Editorial

The truth behind numbers

The essence and aim of every research is to find the truth. As doctors and researchers, it is our duty to pursue, persist and achieve this goal. Sound research enables clinicians to improve their quality of clinical care, thereby reducing the burden of disease. We as researchers, authors, reviewers and editors, have a huge responsibility. What we research, write, review and publish finally alters the standard of care delivered to our patients universally.

STATISTICAL SIGNIFICANCE

The four building blocks of sound research can be given as follows: Establishing the research hypothesis; proper study design; appropriate data collection; and, statistical assessments that include hypothesis testing.[1] Pearson and Neyman developed the theory of hypothesis testing in the late 1920s.[2] The most common form of research compares different materials or techniques in order to find whether a truly significant difference exists between the tested groups, which is ascertained with the help of statistical tools. However, it is important to understand that statistical significance cannot always be extrapolated as a clinically significant finding. This is because most of the statistical tool being employed as well as the used study designs may have certain inherent limitations. Hence, if the researcher is not careful during the design of a study then he/she might end up with results which are statistically significant; however, with poor clinical relevance.

CLINICAL SIGNIFICANCE

The definition of clinical significance varies depending on the specific clinical field being addressed, the size of the effect, the measurement used to evaluate a therapy and the clinical importance of the findings.[3] According to Greenstein,[4] *Clinical Significance* denotes a change that may alter how a clinician will treat a patient, and this value judgment can vary depending on the situation. To arrive at a conclusion that a result is clinically significant, the finding must be clinically meaningful and statistically significant.

The objective of our research should be clearly focused on the clinical endpoint of application. Whether it is a complex randomized controlled trial of alternate restorative or endodontic therapies or a material science study related to shear bond strength analysis; the results of our research should not only be statistically significant but also clinically relevant. In order to be able to avoid statistical errors and correctly achieve results for events which are clinically significant, a researcher should be aware of the following:

1. Standard deviation
2. Effect Size
3. Level of statistical significance
4. Power of the test

STANDARD DEVIATION

This refers to the variation of individual samples from the mean. This value is obtained from the mean values which have been calculated for a particular variable in previous studies. For example, if the aim of our study is to determine the bond strength of a new dentin bonding agent, then we use the standard deviation (SD) of a previously published study with a similar aim, after a thorough process of critical appraisal of the study's methodology. Critical appraisal for this process becomes imperative because the SD obtained from a particular study would determine the truth inferred from our study. A smaller SD from the mean indicates lesser variation from the mean and indicates the behavior of each sample with regards to bond strength.

EFFECT SIZE

Effect size is the expected targeted difference between the test groups, and is determined by the researcher prior to the investigation based on preliminary data or literature review. The smaller the expected difference (effect size), the larger the sample size needed to provide the number of events necessary for a comparison between the test groups. On the other hand, the larger the expected difference (effect size), the smaller will the required sample size be. However, as the sample size based on the effect size may be too large to achieve, researchers sometimes choose a larger effect size than one would normally expect in order to reduce the sample size and minimize the expenditure of time and resources. Further, if the targeted difference between the test groups is larger than the true difference, the study may fail to conclude a difference between the two groups when a smaller, yet meaningful, difference exists. For example, let us consider a research aiming to compare the bond strength efficacy of a newer dentin bonding agent with an established bonding agent having proven shear bond strength of 20 MPa. The effect size in this case has to

Access this article online

Quick Response Code:

Website: www.jcd.org.in

DOI: 10.4103/0972-0707.87191

Table 1: Type I and Type II errors during hypothesis testing[5]

Truth	Study findings	
	Null hypothesis is not rejected	Null hypothesis is rejected
Null hypothesis is true	True negative	Type I error (Alpha) (False positive)
Null hypothesis is false	Type II error (Beta) (False negative)	True positive

be ascertained keeping in mind the clinical correlation. Ideally, you have to arrive at a consensus of what increase in the bond strength offered by the newer material would make it clinically a better option to be used, and thus make the dentists adopt it over the other established brand. In this case, it could be an increase of 4 MPa, which would mean 24 MPa, or an effect size of 20%. Thus, the sample size for this research question would be calculated on the basis of this effect size, and not arbitrarily.

