EuroIntervention 2015;10:1368-1369

Tools and Techniques - Statistical: It's statistically significant, but is it clinically relevant?

Ron T. van Domburg*, PhD; Isabella Kardys, PhD; Matti Lenzen, PhD; Sara Baart, Msc; Eric Boersma, PhD; Sanne E. Hoeks, PhD

Erasmus Medical Center, Rotterdam, The Netherlands

Statistical hypothesis testing is a key element of evaluating the results of medical research and may result in a so-called "significant" finding. However, the value of "statistical significance" must not be overstated, as it describes only one aspect of the results of a study. The interpretation of medical research data from the perspective of clinical relevance, though far less emphasised, is of equal importance. The purpose of this article is to provide clinicians with an outline of the two essential concepts of statistical significance and clinical relevance, or clinical significance as others prefer. We aim to increase awareness of the subject, and we do not intend to provide a complete overview, which can be found elsewhere in the literature[1].

By applying a statistical hypothesis test, the investigator aims to quantify the evidence against the so-called null hypothesis, which usually states that there is no difference (i.e., the effect is "null"). The appropriate statistical test provides the investigator with a probability value (p-value), which indicates the strength of the evidence against the null hypothesis. Mathematically, the p-value is the probability of obtaining the observed effect when the null hypothesis is actually true. For example, if a statistical test shows a p-value of 0.30, then the probability is 30% that the observed difference occurs, whereas in reality there is no true difference. Statisticians explain this phenomenon by the concept of "random sampling error". According to this concept, the study patients constitute a random sample out of a large population. A p-value of 0.30 indicates that, just by natural variation (or "chance"), an effect might be found in three of 10 such samples. If the p-value is very small, then the study data are compatible with a true effect above chance, and the null hypothesis will be rejected. The effect is considered "statistically significant". Mostly, a threshold p-value of 0.05 (5%) is used to declare statistical significance.

However, statistical significance does not automatically mean that the observed effect is also clinically relevant and vice versa. A p-value indicates in an objective way how sure we are that an observed effect is true, but provides no information on the magnitude of that effect. Clinical relevance can be conceptualised as a difference that is large enough to justify clinicians changing the standard of care. Therefore, when evaluating the results of a study, one must address both the statistical significance and the clinical relevance of the findings.

Assume a (hypothetical) placebo-controlled randomised clinical trial to investigate the effect of a new drug for arterial hypertension in a sample of 2*10,000 patients. At the end of the trial, the mean change in systolic blood pressure of the patients randomised to the active treatment turns out to be on average –5 mmHg, compared with –4.5 mmHg in placebo, i.e., a mean difference of –0.5 mmHg.

*Corresponding author: Department of Cardiology, Erasmus Medical Centre, 's- Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands. E-mail: r.vandomburg@erasmusmc.nl

DOI: 10.4244/EIJV10I11A235

The p-value is 0.001, which is much less than 0.05, and, consequently, the null hypothesis of no difference in mean blood pressure reduction can be rejected. However, from a clinical point of view, one might question the clinical relevance of an average effect on systolic blood pressure of –0.5 mmHg. The implementation of the new drug in all future patients who fulfil the trial inclusion and exclusion criteria is not self-evident.

An example from the literature is the RIO-Rimonabant trial[2], in which 1,047 overweight or obese patients with type 2 diabetes were randomised to rimonabant, a drug intended to reduce body weight, or placebo. After one year of follow-up, the mean weight loss was statistically significantly larger in the patients randomised to rimonabant than in placebo (placebo: –1.4 kg; rimonabant: –2.3 kg, p=0.01). Thus, a mean difference appeared of 0.9 kg in favour of the new drug in a sample with an average body weight of 97 kg. One could question here the clinical relevance of this absolute 1% reduction.

The GUSTO IIb trial randomised 1,138 patients presenting within 12 hours of acute myocardial infarction to primary angioplasty (pPCI, n=565) or accelerated thrombolytic therapy with recombinant tissue plasminogen activator (t-PA, n=573). Mortality at 30 days (pPCI: 5.7%, t-PA: 7.0%; p=0.37) was not statistically significantly different[3]. A meta-analysis was performed by Keeley et al, showing with statistical significance that pPCI was better than thrombolytic therapy at reducing overall short-term death (6.9% vs. 9.3%; p=0.0002)[4]. Thus, although the decrease in mortality in the GUSTO IIb trial was not statistically significant and caused no change in treatment, the effect size was big enough to consider it clinically relevant. The meta-analysis had more power to show that this effect was also statistically significant and eventually caused a massive move from thrombolytic therapy to pPCI.

Thus, when evaluating the validity of a study in cardiovascular literature, the reader must consider both the clinical and the statistical significance of the study results. Successful study planning requires an explicit definition of the clinically meaningful primary study endpoint, an estimate of the proposed treatment effect, and an estimate of the sample size necessary to demonstrate the difference of interest. Understanding the direct relationship between sample size and power is crucial for the critical judgement of any study conclusion. An inadequate sample size will fail to detect clinically important differences, whereas an excessively large sample size may show significant differences which are far from clinically relevant.

In conclusion, a good notion and awareness of both statistical significance and clinical relevance is crucial for a correct interpretation of clinical trial results.

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

1. Kaul S, Diamond GA. Trial and error. How to avoid commonly encountered limitations of published clinical trials. *J Am Coll Cardiol.* 2010;55:415-27.

2. Scheen AJ, Finer N, Hollander P, Jensen MD, Van Gaal LF; RIO-Diabetes Study Group. Efficacy and tolerability of rimonabant in overweight or obese patients with type 2 diabetes: a randomised controlled study. *Lancet.* 2006;368:1660-72.

3. The Global Use of Strategies to Open Occluded Coronary Arteries in Acute Coronary Syndromes (GUSTO IIb) Angioplasty Substudy Investigators. A clinical trial comparing primary coronary angioplasty with tissue plasminogen activator for acute myocardial infarction. *N Engl J Med.* 1997;336:1621-8.

4. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet.* 2003;361:13-20.