IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 16, Issue 3 Ser. IV (May – June 2020), PP 47-51
www.iosrjournals.org

Confidence Interval as a Better Alternative to P-Value for Clinical Significance

KazeemAdewaleOsuolale and AdesolaZaidat Musa
Monitoring and Evaluation Unit (Biostatistics), Nigerian Institute of Medical Research, Lagos, Nigeria

Abstract: *Quite manyresearchers find understanding p-values and confidence interval (CI) difficult which are essential for the evaluation of results of any study. A test of the hypothesis does not indicate what the difference is or how large it is. Simple statements in a study report such as 'p < 0.05' or 'p =Not Significant(NS)' do not explicitly describe the results of a study. Complementing the hypothesis test with a CI will indicate the magnitude of results and help researchers to decide whether the difference is of interest clinically. This study provides useful information for the interpretation of these statistical concepts to avoid misleading results. The use of these two statistical concepts and the differences between them are discussedbased on a comprehensive and selective literature search on the methods in scientific studies. While thep-values are used to determine whether a null hypothesis is to be accepted or rejected and also enable the recognition of statistically significant findings, the CI provides an estimate of the precision with which a statistic estimates a population value. For instance, in a clinical trial of a placebo versus a hypotensive agent, each group with 10 patients, the change in blood pressure for the placebo was 17 mmHg and that of the hypotensive drug was 30 mmHg. If the pooled standard deviation was 15.5 mmHg by the two-sample t-test: t = 1.9, degree of freedom (df) = 18 and p = 0.06. This fails to reach the conventional 5% significance level and maybe declared not statistically significant. However, the potential benefit from a reduction in blood pressure of 13 mmHg is substantial and so the result should not be ignored. In this case, it would be misleading to state that 'There was no significant difference between the drug A and B, and it would be better to quote the extra gain achieved of 13 mmHg, together with a 95% CI of −2 to 28 mmHg. In this way, we can truly judge if the trial results are indicative of no difference or that, in a larger trial, the clinically important benefit of 17 mmHg indicated may be proven right. This enables conclusions to be made about the statistical significance and clinical importance of the study findings. This study, therefore, concludes that the presentation of both the p-value and CI is desirable since they provide supplementary information. But if only one is to be reported, CI must be given preference over the p-value for the sake of clinical significance.*

Keywords: *Confidence interval; P-value; Clinical importance; Preference, Significance*

Date of Submission: 04-06-2020 Date of Acceptance: 20-06-2020

I. Introduction

P-values and Confidence intervals are widely used among scientific researchers to summarisesignificance, however, theyare often misused and misinterpretedeven though they are essential for the evaluation of results of any scientific study. Some researchers wonder why a p-value is given as a measure of statistical probability in certain studies, while other studies give a confidence interval and still others give both. This article looked at definitions, appropriate use and interpretation of P-values and Confidence intervals with examples of applications. The study aims to provide a useful guidance for the interpretation of these statistical concepts to avoid misleading reports and guard against statistical fallacies.

There are two basic divisions of statistical inference such as**estimation**, which is associated with confidence intervals while **hypothesis testing**is connected with *p*-values (Machin, et al., 2007). Both p-value and CI are common statistics measures, which provide complementary information about the statistical probability and conclusions regarding the clinical significance of study findings (Harari, 2014). The notion of p-values as recently stated by Nguyen (2019) provides some useful information from the observed data but it is not enough to use it alone to make decisions as statistical significance cannot convey the complete picture of the effectiveness of an intervention. Complementing the hypothesis test with a CI will indicate the magnitude of results and enable researchers to decide whether the difference is of interest clinically.

What is P-Value?

The *p*-valueue is the probability of rejecting or failing to reject the null hypothesis, H_0(Boos and Stefanski, 2011).The H_0 is the hypothesis that there is no difference between two groups for a specific variable. The "*p*" in p-value stands for probability. It measures the strength of evidence against H_0(Matthew, 2016).

