ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 10, Issue, 07(E), pp. 33658-33659, July, 2019

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

NEED OF REPORTING EFFECT-SIZE MEASURES IN CLINICAL RESEARCH

Dinesh Kumar Bagga

Department of Orthodontics School of Dental Sciences Sharda University Knowledge Park III Greater Noida (UP) India

DOI: http://dx.doi.org/10.24327/ijrsr.2019.1007.3717

ARTICLE INFO

Article History:
Received 06th April, 2019
Received in revised form 14th May, 2019
Accepted 23rd June, 2019
Published online 28th July, 2019

Key Words:

Statistical significance; *p* value; clinical significance; effect-size; clinical research

ABSTRACT

The statistical approaches to medical inquiry involve *p* value to interpret and report the research results. The research result is supposed to impact clinical decision-making in clinical practice. This research outcome must be of adequate magnitude to have clinical importance. The statistical significance as evaluated by the *p* value does not reveal information about the clinical importance of the research outcome. This puts challenges in proper interpretation of research results in order to incorporate the research outcome into clinical practices. In evidence based practice, the decision to make research a reality needs to go beyond statistical significance. To overcome the limitations with *p* value, it has been suggested to include effect-size metrics so that the research outcome can be viewed more objectively along with research results obtained in the past.

INTRODUCTION

The *p* value remains the cornerstone of the quantitative clinical studies. It takes the binary approach of statistical significance to divide the entire data into statistically "significant" and "insignificant" outcomes based on the arbitrary cut off i.e. level of significance generally set at 0.05. This statistical significance determined by *p* value is quite limited in clinical research because a statistically significant research outcome may be clinically insignificant whereas a statistically insignificant research outcome may be clinically significant. A clinically significant research outcome must be incorporated in clinical decision-making as dictated by the evidence-based practice[1]. Statistical significance alone may not reveal all information about the clinical significance[2,3,4,5]. Therefore, it is not enough only to understand what the *p* value is but it is also important to understand what the *p* value is not[6]. The statistical significance indicates whether the effect exists whereas clinical significance makes the researcher understand the size of the effect letting him determine the clinical meaningfulness[7,8,9]. In other words, the clinician is able to decide whether the treatment is effective enough in order to apply it into clinical practice based on the magnitude of the effect.

Determining statistical significance

The *p* value gathers evidence to reject the null hypothesis. The level of significance or α value (which is acceptable limit of Type I or α error when the null hypothesis is true); is generally fixed at 0.05. The $p \leq 0.05$ provides adequate evidence to reject the null hypothesis. Type II or β error is failure to reject null hypothesis when it is actually false. In other words, it is failure to detect the difference between groups when one exists. Therefore the probability of detecting the difference, when one exists, by rejecting null hypothesis (or by not making a β error) is the statistical power denoted as (1-β). The statistical power primarily depends upon the effect-size and the sample size. With bigger effect-size, we need smaller sample size to achieve the desirable statistical power. The power analysis calculates the minimum sample size to ensure the adequate statistical power to the study. Statistical significance is set at 0.05. Statistical power of 80% (0.8) is generally accepted. The sample size can be calculated based on the effect size obtained from the previous studies or the pilot study on a few samples of the main study. A smaller sample size may demonstrate statistically non-significant results which may be clinically important ones. On the other hand, the larger sample size may show statistically significant differences which may be of clinically little significance. An increase in sample size may make the 'insignificant' difference into the 'significant' difference due to an increase in statistical power[10].

Corresponding author:* **Dinesh Kumar Bagga
Department of Orthodontics School of Dental Sciences Sharda University Knowledge Park III Greater Noida (UP) India

This dependence of statistical significance testing on the sample size and abuse of *p* value due to the arbitrarily set level of significance are the main limitations of hypothesis testing. To add, small samples and heterogeneity in clinical studies put further limitations. In view of these limitations with the use of statistical significance testing alone, it is suggested to incorporate other clinically relevant measures such as effect-size while reporting the research results[11].

