Interpreting Research Findings with Confidence Interval

**Neeraj Hirpara,
Sandesh Jain,
Alpana Gupta,
Soumya Dubey intern**

Govt College of dentistry Indore, M.P India

Abstract

An understanding of p-values and confidence intervals is necessary for the evaluation of results of any study. The purpose of this article is to provide useful information for the interpretation of these two statistical concepts. The results based solely on P value can be misleading as it only tells that whether the results are significant or not but the confidence interval provides a range of probabilities of effect sizes within which the true value would lie 95% or 90% of the time, depending on the precision desired. CI provide a slightly different estimate of statistical truth, it provide a range of the true effect observed in a study.

Key words: Confidence interval; p value; Precision; Power; Sample size

Corresponding author:
Neeraj Hirpara

✉ drneerajhirpara@yahoo.in

Govt College of dentistry orthodontics, Indore, M.P, INDIA

Tel: 09425045455

Introduction

Confidence interval is the range of values, a variable or outcome measure calculated from data within which true value of parameter lies with some specific probability. Studies data can be assessed by calculating probability (p-value) and also by calculating confidence interval. The specified probability is called the confidence level, and the end points of the confidence interval are called the confidence limits' [1]

P-value and Confidence Interval are common statistics measures which provide complementary information about the statistical probability and conclusions regarding the clinical significance of study findings [2]. In situations involving normally distributed data or large sample of data from other distribution, the normal approximation may be used to calculate confidence interval but in clinical trials where number of observed adverse events is small the criteria for approximate normality is calculated [3].

The decision of hypothesis testing is in dichotomy of significant or non significant. There is a cut-off point and on that basis the result is classified as significant or non significant. In contrast, confidence interval provides a range of observed effect size which is likely to represent true effect size. It is conventional to create a confidence interval at 95% level which means that 95 out of 100 times it would contain a true value of the variable [4]. Statistically significant result can be inferred from confidence interval along with largest upper bound & smallest lower bound effect size thus providing additional information [4].

P value is calculated to assess whether the difference or the result is by chance or not. P value simply provides a cut-off beyond which we assess that the findings are statistically not significant (p<0.5). However, the results of a statistical testing are highly influenced by the sample size and sample variability (within sample) or standard deviation (**Table 1**) [5,6].

Smaller sample size leads to higher p value (non significance difference) known as – Type II error (false negative results). Keeping everything else constant if sample size is larger, then statistically significant result can be obtained. Similarly when the mean outcome value and sample size are constant, there is an increase in the variability (large standard deviation) which leads to a higher p value (Type II or β error is missing the real difference when it is there.) [5,6] For similar effect size in two different studies which has large sample size and little variability of sample hypothesis testing leads to statistically significant result compared to similar study with small sample size and greater variability.

P value indicates nothing about effect size, it only indicates that the difference is not by chance. It is the probability of committing Type I error (false positive results) Statistical significance alone cannot convey the complete picture of effectiveness of

Table 1 Relationship of p value with sample size and standard deviation.

P value	Sample Size	Standard Deviation
Significant	More	Less
Non-Significant	Less	More

an intervention. Statistically significant result may not have meaningful impact in clinical settings while the use of confidence interval helps researchers to interpret both statistical as well as clinical significance. Confidence interval provides information about magnitude of effect and uncertainty surrounding it. Thus clinical significance can be calculated using confidence interval and minimal clinical important difference (MCID). We can know the clinical significance of study [5]. A study is clinically significant if its 95% confidence interval is higher than minimal clinically important difference (MCID). Value of statistical significance alone cannot convey the effectiveness of intervention both statistical significance and clinical significance both are important for interpreting clinical research.

Confidence interval is much more helpful than simple p value. The upper & lower bound of confidence interval are important. Just because we have not found a significant treatment effect (p>.05), it does not mean that there is no treatment effect to be found. We have to carefully interpret the findings.

Importance of calculating CI

Suppose we have to calculate average height of male population. Since we cannot measure each individual, we select a sample. Sample mean comes out to be X feet. Because this is derived from sample we are uncertain on the accuracy of this measure. We now calculate 95% confidence interval. This will indicate the range of values that we expect to represent the true population mean. The narrower the confidence interval the less is uncertainty. Thus helps us to conclude with some level of confidence.

For e.g. if we have to find a difference in molar distalization between two groups. If the 95% confidence interval range of difference is 0.2-3mm. It means if we repeat this study 100 times, then 95% this value will fall between 0.2-3 mm. We can interpret this as high degree of uncertainty, because value is representing a wide range from 0.2 mm which is of no clinical value to 3 mm which is a clinically significant amount.

Confidence interval aids in interpreting the study by giving upper and lower bounds of effects. E.g. - 95 confidence interval of risk ratio is 0.78 (0.70-0.86). Its intervention is as follows – since the confidence interval does not embrace risk ratio one (0.70-0.86) this observed risk is statistically significant at 5% level. Secondly we can observe from this range that there is as much as 30% reduction in risk or little as 14% reduction of risk.

CI is valuable if non significant differences are found by p values, hence a clinical judgement can be made even in non significant results. For e.g. if the risk ratio is 1, indicating no difference in experimental and control group (p>0.05). If its 95% confidence interval is 0.85-1.26 indicating an effect range from 15% protective to 20% harmful. In this case although the p value indicate no difference between the two groups but the confidence interval tells us different clinical story. So we not only know that the result is significant or not significant but we also know how large or small a true difference might be.