Ronald Fisher, who introduced the *p*-value, intended it as an informal way to judge whether the evidence was significant in the sense of being worthy of a second look (Nuzzo, 2014). A very small *p*-value indicates that the null hypothesis is very incompatible with the data that have been collected. However, we cannot say with certainty that the null hypothesis is not true, or that the alternative hypothesis must be true (Van Rijn et al., 2017).

What is Confidence Interval?

A confidence interval is defined as 'a range of values for a variable of interest constructed so that this range has a specified probability of including the true value of the variable, as well as about the direction and strength of the demonstrated effect.The specified probability is called the confidence level, and the endpoints of the confidence interval are called the 'confidence limits'.(Last, 1988, Sandeep K. Gupta).

II. Methods

The uses of these two statistical concepts and the differences between them are discussed based on a relevant and selective literature search on the subject. The diagram presented in Figure 1 was adopted from Machin, et al. (2007) to expatiate the difference between statistical significance and clinical importance.

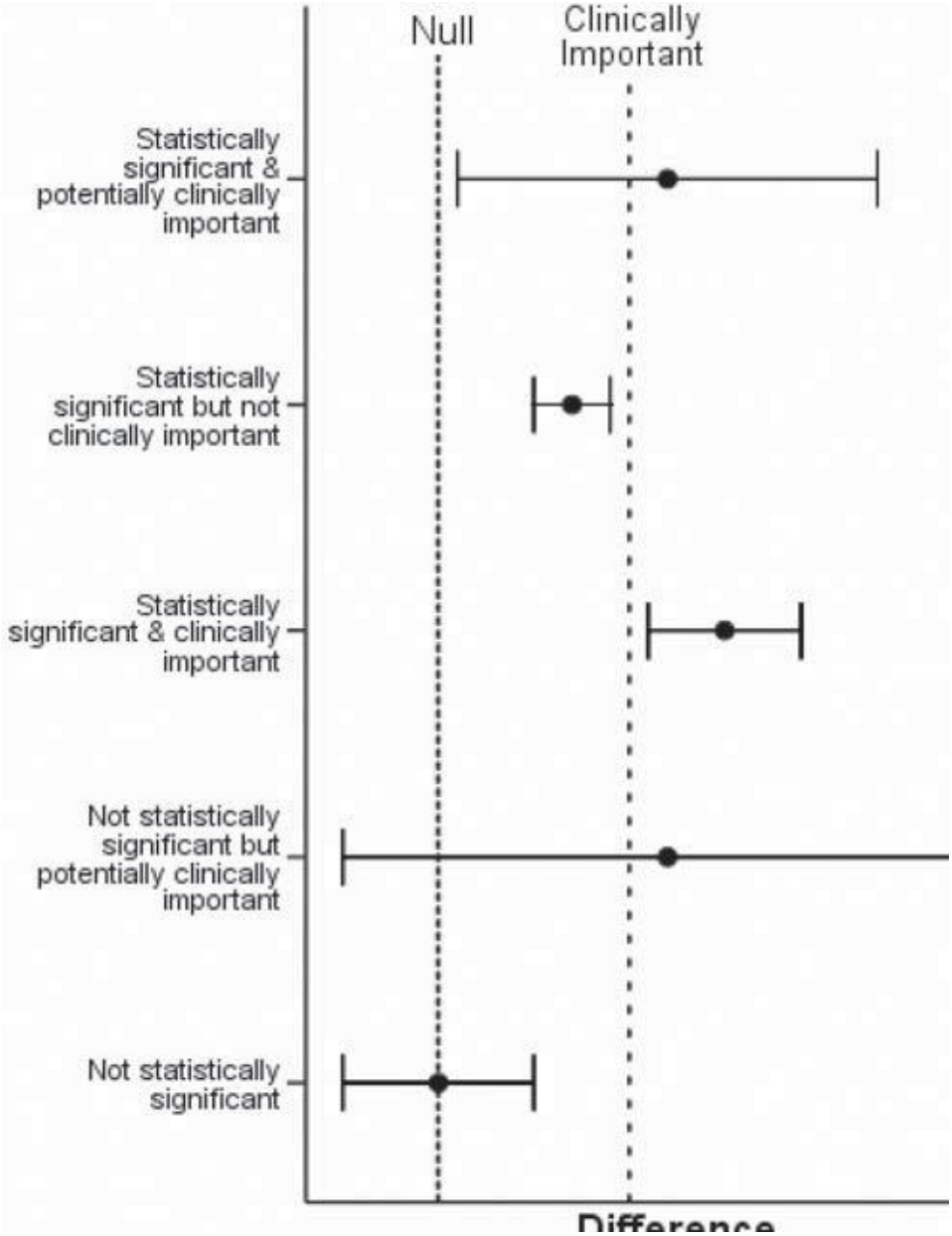

Figure 1: Use of confidence intervals to help differentiate statistical significance from clinical importance

Thus, clinical significance can be calculated using confidence interval and Minimal Clinically Important Difference (MCID). A study is clinically significant if it is 95% CI is greater than MCID.

Applications with Examples

This section delves into the applications of confidence interval (CI) with a follow-up example for a better understanding of its usefulness. Medical researchers use CI as a numerical range to describe research data of a specific study.The CI can be calculated for the mean and proportion with the assumptions that sample mean, $\bar{x}$ follows a Normal distribution for large sample size and the sampling distribution of a proportion follows a Binomial distribution and approximately normal for reasonably large sample size, n. The 95% confidence interval for themean is estimated by:

