the need for reporting of the estimated effect size and its precision (such as 95% confidence interval) for each primary and secondary outcome.[5] Readers should bear in mind that interpretation of study results should take into account the clinical significance by looking at the actual treatment effect (with confidence intervals) and should not just be based on "P" values and statistical significance.

REFERENCES

1. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals. Perspect Clin Res 2015;6:116-7.
2. LeFort SM. The statistical versus clinical significance debate. Image J Nurs Sch 1993;25:57-62.
3. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. Aust Crit Care 2010;23:93-7.
4. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2007;25:1960-6.
5. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152:726-32.