Level of statistical significance

The science of statistics finds it easy to disprove things than to prove anything. Hence, in order to assess whether the difference between the groups is due to a genuine difference or due to chance alone; we set a *null* hypothesis that there is no difference between the test groups. Following this, we can go ahead to determine the probability (*P value*), which is a predetermined significance level. If the result of the comparison between the tested groups is below this chosen *p* value then the null hypothesis is rejected (that there is a statistically significant difference between the groups), and if the results produce findings *above* the *P* value, the null hypothesis is accepted (that there is no statistically significant difference between the groups).

Hypothesis testing is not a perfect tool and can lead to erroneous conclusions if the study design is not carefully planned. The two most common errors during hypothesis testing are the Type I Error and Type II Error [Table 1].[5]

Type I error or false positive

It is an error in concluding that there is a difference between the groups where as in reality no such difference exists. In other words, a Type I Error is a false positive error of rejecting the null hypothesis; thereby assuming that there is a statistically significant difference between the test groups. [Table 1] It is of clinical significance to ensure that the result of a research does not conclude that an ineffective group is beneficial. A numerical boundary, which is set to limit the likelihood of making this error, is known as Alpha, α or Type I error probability. By precedence, the level of Type I error probability is set at 5% or 0.05. However, in certain situations, it is recommended to set a *P* value as low as 0.01 or 0.001. The other clinically relevant information is the fact that when a study does not demonstrate a statistically significant difference between the groups, it does not mean that there is equivalence between the groups. This is however a commonly overlooked aspect amongst most readers of scientific articles.

Power of the test

Type II error or false negative

It is an error in concluding that there is no difference between the groups where as in reality, such a difference genuinely exists. In other words, a Type II Error is a false negative error of accepting the null hypothesis; thereby assuming that there is no statistically significant difference between the test groups. [Table 1] The power of a test is defined as: 1 minus the probability of Type II error).[5] The Type II error is concluding at no difference (the null is not rejected) when in fact there is a difference, and its probability is named b. Therefore, the power of a study reflects the probability of detecting a difference when this difference exists. It is also very important for dental research that studies are planned with an adequate power so that meaningful conclusions can be arrived at, even if no statistical difference has been shown between the groups being compared. Increasing the power of a study ensures that we minimize the risk of Type II errors and increase the chances to detect a difference between the groups when it exists. By precedence, the power of a study is usually set at 80% and in certain cases even as high as 90%. It is evident that the sample size increases as we increase the power of a study.

Sample size

The number of test specimens or the sample size is one of the most critical parameters to ensure clinically significant results from the statistical analysis of the test results. This assessment has to be done prior to the commencement of the study in consultation with a bio-statistician. A thorough review of the literature is recommended in order to establish the appropriate effect size. Sample size calculation should then be based on the effect size, appropriate *P* value and as high a power of the *test* as feasible. In a nutshell, dental research is for the benefit of our patients. Research, which does not fulfill this, cannot be termed significant. Statistics constitutes an integral component of research analysis. However, the role of statistics is best summarized by Altman[6] who famously quoted *"Statistical analysis allows us to put limits on our uncertainty, but not to prove anything."*

Significantly Yours'

Velayutham Gopikrishna

Editor in Chief,
Journal of Conservative Dentistry,

Professor, Department of Conservative Dentistry and Endodontics, Thai Moogambigai Dental College, Chennai, India
E-mail: hi_gopikrishna@hotmail.com

REFERENCES

1. Kramer MS. Clinical epidemiology and biostatistics. New York: Springer-Verlag; 1988. p. 137-45.
2. Pearson J, Neyman ES. On the use, interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika 1928;20A:175-240.
3. Lindgren BR, Wielinskyi CL, Finkelstein SM, Warwick WJ. Contrasting clinical and statistical significance within the research setting. Pediatr Pulmonol 1993;16:336-40.
4. Greenstein G. Clinical versus statistical significance as they relate to the efficacy of periodontal therapy. J Am Dent Assoc 2003;134:583-91.
5. Biau DJ, Kerneis S, Porcher R. The importance of sample size in the planning and interpretation of medical research. Clin Orthop Relat Res 2008;466:2282-8.
6. Altman DG. Practical statistics for medical research. london, UK: Chapman and Hall; 1991.

How to cite this article: Gopikrishna V. The truth behind numbers. J Conserv Dent 2011;14:327-9.