$$\left(\bar{x} - \left(1.96 \text{ x } \frac{\sigma}{\sqrt{n}}\right), \bar{x} + \left(1.96 \text{ x } \frac{\sigma}{\sqrt{n}}\right)\right)$$

where$\bar{x}$ is the sample mean, σ is the population standard deviation and n is the sample size.**Should** the underlying data are not normallydistributed or the population standard deviation is unknown,the sample mean follows a t-distribution **and** the 95% confidence interval for the meanis calculated as:

$$\left(\bar{x} - \left(t_{0.05} \text{x } \frac{s}{\sqrt{n}}\right), \bar{x} + \left(t_{0.05} \text{x } \frac{s}{\sqrt{n}}\right)\right)$$

where$t_{0.05}$ is the **percentage point** of the t-distribution with (n - 1) degrees of freedom which gives a two-sided probability of 0.05.The 95% confidence interval for the proportion is also estimated by:

$$\left(p - \left[1.96 \text{ x } \sqrt{\frac{p(1-p)}{n}}\right], p + \left[1.96 \text{ x } \sqrt{\frac{p(1-p)}{n}}\right]\right)$$

The p in the above formula is the sample proportion and is estimated by $p = \frac{x}{n}$ (where $\boldsymbol{x}$ is the number of individuals in the sample with the characteristic of interest) and$(1-p)$ is replaced by $(100-p)$if p is expressed as a percentage.

Example I

The weight of 10 babies born in a hospital in Nigeria was measured and recorded in kilogram (kg) as $x_1 = 4.94, x_2 = 4.56, x_3 = 4.00, x_4 = 4.69, x_5 = 3.72, x_6 = 4.13, x_7 = 4.13, x_8 = 3.44, x_9 = 3.97$ *and* $x_{10} = 3.00$.

Table 1: Descriptive analysis of baby weight in a hospital

Measures		Statistic	Std. Error
Mean		**4.06**	0.18
95% Confidence Interval for Mean	Lower Bound	**3.64**	
	Upper Bound	**4.47**	
5% Trimmed Mean		4.07	
Median		4.06	
Variance		0.34	
Std. Deviation		0.58	
Minimum		3.00	
Maximum		4.94	
Range		1.94	
Interquartile Range		0.94	

Going by the results in Table 1,a claim that the true average baby weight of 10 babies born in a hospital in Nigeria is more than 4.0 kg given that the mean calculated from their respective weights is 4.06kg and the 95% CI for the true average weight is (3.64, 4.47). The sample average of 4.06kg is more than 4.0kg and therefore the claim is not statistically significant since CI includes weights lower than 4.0kg. The claim that the true average baby weight is over 3.0kg is a statistically significant conclusion since it holds for all values in the CI.

