ORIGINAL RESEARCH PAPER
Volume - 10 | Issue - 01 | January - 2021 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

WHEN AND WHERE P-VALUE IS REQUIRED: MISNOMERS IN CLINICAL AND STATISTICAL SIGNIFICANCE

Community Medicine

Dr. Sukhvinder Singh Oberoi* B.D.S., M.D.S.Associate professor,Public Health Dentistry,ESI dental college, Rohini, Indraprastha University.*Corresponding Author

Dr. Mansi Atri B.D.S., M.D.S.Associate professor,Public Health Dentistry,ESI dental college, Rohini, Indraprastha University.

ABSTRACT

The interpretation of the p-value has been an arena for discussion making it difficult for many researchers. The p-value was introduced in 1900 by Pearson. Though, it is very difficult to comment about the demerits of the *p*-values and significance testing which has not been spoken in a long time because of the practical application of it as a measure of interpretation in clinical research. The usage of the confidence intervals around the sample statistics and effect size should be given more importance than relying solely upon the statistical significance. The researchers, should be consulting a statistician in the initial stages of the planning of the study for avoidance of the misinterpretation of the P-value especially if they are using statistical software for their data analysis.

KEYWORDS

Clinical research, Effect size, P-value, Statistical software

INTRODUCTION

The interpretation of the p-value has been an arena for discussion making it difficult for many researchers in the light of the clinical effects giving a varied picture than the statistical interpretation.[1]

Pearson introduced the P-value in 1900,[2] being the one of the method routinely for summarizing the results in medicine and dentistry. As P-value is derived from the tests of significance, it is interpreted to by many researchers to be solely relevant zest for the statistical interpretation.

Though, it is very difficult to comment about the demerits of the *p*-values and significance testing which has not been spoken in a long time because of the practical application of it as a measure of interpretation in clinical research.[3,4]

Many a times, p-value is the sole basis for establishing an association will be there or not. The question is whether it translates the scientific or practical significance in a similar manner. There are a few issues of concern when looking at statistical significance. These issues include choosing the alpha, statistical analysis method, and clinical significance. There is a key distinction in statistical significance versus clinical significance. Statistical significance determines if there is mathematical significance to the analysis of the results. Clinical significance means the difference is important to the patient and the clinician.[5]

Our intention in this article is to evaluate the relevance of p-value and how it could be used in a cautious way with modifications of the conventional threshold of 0.05. We also discuss how to interpret statistical significance keeping in mind its relevance to the statically significance.

Understanding the concept of p-value

The p-value indicates the probability, that for a statistical model, the null hypothesis stated is true or not, the summary of the statistics applied was found to be equal to or more extreme than the actual observed results.[6] That's why, the *P* values give an indication of the incompatibility the data with respect to a particular statistical model used (usually along with a null-hypothesis). *P*-value being lesser, indicating that more is the incompatibility of the statistical data with respect to the null hypothesis.[1]

Though, the medical journals commonly use the terms like: "statistically significant", "unlikely due to chance", "not significant," "due to chance", or common notations like, "P > 0.05", "P < 0.05", indicating the results of a test being having the statistical significance or not which depends upon the P value is the subject of argument between the statisticians. Since, its origin nearly 60 years back, it has been a controversial issue.[7-9]

The P stands for probability and measurement of any observed difference that occurs between the groups is by chance. Probability or p can have value between 0 and 1 with values nearer to 0 giving an indication that this difference is less likely to be due to chance, whereas value nearer to 1 suggesting no difference occurring between the groups for reasons other than chance. So, commonly used terms such as the "highly significant" or "very significant" appearing in the medical literature after citing p-value on the basis of the closeness of eth value to zero.[10]

Fallacies of P-Value

However, p-value might be subject to some fallacies due to some reasons such as that p-value when lesser than 0.05 is said to be surely indicative of the condition being definitely present.[11] The inherent subjective nature of the Fisher's P-value approach along with poor interpretation for it by the medical researchers is the reason that P value gets into a lot of false interpretations. And then it has been considered to be "passports to publication" has led to the scenario being more mis-directed.[12]

The historical evolution of the p-value has made the medical research dependent on the p-value with its limitations. These limitations can be overcome by combining it with the conceptual framework of the confidence interval or the Bayes factor.[10]

Alternatives to p-value

The other way of approach for statistical significance is the "Bayesian theory", to frequentism. Whereas, the frequentists have the method of calculating the probability of the data in light of the null hypothesis, Bayesians have their interpretation on the basis of the probability of the hypothesis with the available data. Though, the theory of frequentism is much more commonly used than Bayesianism, Bayesian inference is much more instinctive: it gives a probability to a hypothesis on the basis of the chances it carries to be true.[13]

