

## Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results

### ABSTRACT

It is incredibly essential that the current clinicians and researchers remain updated with findings of current biomedical literature for evidence-based medicine. However, they come across many types of research that are nonreproducible and are even difficult to interpret clinically. Statistical and clinical significance is one such difficulty that clinicians and researchers face across many instances. In simpler terms, the P value tests all hypothesis about how the data were produced (model as whole), and not just the targeted hypothesis that it is intended to test (such as a null hypothesis) keeping in mind how reliable are the of the research results. Most of the times it is misinterpreted and misunderstood as a measure to judge the results as clinically significant. Hence this review aims to impart knowledge about "P" value and its importance in biostatistics, also highlights the importance of difference between statistical and clinical significance for appropriate interpretation of research results.


**Key words:** Bias; biostatistics; clinical significance; research design; sample size; statistical significance

### Introduction

Currently, in the publish or perish era where most of the researches are judged based on their statistically significant findings, it is often difficult for young researchers to interpret the correct findings of the research. The recent development of high-speed and more sophisticated computing power, utilizing high-end computers and statistical software packages, has resulted in a significant increase in the use of statistical methods, tests for hypothesis testing and reporting to the health literature. Unluckily, the appropriate interpretation of research results from the clinical point of view not received similar interest.<sup>[1]</sup> This imbalance from decades to determine the actual importance of statistical and clinical significance and publication of such results in reputed

indexed journals had led researchers to consider statistically significant results also as a clinically important one. It is essential to understand that publications in reputed indexed journals do not indicate that appropriate study design or statistics methods were used. This dilemma of the young researchers creates obstacles in their clinical decision-making and ultimately affects their role in Evidence-based practice.<sup>[2]</sup>

Researchers must realize that a clinical study is valuable and is of importance to clinical practice when the results are appropriately interpreted. Every year hundreds of studies and clinical trials are conducted to test different hypothesis. These trials are entirely dependent on appropriate statistical

Access this article online	
<b>Website:</b> www.saudija.org	<b>Quick Response Code</b> 
<b>DOI:</b> 10.4103/sja.sja_158_21	

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Sharma H. Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Saudi J Anaesth 2021;15:431-4.

### HUNNY SHARMA

Department of Community and Family Medicine, All India Institute of Medical Sciences, Raipur, Chhattisgarh, India

**Address for correspondence:** Dr. Hunny Sharma, MD 264, Phase 4, Near AIIMS Residential Complex, Kabir Nagar, Raipur - 492 099, Chhattisgarh, India. E-mail: smilerecoverydc@gmail.com

**Submitted:** 04-Mar-2021, **Revised:** 04-Mar-2021, **Accepted:** 04-Mar-2021, **Published:** 02-Sep-2021

tests to assess whether new therapies or treatment protocol are better in clinical practice as compared to the usual approach or methods. Researchers should understand what is the importance of both statistical and clinical significance.<sup>[3]</sup>

When looking from a clinical point of view, the statistically significant difference among groups is not of prime importance. If a well-conducted study shows a difference in treatment options within two groups, it is of prime importance to know whether that difference is of clinically importance or not.<sup>[4]</sup> Since sample size and measurement variability can easily influence the statistical results, a nonsignificant outcome does not imply that the new therapy or treatment protocol is not clinically useful.<sup>[5,6]</sup>

Hence this review aims to impart knowledge about “*P*” value and its importance in biostatistics, also highlights the importance of difference between statistical and clinical significance for appropriate interpretation of research results.

#### What does *P* value infer?

In simpler terms, the *P* value tests all hypothesis about how the data were produced (the whole model), not just the targeted hypothesis that it is intended to test (such as a null hypothesis).<sup>[7]</sup>

The *P* value is the likelihood that if every model assumption, including the test hypothesis, were correct, the chosen test statistic would have been at least as large as its observed value.<sup>[7]</sup>

The most common threshold value for the “*P*” we find in biomedical literature is 0.05 (or 5%), and most often the *P* value is distorted into a dichotomous number where results are considered “statistically significant” when *P* falls on or below a cut-off (usually 0.05) and otherwise its declared “nonsignificant”.<sup>[7]</sup>

#### Why are “*P*” values not enough?

According to Ron Wasserstein, ASA’s executive director, the *P* value was never meant to substitute the scientific reasoning, which is of greater interest. *P* value, which is a number whose value can range from zero to one in relation to a threshold value, represents the probability that the difference between the groups is not by chance. A well-reasoned and scientifically driven explanation will always remain the basis of reporting scientific outcomes.<sup>[8]</sup>

#### On what factors does the “*P*” value depend?

It should be borne in mind that the “*P*” value only represents that to what extent the data are inconsistent or incompatible with a given specific statistical model (i.e., null hypothesis).

