EDITORIAL

Understanding Grows When "P Value" Is Replaced with the "Confidence Interval" in Medical Toxicology

REZA AFSHARI[*]

Editor-In-Chief

Addiction Research Centre, Mashhad University of Medical Sciences, Mashhad, Iran

While working on the articles submitted to Asia Pacific Journal of Medical Toxicology, I realised that statistical approach and notably the concept of "*P value*" is somehow poorly understood and has taken surrealistic credibility within the community of medial toxicologists. In this editorial, common flaws on the subject have been summarized.

P value, which is ranged from 0 to 1, is the probability of observing the result found from a study at least as extreme as the one that was actually observed when the null hypothesis (H_0) is not rejected (1). The famous arbitrary cut-off point of 0.05 (alpha level) for separating the probable from the improbable was suggested by Sir Ronald Fisher in 1926 and still stands to date (2). This level was selected to distinguish real differences from random variability, and to minimise the contribution of random variability (3). Despite its advantages, P value realm of importance has recently been shaken in clinical studies. A significant P value means that there is a sufficient evidence to reject the null hypothesis, when the statistical design has been formulated before conducting the study (4). The P value ability to determine the truth about measured effects; however, is usually quite limited (5), as P value is confounded by a number of factors. If the sample size increases or the standard deviation decreases, the P value becomes smaller for the same mean difference (6). Any difference in very large samples would eventually become significant unless zero difference exists.

P value has been widely misinterpreted in medical toxicology, especially when rare outcomes (number of deaths in each group) in very large samples (number of poisonings) are studied. P value is just hypothesis testing which makes it less interesting in clinical settings (7). Interpretation of P value is rather an area of philosophical uncertainty (5).

The following notions are incorrect - as are commonly stated - to assume that P value (a) demonstrates a null hypothesis is true or false (b) excludes false significance (a P value of 0.01 means that repeated experiments might even lead to a non-significant P value) (c) likely reflects a causal relationship. In addition, P value provides no insight into the clinical relevance. It cannot tell how large or small the observed effect was (1). P value does not provide usefulness compared to the magnitude of differences between two findings (6). No theoretical basis exists for 5% probability or less to be strong enough evidence to reject all null hypotheses (5). P value should not be treated as "absolute". The actual difference between 0.049 and 0.051 is very limited.

When multiple independent hypotheses are tested in one study the risk that one of the significant results is false positive increases. Comparison of P values across repeated studies with similar design is also not correct. It should be kept in mind that a low P value is not equivalent to a high level of precision (6). Reviewers should be aware that P value can be manipulated. By changing the arbitrary cut-off points in continuous data, statistical analysis may provide more than one answer, implying "data interpretation is more an art than science" (8). Authors may also push a "near significant P value" to a level that is considered significant (9), as non-significant results might be subjected to publication bias during the review process.

Confidence interval (CI) is a type of interval estimate of a population parameter. CI is used to indicate the reliability of an estimate. Unlike P value, CI gives a range of outcome values, which is mathematically constructed using the standard deviation as the principal determinant of width and a user-determined coverage percentile (usually 95%) (5). CI provides information about a range in which the true value lies with a certain degree of probability. It also reveals the direction and strength (effect size) of the demonstrated effect. CI is independent from the sample size. This enables conclusions to be drawn about both the statistical plausibility and clinical relevance of study findings (4). CI is easy to understand and gives a much better and direct insight into the observed clinical results and so it is more interesting than P value (10). CI is a better indicator of precision. A given finding may have a low P value but wide confidence intervals, suggesting low precision and wide variance in potential results (6). It shows what effects are likely to exist in the population, and means values excluded from the CI are thus not likely to exist in the population (10).

Despite all these, how the products of a research should be presented? It seems that descriptive statistics should be focused. Analysis of data and the need for boundaries in our studies; however, are understandably important. In the past, P value has been successfully used to minimise the potential incorrectness in reporting of the findings. It is fanatic to wish "P value" fade away for now. It keeps helping us in making decisions. CI has many advantages over P value. Hence, widespread misinterpretation of P values in medical toxicology substantiates asking for CI in our articles.

REFERENCES

1. Pandis N. The P value problem. Am J Orthod Dentofacial Orthop 2013 Jan;143(1):150-1.
2. Maher BHDuffy ME, Munroe BH, Jacobsen BS. Key principles of statistical inference. In: Maher BH, ed. Statistical Methods for Health Care Research. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p.73-105.
3. Forbes DA. What is a p value and what does it mean? Evid Based Nurs 2012 Apr;15(2):34.
4. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009 May;106(19):335-9.
5. Youngquist ST. Part 19: What is a P value? Air Med J. 2012 Mar-Apr;31(2):56-71.
6. Cook C. Five per cent of the time it works 100 per cent of the time: the erroneousness of the P value. J Man Manip Ther 2010 Sep;18(3):123-5.
7. Gupta SK. The relevance of confidence interval and P-value in inferential statistics. Indian J Pharmacol 2012 Jan;44(1):143-4.
8. Dalman MR, Deeter A, Nimishakavi G, Duan ZH. Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics 2012 Mar 13;13 Suppl 2:S11.
9. Gadbury GL, Allison DB. Inappropriate fiddling with statistical analyses to obtain a desirable p-value: tests to detect its presence in published literature. PLoS One 2012;7(10):e46363.
10. Ranstam J. Why the P-value culture is bad and confidence intervals a better alternative. Osteoarthritis Cartilage 2012 Aug;20(8):805-8.

[*] Correspondence to: Reza Afshari, MD, PhD, MSc. Addiction Research Centre (ADRC), Mashhad University of Medical Sciences, School of Medicine, Mashhad, Iran.
Tel/Fax: +98 511 859 8973, E-mail: afsharir@mums.ac.ir