Determining clinical significance

Robert Abelson mentioned the "MAGIC criteria" (an acronym for magnitude, articulation, generality, interestingness & credibility)[12]. This magnitude of the effect also known as effect-size is a very important measure of clinical significance. There is a quest to find out beyond 'whether the difference exists' to 'how big this difference could be'.

The *p* value is limited only to find out the probability of the difference between the groups. Lower *p* value merely indicates that lower is the probability of rejecting null hypothesis when it is true but does not indicate the bigger effect-size as taken by many researchers due to their misconceptions about *p* value[13,14].

Effect-size refers to the magnitude of the observed effect/ difference or strength of the relationship between the variables in contrast to the statistical significance judged by the *p* value merely reflecting the likelihood of the difference existed by chance only.

The effect-size is a valuable tool in reporting and interpreting the clinical significance. The inclusion of effect-size in addition to the *p* value broadens the concept of statistical approach to the scientific problem solving methods. The reporting of effect-size in original research is still not a common practice. This remains limited to the meta-analysis which makes use of pooling the sample data from different studies and putting them to the statistical tests for drawing statistical inferences. Although the origin of such useful measure can be traced back to 1960[15], its significance came into recognition during 1990[7]. Still reporting of effect-size measures is not a routine practice in the research studies and the interpretation remains limited to the *p* value.

Effect-size is a standardized unit free measure which helps a researcher to understand the magnitude of the research outcome in his study in the backdrop of the previously conducted studies. Furthermore, it lets the researcher to make a decision regarding the incorporation of the research outcome into clinical practice. Here, it is important to mention that *p* value and effect-size are complementary to each other. Hence, both should be reported to get to a larger picture of the research outcome and reach to a more meaningful conclusion.

CONCLUSION

The *p* value represents the statistical significance in terms of the probability of rejecting null hypothesis when it is true. This does not refer to the magnitude of the observed difference between the groups studied in the experiment. The effect-size refers to the magnitude of the difference.

This helps in interpreting the research outcome along with the studies done in past and in making a decision for incorporation of the clinical research into the clinical practice based on clinical significance rather than just statistical significance.

The statistical significance and the effect-size are complementary to each other. Therefore, both must be reported and utilized for interpretation of the research study.

The use of effect-size is limited to the meta-analysis but it is highly recommended that this needs to be reported in the original studies, too.

References

1. Kirk RE. Practical significance: a concept whose time has come. Educ Psychol Meas 1996; 56:746-59.
2. Fan X. Statistical significance and effect size in education research: two sides of a coin. *J Educ Res* 2001; 94:275-82.
3. Vaske JJ. Communicating judgments about practical significance: Effect size, confidence intervals and odds ratios. Human Dimens Wildl 2002; 7:287-300.
4. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007; 82(4):591-605.
5. Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pract 2009;40:532-8.
6. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. *J Grad Med Educ* 2012; 4:279-82.
7. Thompson B. AERA editorial policies regarding statistical significance testing: three suggested reforms. Educ Res 1996; 25:26-30.
8. Hojat M, Xu G. A visitor's guide to effect sizes-statistical significance versus practical (clinical) importance of research findings. Adv Health Sci Educ 2004; 9:241-49.
9. Pandis N. The effect size. *Am J Orthod Dentofacial Orthop* 2012; 142:739-40.
10. Maher JM, Markey JC, Ebert-May D. The other half of the story: Effect size analysis in quantitative research. CBE-Life Sciences Education 2013;12: 345-51.
11. Page P. Beyond statistical significance: Clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther* 2014; 9(5):726-36.
12. Abelson RP. Statistics as principled argument. Mahwah, NJ: Lawrence Erlbaum Associates, 1995
13. Thompson B. The use of statistical significance tests in research: bootstrap and other alternatives. *J Exp Educ* 1993;61:361-77.
14. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. Psychol Methods 2000;5:241-301.
15. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum, 1988.

****