Hence it only aims to accept or reject the null hypothesis rather than focusing on the research hypothesis. From a statistical point of view, it measures the strength of evidence against the null hypothesis.<sup>[9]</sup>

With the advancement in biostatistics, it is now clear that the “*P*” value can easily be affected by various factors like sample size, the magnitude of the relationship and error. Each of these factors can work independently or in combination to distort the study findings based on “*P*” values.<sup>[10]</sup>

#### (1) Effect of error on “*P*” values

In general, two types of errors that is, systematic and random error effects the “*P*” value.

“Systematic errors,” that is, “Non-random errors” of certain significant magnitude distorts the research results towards a specific direction or can result in altered observed association in either direction. This type of error generally occurs when a single examiner takes the measurement leading to an unintended bias of deviating the research results to his/her expectations or may also result from modification of the measuring technique. Hence, Systematic error is a systematic flaw in the measurement of a variable due to methodological error leading to underestimation or overestimation of measurements. The extent of systematic errors can be determined by re-examination and re-measurement of a certain sufficient number (i.e., 20%, not always applicable) of individuals again by material and method used in the agreement. Some statistical tests like paired t-test, the intraclass correlation and the Bland-Altman method can also help in the determination of systematic errors.<sup>[10-12]</sup>

A “random error” is defined as a variability of the data which cannot be explained. Random errors of high magnitude means trouble in reproducibility of the measurements, which may result in questionable methodology and questionable examiners’ ability. This occurs randomly across the population, ultimately distorting the results. Random errors can be minimized by taking a large number of samples or measurements. Let us understand this by taking an example of measuring Mid-Upper Arm Circumference (MUAC) of the population. While measuring the MUAC of each individual in the population, random error may exhibit itself in the form of random MUAC among individuals that is, less or more MUAC measured as compared to the actual measurement. This may be a result of how the tape was held while taking the measurement, at what position it was measured (ideally midway between the olecranon process and the acromion), and who was the researcher who took the measurement. Random error can be reduced by incorporating a large number of samples or measurements that is, the more study

participants are included in these measurements, the smaller the effect of random error will become.<sup>[10]</sup>

### (2) Effect of sample size on “P” values

It is well known that the *P* value depends on the sample size to a vast extent. More the sample size less will be the variability of the measurement or data, and more precise will be the measurement for the study population. With an increase in sample size, the magnitude of random error decreases and the study is more likely to find a significant relationship if it exists.<sup>[10]</sup>

### (3) Effect of magnitude of relationship between groups on “P” values

*P*-value also relies on the magnitude of difference or relationship between the groups compared. In simpler terms, if the magnitude of difference between the two groups is more substantial, then it will be easy to detect and will have a small *P* value.<sup>[10]</sup>

### What are the American Statistical Association (ASA) principal statements on statistical significance and *P* values?

ASA on 8<sup>th</sup> March 2016, in the event of the growing concern of misuse and misinterpretation of *P* values, gave six principal statements to improve conduct and interpretation of quantitative research and increase research reproducibility. The six principal statements issued regarding significance and *P* value which are as follows:

1. *P*-value shows the extent of incompatibility of the data with the stated statistical model.<sup>[8]</sup>
2. *P*-value is neither the measure of the probability of the studied hypothesis being true nor is the representation of the probability that study data were produced by random chance alone.<sup>[8]</sup>
3. It is extremely important to note that any business model, policy decision, or conclusion related to any scientific study or experiment should not be based on the *P* value and merely on the fact whether it passes a specific threshold or not.<sup>[8]</sup>
4. It is the moral duty of the authors and researchers to report the research or experimental findings to its full extent and with transparency.<sup>[8]</sup>
5. A *P* value is neither represents the importance of research results nor is the representation of the effect size of the study.<sup>[8]</sup>
6. *P*-value does not give a sufficient measure of evidence regarding a model or “hypothesis”.<sup>[8]</sup>

### What are clinically significant outcomes?

The term “clinically significant” can be used for the researches in which clinically relevant results or outcomes are used to

assess the effectiveness or efficacy of a treatment modality. When used the term “clinically significant” findings are those who make the patient improves the quality of life and makes him/her feel, function well.<sup>[13]</sup>

Clinically significant findings are those which improve medical care resulting in the improvement of individual’s physical function, his/her mental status, and ability to engage in social life. The term improvement of quality of life in medical care deals with both subjective and objective terms. Here the term objective deals with improvement in performance status, duration of remission of disease, and prolongation of life-span, while subjective improvement in quality-of-life deals with improved mood, attitude, physical and social activity, feeling of general wellbeing, and the alleviation of distressing symptoms like pain, weakness, and discomfort.

