Synchronous use of two markers of research: p-value and confidence interval

Krishna Subedi*[1], Nuwadatta Subedi[2]

[1]Department of Community Dentistry, [2]Department of Forensic Medicine, Gandaki Medical College Teaching Hospital, Kaski, Pokhara, Nepal

Journal of GANDAKI MEDICAL COLLEGE-NEPAL (JGMC-N)

JGMC-N | *Volume* **15**| *Issue* **02**| *July-December* **2022**

***Correspondence:**
Dr. Krishna Subedi
Department of Community Dentistry
Gandaki Medical College Teaching Hospital, Pokhara, Nepal
Email: drkrishnasubedimdsphd@gmail.com
ORCID iD: https://orcid.org/0000-0001-5409-1751

Submitted: December 22, 2022
Accepted: December 25, 2022

To cite: Subedi K, Subedi N. Synchronous use of two markers of research: p-value and confidence interval. JGMC Nepal. 2022;15(2):100-2.
DOI: 10.3126/jgmcn.v15i2.48317

In medical research, p-values are widely used as markers of statistical significance.[1] Researcher frequently compete with the p-value. There is plenty of argument regarding the misuse and misinterpretation of a p-value from the past to the present. However, it is still the most common value that is handled by researchers.

The p-value was first introduced by Fisher in 1922 where it is defined as the probability p under that hypothesis of as great or greater departure than that observed. It was later modified by Neyman et al. in 1933 to a procedure whereby an observed departure with p <5% (or other small value) would lead to rejection of the null hypothesis.[2]

The p-values are used in significance testing for the null hypothesis, which can be accepted or rejected. The null hypothesis states no significant relationship exists between the two variables. Rejecting the null hypothesis means accepting the alternative hypothesis (a significant relationship exists between two variables). The p-value estimates the probability that findings will be at least as extreme as the ones observed when the null hypothesis holds true. Then, it is measured against a predetermined significance level (α). The result is regarded as statistically significant if the stated p-value is less than α. Typically, in biomedical research α is set at 0.05. Sometimes significance levels 0.01 and 0.001 are also used.[3]

There are various misconceptions regarding the interpretation of the p-value, some of which are listed here.

1. Clinical versus the statistical significance of the effect size

The differences between the groups are relevant when the p-value is very low. However, the actual difference may be too small to be clinically important as the p-value carries no information about the magnitude of an effect, which is captured by the effect estimate and confidence interval.

2. Non-significant significance (p >0.05) values

The new therapy has no impact if the p-value is higher than 5%. No matter how significant an impact may be, it is always the observed effect that is best supported by the data from a given experiment.

3. Overinterpreting a non-significant p-value that is close to 5%:

A p-value of 0.06, for example, should not be interpreted as a trend toward a difference. A p-value of 0.06 indicates that there is a 6% probability of getting the result by chance when the treatment has no discernable effect. The null hypothesis shouldn't be rejected because we put the significance level at 5%.

4. Effect sizes versus p –values

The p-value, according to many researchers, is the most crucial figure to report.

However, we should emphasize the effect size. The mean values for each group, the difference, and the 95% confidence interval should be reported instead of just the p-value.[4,5]

5. Highly significant when the p-value is <0.001

When the p-value is very small many researchers believe that the effect is highly significant but it does not hold always. The p-value does not describe the effect size. So it is better to avoid terms like highly or moderately significant based on the p-value.[6] The isolated use of p-value is a matter of debate and raised issues like replicability, veracity, and reliability of the generated conclusions.[7] Therefore, there are some alternatives/complementary to the use of p-value. Complementing p-value with other statistics such as confidence interval, effect sizes, and Bayes factors is recommended.[8]

Confidence interval (CI)

The confidence interval was first introduced by Neyman in 1930.[9] The use of CI in the scientific journal started in the 1980s.[10] However, still many journals are not using CI but only reporting p-value. The relative use of CI is low in clinical and biomedical research.[7]

A CI is a degree of uncertainty around the effect estimate. It consists of an upper and lower limit, which shows that the true (unknown) effect may fall within this interval. The effect stated in the research paper must always lie within the CI reported, and the width of the interval portrays the precision of the effect estimate. For precise effect estimates, CI needs to be narrower. Sample size (n) and heterogeneity (standard deviation [SD] or standard error [SE]) both affect the CI width. There is an inverse relationship between the degree of uncertainty and the sample size, the bigger the sample size, the smaller the width of CI indicating a lower degree of uncertainty. However, the degree of uncertainty is directly proportional to heterogeneity. Studies with lower SDs or SE have narrower CI.[11]

There is a rule for same-sized samples: the smaller the confidence level is, the higher the estimate accuracy.[12] Only the studies with a large sample will give a very narrow confidence interval, which points to high estimate accuracy with a high confidence level.[12]

