

Statistical Significance: Is it really that significant?

Zulfiqar Ahmed

Department of Anesthesiology, Children's Hospital of Michigan, Wayne State University, Beaubien.

The demands of evidence based medical practice include capability of critical appraisal of peer reviewed journal articles (as part of peer review). In order to understand these articles, some knowledge of statistical methodology is essential. Often results are reported to be statistically significant or not. Basically defined, when the result of a scientific experiment is statistically significant ($p < 0.05$), it means that there is a 95% or higher chance that this result is not by any error or coincidence. This translates to minimizing (not eliminating) bias, chance, errors and luck.

The magical "p value" in the results section of scientific publications is highly revered and often heralded as a standard. P value below 0.05 is often mentioned while submitting a paper to a journal or recommending/selling certain therapy for medical or surgical diagnosis and or treatment. This magic bullet is coming under increasing scrutiny. The statistically significant treatment effect may be 10% to 200% but this difference may or may not matter when it comes to real time patient management. For example:

"Patients receiving Clonidine showed a statistically significant, but clinically trivial, decrease in heart rate (116 ± 32 versus 112 ± 28) and mean arterial blood pressure (92 ± 18 versus 87 ± 21) during the recovery period".¹

Hence it is misnomer that if the results are statistically significant then the therapy is of clinical significance.

In his editorial Kain et al² discuss this issue. According to Kain, with a large enough sample size, statistically significant results may be achieved while being a clinically ineffective therapy (i.e, small size effect). At the same time, it is incorrectly assumed that the higher the statistical significance ($p=0.0004$) vs ($p=0.03$), the clinical effect is proportionally higher. This will be correct if the sample size is the same. Hence Kain suggest that, while examining a scientific report, the following three considerations should be made:

- 1) Are the findings solely a result of chance occurrence? (i.e., statistical significance)
- 2) How large is the difference between the primary

endpoints of the study groups? (i.e., treatment impact and sample size)

3) Is the clinical difference of the therapy meaningful to the patient?(i.e., clinical significance)

We have to remember that the magical value of $p=0.05$ was designated arbitrarily and has no proven significance. In other words, this chance we have imposed upon ourselves to accept or reject certain results. Statistical significance may be considered borderline by some when $p=0.05$ especially when the authors are paying attention to the risk estimation aspect rather than hypothesis testing aspect of the statistical analysis.

In order to tackle the limitations of p value, "confidence interval" and "effect size" are increasingly been used.

Confidence interval (CI) is considered reciprocal to statistical significance. That is $p < 0.05$ is the same as having a 95% CI that does not overlap zero. CI can some times be used to estimate the size of the difference as well as indicating the presence or absence of statistical significance. We should also add the case of checking significance by 95% CI when it does not contain.¹

Effect sizes measures the magnitude of difference between the two groups. Hence in other words, it can quantify the amount of difference between the two groups. It does not simply rely on sample size but instead relies on the strength of the intervention.

As a conclusion, evidence based medical practice is a complicated yet essential tool in modern medical practice. The peer-reviewed articles are to be read with care. One should gain proficiency in statistical methodology and critical appraisal of literature. Evidence based medicine should be an essential part of medical education and training.

References

1. Tesoro S, Mezzetti D, Marchesini L, Peduto VA. Clonidine treatment for agitation in children after sevoflurane anesthesia. *Anesth Analg* 2005; 101: 1619-22.
2. Kain ZN. The legend of the P value. *Anesth Analg* 2005; 101: 1454-6.