The Bayes factors has quite of designs that can be operated on the web-based calculators (e.g. http://pcl.missouri.edu/bayesfactor) and have availability as a new package for R known as the Bayes Factor. One of the issues with the Bayesian approach requirement for specifying the strength of the effect under study before the beginning of the experiment (prior proposition of an alternative hypothesis). Thus, making the choice subjective for 'prior' influences affecting the results of the data analysed.[13]

Schonbrodt et al.[14] says that with over-exaggeration of the Bayesian statistics because the influence of the prior is limited when a reasonable prior distribution is used. The influence of the prior assessment can be on the basis of a "simple sensitivity analysis" in which, the analysis is done on a fixed range of realistic probability decided beforehand only.[15]

Though, the Bayesian analyses have much more involvement in comparison to the null hypothesis of testing the significance, and

having the specification of the prior has addition of some subjective assessment. The p-value can be converted to the Bayes factor, by application of a simple formula with no other information being required. This simplification of the Bayes factor, is also called to be upper bound, which makes the statement that alternative hypothesis is true rather than the null hypothesis beyond any reasonable distribution priorly. (Benjamin and Berger's comment being annexed to [6] and Goodman [16]).

The P values do not tell how 2 groups are different. The degree of difference is referred as 'effect size'. Statistical significance is not equal to scientific significance. P-values being lesser is not an indication that a important effect exists, whereas the higher P-values also cannot rule out any relevant findings. Even with the same effect size, the P values are totally different, based on the sample size. When the sample size is not large enough to find any difference between the groups (a situation of weak statistical power), the P value becomes larger, which makes researchers unable to find any differences between the groups. Any effect, even if it is very tiny, can produce a small P value ($P < 0.05$) with a bigger size of eth sample, and bigger effects can lead to the non-impressive P values ($P > 0.05$) in case of a smaller sample size. To prevent this phenomenon, it is essential to clarify the process of adequate sample size calculation.[1]

The statistical approach has to be modified in light of the previously recognized facts and the newly realized issues pertaining to give less emphasis upon the p-value and more emphasis on the clinical interpretation. Also, whenever possible, the effort should be made to combine the bayes theory to the already obtained results. The science of p-value remains but requirement of the careful interpretation.

CONCLUSION

The benchmark of the p-value lesser than 0.05 is not sufficient enough to decide the clinical significance though statistical significance is achieved. The scientific validity of this though doubtful is still usable for the practical purposes if combined with the clinical validation of the results obtained. The p-value despite so many controversies is still most commonly used as it is much easier to apply with quick results and interpretation.

REFERENCES

1. Nahm FS. What the P values really tell us. Korean J Pain. 2017;30(4): 241-2.
2. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag. 1900;50:157-75.
3. Ziliak, S. (2019), "How Large Are Your G-Values? Try Gosset's Guinnessometrics When a Little 'P' is Not Enough," *The American Statistician*, 73.
4. Hubbard, R. (2016), *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science*, Thousand Oaks, CA: Sage.
5. Tenny S, Abdelgawad I. Statistical Significance. [Updated 2020 Jul 10]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459346/.
6. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat. 2016;70:129-33.
7. Goodman SN. P value hypothesis and likelihood: implications for epidemiology of a neglected historical debate. Amer Journ Epidemiol. 1993;137:485–96.
8. Lehmann El. The Fisher, Neyman-Pearson theories of testing hypothesis: one theory or two? Journ Amer Stat Assoc. 1993;88:1242–9.
9. Goodman SN. Toward evidence-based medical statistics: the P-value fallacy. Ann intern Med. 1999;130:995–1004.
10. Dahiru T. P - value, a true test of statistical significance? A cautionary note. Ann Ib Postgrad Med. 2008;6(1):21-6.
11. Jekel JF. Should we stop using the P value in descriptive studies? Paediatrics. 1977;60:124–25.
12. Mainland D. Statistical ritual in clinical journals: is there a cure?-I. Br Med Journ. 1984;288:841–43.
13. Halsey LG. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? Biol Lett. 2019;15:20190174.
14. Schonbrodt FD, Wagenmakers E-J, Zehetleitner M, Perugini M. Sequential hypothesis testing with Bayes factors: efficiently testing mean differences. Psychol Methods. 2017;22:322.
15. Spiegelhalter D, Rice K. Bayesian statistics. Scholarpedia. 2009;4:5230.
16. Goodman SN. Of P-values and Bayes: a modest proposal. Epidemiology. 2001;12:295-7.