Interpretation of CI and p values

95 % CI means that there is 95 % chance that range contains the true population mean or in other words we can interpret that if we repeat the experiment/study 100 times from many samples then we get the same values in 95 % of cases. Rest 5 % may not include true population mean value. Similarly if CI is 99 % then our confidence level will be 99 %

We can assess significance of result from confidence interval in the following ways (**Table 2**).If confidence interval captures the value of no effect (i.e 1 in case of risk ratio or odds ratio and 0 in case of rate difference or distance in mm) this represents a statistically non significant result. If confidence interval does not include the value of no effect, then this represents statistically significant result.

Relationship of CI with precision and range

Precision is a measure of consistency and is a function of random error and confidence required. At 99% confidence interval excluding the outline results become nearer to actual measurement. Similarly with 95% confidence interval the results are more precise (**Table 3**)

Width of CI

Width of confidence interval is associated with sample size (**Table 4**). Narrow width of CI means there is small range of effect size in the study indicates study size is quite large since the range of effect size is narrow & hence the study has reasonable certainty. Wide

Table 2 Interpreting CI and p values for measures of effect [6].

Measure of effect	No effect value	Confidence interval	P value
Risk or rate difference or distance in mm	0	95% CI includes 0 Non significant	p>0.05 Non significant
Risk or rate difference or distance in mm	0	95% CI does not include 0 Significant	P<0.05 Significant
Relative risk, relative ratio or odds ratio	1	95% CI includes 1 Non significant	p>0.05 Non significant
Relative risk, relative ratio or odds ratio	1	95% CI does not include 0 Significant	P<0.05 Significant

Table 3 Relationship of CI with precision and range.

CI	Precision	Range of effect size	Confidence level
95 CI (±1.96 SD)	More	Narrow	95 %
99 CI (±2.58 SD)	Less	Large	99 %

Table 4 Width of confidence interval and its relation to precision and power.

Width of CI		Precision	Power
Narrow	Larger studies with sufficient sample size	Precise	Sufficient power
Wide	Smaller studies with sufficient sample size	Relatively less precise	Low power

or diverse range of effect size & hence the estimate is not precise. Study has no reasonable power to detect an effect.

Confidence interval like p value guides us to help interpreting the research findings in the light effect of chance.

The findings of any study are related to patient included in the study. The findings may not be applicable to other groups of patients (external validity). Assessment of this external validity should be made. Neither confidence interval nor p value is of much help for this judgement.

Pitfalls in interpretation

1. Statistically significant does not necessarily mean that the effect is real/true. One in 20 significant findings will be spurious (type I error). However, on the other hand just because we are unlikely to observe such a large difference simply by chance, this does not mean that it will not happen. One in 20 may be by chance and will mislead us.

2. Statistically significant does not necessarily mean clinically important. It is the size of effect that determines clinical importance and not the presence of statistical significance. A large study may identify a fairly small difference as statistically significant which may or may not be clinically significance. It is the size of effect and not just the size of significance that is important to make clinical difference.

3. Non significant does not mean no effect. Small studies will often report non significance even when the difference is real & important (type II error). A non significant confidence interval simply tells us that the observed difference is consistent with there being no difference between the two groups.

Before interpreting result we must first check that the study is not biased. There are six key methodological criteria (quality items) which are used for assessment of bias in the study: sample size calculation, random sequence generation, allocation concealment, reporting of withdrawals, blinding of measurement assessment, and the use of intention to treat analysis [10]. Depending on the number of quality items fulfilling the criteria, the study can be classified as low (5 or more), medium (3 or more) and high risk (less than 3).

Conclusion

Interpretation of results based only on P values can be misleading. Confidence interval conveys more information than P values. It provides magnitude of effect as well as its variability. Confidence interval should be calculated for each variable especially if P values are insignificant.

References

1 Last JM. A dictionary of epidemiology. Oxford: International Journal of Epidemiology, 1988

2 Harari G1 (2014) [Confidence interval or p-value--similarities and differences between two important methods of statistical inference of quantitative studies]. Harefuah 153: 48-51, 63.

3 Harari G, Herbst S (2014) [Confidence interval calculation for small numbers of observations or no observations at all]. Harefuah 153: 289-291, 304, 303.

4 Gardner MJ, Altman DG (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J (Clin Res Ed) 292: 746-750.

5 Jain S, Gupta A, Jain D (2015) Estimation of sample size in dental research. Int Dent Med J Adv Res 1: 1-6

6 Jain S, Sharma N, Jain D (2015) Basic fundamentals of designing a quality research. J Adv Med Dent Sci Res 3: 88-95

7 Jain et al. (2015) Relevance of statistical and clinical significance in dental research. Int Dent Med J Adv Res 1: 1-3.

8 Pandis N (2013) Confidence intervals rather than P values. Am J Orthod Dentofacial Orthop 143: 293-294.

9 Theodore Eliades (2013) Research methods in: A Guide to Understanding Orthodontic Research, 179p

10 Fleming PS1, Johal A (2010) Self-ligating brackets in orthodontics. A systematic review. Angle Orthod 80: 575-584.

11 Jain S, Jain S, Punyani PR, Jain D (2015) Basics of interpreting results. Int Dent Med J Adv Res 1: 1-4.

12 Jain S, Gupta A, Jain D (2015) Common Statistical Tests in Dental Research. J Adv Med Dent Scie Res 3: 38-45.