Finally, the size of the confidence interval is influenced by the selected level of confidence (probability). A 99% CI is wider than a 95% CI. The confidence (probability) level (i.e., 95%) of the CI represents the accuracy of the effect estimate. For example, the 99% CI is more accurate than the 95% CI, because it captures a broader spectrum of the data distribution. In general, with a higher probability to cover the true value the confidence interval becomes wider.[11,13,14] However, the trade-off is that the 99% CI is less precise than the 95% CI. The decision of using a certain confidence level should consider a balance between accuracy and precision. In health sciences, the 95% confidence level is most often used which relates to the level of statistical significance $p< 0.05$. Very less often 90% and 99.9% are also used.[11]

The confidence interval provides a reasonable approximation of the (in) precision and sample size of a given study. As a result, CI can be used to assess the quality of research. Nowadays many journals are recommending the reporting of CI.[12] For continuous data, a CI is the range of values statistically consistent with the value observed in the study. For binary data (presence or absence of disease, prevalence), the CI is determined for the rate of the event. It shows the range of values for the rate of the events that are statistically consistent with the observed rate.[15]

Limitations of CI

Although CIs can be used to enhance the interpretation of a study, they have several limitations. For example, a 95% CI does not have a 95% probability of containing the true value of interest (eg, the true treatment effect), even though it is commonly described that way. Creating an interval that does have a specified probability of containing the true value is termed a probability interval, which further requires a bayesian analysis. In addition, the values within a 95% CI are not the only values that could lead to the current data and model results; they are simply the most compatible values.[16]

The p-value and the confidence interval are two complementary statistical measures. Therefore, it is better to use both p-value and confidence intervals in research. While reporting p-value, do not report a "$p< 0.05$" or "$p \geq 0.05$". It should be reported the actual values. We also encourage the researchers to keep in this mind while submitting their research papers to journals.

REFERENCES

1. Cristea IA, Ioannidis JPA. P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. PLOS ONE. 2018;13(5):e0197440. DOI: 10.1371/journal.pone.0197440 PMID: 29763472.
2. Fraser DA. The p-value function and statistical inference. Am Stat. 2019;73(sup1):135-47. DOI: 10.1080/00031305.2018.1556735
3. Vidgen B, Yasseri T. P-Values: Misunderstood and Misused. Front Phys. 4:6. DOI: 10.3389/fphy.2016.00006

4. Ferreira JC, Patino CM. What does the p value really mean. Jornal Brasileiro de Pneumologia. 2015;41(5)485. DOI: 10.1590/S1806-37132015000000215 PMID: 26578145.
5. Goodman S. A dirty dozen: Twelve p-value misconceptions. Semin Hematol. 2008;45(3),135-40. DOI:10.1053/j.seminhematol.2008.04.003 PMID: 18582619.
6. Madjarova SJ, Williams RJ, Nwachukwu BU, Martin RK, Karlsson J, Ollivier M, et al. Picking apart p values: common problems and points of confusion. Knee Surg Sports Traumatol Arthrosc. 2022;30(10):3245-48. DOI: 10.1007/s00167-022-07083-3 PMID: 35920843.
7. Amaral E de OS, Line SRP. Current use of effect size or confidence interval analyses in clinical and biomedical research. Scientometrics. 2021;126(11):9133-45. DOI: 10.1007/s11192-021-04150-3 PMID: 34565930.
8. Lakens D. The Practical Alternative to the p Value Is the Correctly Used p Value. Perspect Psychol Sci J Assoc Psychol Sci. 2021;16(3):639-48. DOI: 10.1177/1745691620958012 PMID: 33560174.
9. Neyman J. Fiducial argument and the theory of confidence intervals. Biometrika. 1941;32(2):128-50. DOI: 10.1093/biomet/32.2.128
10. Gardner MJ, Altman DG. Estimating with confidence. British Med J. 1988;296(6631):1210. DOI: 10.1136/bmj.296.6631.1210 PMID: 3133015.
11. Hespanhol L, Vallio CS, Costa LM, Saragiotto BT. Understanding and interpreting confidence and credible intervals around effect estimates. Braz J Phys Ther. 2019;23(4):290-301. DOI: 10.1016/j.bjpt.2018.12.006 PMID: 30638956.
12. Simundic AM. Confidence interval. Biochem Medica. 2008;18(2):154-61. DOI: 10.11613/BM.2008.015
13. Du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2009;106(19):335. DOI: 10.3238/arztebl.2009.0335 PMID: 19547734.
14. Doll H, Carney S. Statistical approaches to uncertainty: p values and confidence intervals unpacked. BMJ Evid-Based Med. 2005;10(5):133-4. DOI: 10.1136/ebm.10.5.133
15. Young KD, Lewis RJ. What Is Confidence? Part 2: Detailed Definition and Determination of Confidence Intervals. Ann Emerg Med. 1997;30(3):311-8. DOI: 10.1016/S0196-0644(97)70167-7 PMID: 9287893.
16. Hawkins AT, Samuels LR. Use of Confidence Intervals in Interpreting Nonstatistically Significant Results. JAMA. 2021;326(20):2068. DOI: 10.1001/jama.2021.16172 PMID: 34812882.