Since statistical significance results do not necessarily mean that the results are clinically relevant and lead to improvement in the quality of life of the individuals. Hence, many outcomes can be statistically significant but not clinically relevant in a clinical point of view. Hence, clinicians and researchers should give importance to both statistical and clinical significance.<sup>[13]</sup>

A clinically relevant intervention justifies the effects which over-benefits the associated costs, harm, and the inconveniences caused to the individuals for whom it is targeted. The main difference between statistical and clinical significance is that the clinical significance observes dissimilarity between the two groups or the two treatment modalities, while statistical significance implies whether there is any mathematical significance to the carried analysis of the results or not.

Different studies can have a similar statistical significance but may differ significantly in clinical significance. Let’s consider an example of two different chemotherapy agents for cancer. The first study estimates to increase the survival of treated patient with Drug A (Less Expensive than usual chemotherapeutic agents) by five years ( $P = 0.01$ ) and alpha being 0.005, similarly a Second study utilizing Drug B (Expensive than usual chemotherapeutic agents) estimates to increase the survival of treated patient by mere five months ( $P = 0.01$ ) and alpha being 0.005. In both cases, the statistical test is significant, but Drug B only increases the survival by only five months which is not clinically significant as compared to Drug A which increases survival by five years, nor useful in terms of cost-effectiveness and superiority when compared to already available chemotherapeutic agents.<sup>[14,15]</sup>

## Conclusion

Hence from the above description of statistically significant and clinically significant results, it is clear that both the notations have the importance of their own. The statistically significant results may not be of clinical importance, vice versa the results which are of clinical importance may not be statistically significant. It is high time now that the researchers, journal editors, and readers should take a keen interest in looking beyond the threshold “*P*” value and also consider the results from a clinical point of view rather than just assessing the worth of research by considering the “*P*” values. All the researchers should also take into account the design, sample size, effect size of the study, bias incorporated, and reproducibility of the study while analyzing the study results. An aware researcher with a logically and critically thinking mind is in the best position to evaluate research results and thereby applying them to practice evidence-based medicine. Logically, discussion of the clinically significant research results will increase discussion and understanding of the new treatment modalities and will help in the upliftment of evidence-based practice.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References

1. Arnold LD, Braganza M, Salih R, Colditz GA. Statistical trends in the Journal of the American Medical Association and implications for training across the continuum of medical education. *PLoS One* 2013;8:e77301.
2. Grabowski B. “*P*<0.05” might not mean what you think: American statistical association clarifies *P* values. *J Natl Cancer Inst* 2016;108:djw194.
3. Page P. Beyond statistical significance: Clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther* 2014;9:726-36.
4. Bhandari M, Joensson A. Part IIC: Understanding Treatment Effects. *Clinical Research for Surgeons*. Thieme Electronic Book Library. 333, Seventh Avenue, New York, NY 10001, USA: Thieme Publisher; 2009. p. 139-44.
5. Sullivan GM, Feinn R. Using effect size-or why the *P* value is not enough. *J Grad Med Educ* 2012;4:279-82.
6. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1:50-7.
7. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, *et al*. Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *Eur J Epidemiol* 2016;31:337-50.
8. “*P*-Values under Question.” American Psychological Association, American Psychological Association. Available from: <https://www.apa.org/science/about/psa/2016/03/p-values>. [Last accessed: 30 December 2020].
9. Nahm FS. What the *P* values really tell us. *Korean J Pain* 2017;30:241-2.
10. Thiese MS, Ronna B, Ott U. *P* value interpretations and considerations. *J Thorac Dis* 2016;8: E928-31.
11. Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod* 1983;83:382-90.
12. Cañado RH, Lauris JR. Error of the method: What is it for?. *Dental Press J Orthod* 2014;19:25-6.
13. Armijo-Olivo S. The importance of determining the clinical significance of research results in physical therapy clinical research. *Braz J Phys Ther* 2018;22:175-6.
14. Tenny S, Abdelgawad I. Statistical Significance. [Updated 2019 May 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK459346/>.
15. Man-Son-Hing M, Laupacis A, O’Rourke K, Molnar FJ, Mahon J, Chan KB, *et al*. Determination of the clinical importance of study results. *J Gen Intern Med* 2002;17:469-76.