Example II (extracted from Machin, et al., 2007)

In a clinical trial of two hypotensive agents, with 500 subjects on each treatment, one treatment reduced blood pressure on average by 30 mmHg and the other by 32 mmHg. Suppose the pooled standard deviation was 15.5 mmHg, then the two-sample z-test gives z = 2.04, p = 0.04. This is a statistically significant result which may be quoted in the Abstract of the scientific report as A was significantly better than B without any mention

that it was a mere 2 mmHg better. Such a small difference is unlikely to be of any practical importance to individual patients. Thus, the result is statistically significant but not clinically important (Given a large enough study, even small differences can become statistically significant).

In a clinical trial of a placebo versus a hypotensive agent, each group with 10 patients, the change in bloodpressure for the placebo was 17 mmHg and for the hypotensive drug, it was 30 mmHg. If the pooled standard deviation were 15.5 mmHg, by the two-sample t-test: t = 1.9, df = 18 and p = 0.06. This fails to reach the conventional 5% significance level and may be declared not statistically significant. However, the potential benefit from a reduction in blood pressure of 13 mmHg is substantial and so the result should not be ignored. In this case, it would be misleading to state that 'There was no significant difference between the drug A and B', and it would be better to quote the extra gain achieved of 13 mmHg, together with a 95% CI of −2 to 28 mmHg. In this way, we can truly judge if the trial results are indicative of no difference or that, in a larger trial, the clinically important benefit of 13 mmHg indicated may be proven right.

Scientific Relevance

Similar to other inferential statistics, confidence intervals (CIs) are a powerful tool that medical researchers use to analyze and interpret their data. The CIs allow researchers to make statements about how a small experiment relates to a larger one. Confidence interval also provides possible estimates about the magnitude of the effect in the population from which samples were drawn rather than merely relyingon statistical significance using the only p-value. The upper and lower limits of a CI serves as a means of assessing whether the results are clinically important or not. Researchers can check if a hypothesized value for the population parameter falls within the estimated confidence interval. If a hypothesized value falls within the CI, then the results are consistent with this hypothesized value. Otherwise, it is unlikely that the parameter has this value. The use of CI to complement *p*-value will afford medical researchers the opportunity not to ignore findings that are potentially clinically relevant.

III. Discussion

A *p*-value > 0.05 only implies that the evidence is not sufficient to reject the null hypothesis

That there is no difference between two alternative treatments but there is a need to complement this a confidence interval to indicate the magnitude of results to enable researchers to take appropriate decisions of clinical importance. Reputable international journals of medical sciences, such as the Lancet and the British Medical Journal, as well as the International Committee of Medical Journal Editors (ICMJE), recommend the use of confidence intervals in interpreting the results of randomized clinical studies and meta-analyses (Altman, 2002; Jean-Baptist, et al., 2009). The confidence intervals are considered very important that they are used in three out of four medical papers published in reputable journals (Harris and Taylor, 2003).

For instance, a claim that the true average baby weight of 10 babies born in a hospital in Nigeria was more than 4.0 kg posed in Example I with the mean of 4.06kg and the 95% CI for the mean weight of (3.64, 4.47) implied that the claim was not statistically significant since CI includes weights lower than 4.0kg. A statistical significance, in this case, will hold for a claim that falls within the lower and upper confidence limits without the inclusion of zero in the interval.

In the first part of Example II, the result is statistically significant since the p-value was less than 0.05 but not clinically important due to a small difference of 2 mmHg reduction in the blood pressure which is unlikely to be of any practical importance to individual patients. However, in the second part of Example II, there was no significant difference between the drug A and B since the p-value is greater than 0.05 and the CI includes zero but clinically important due to the clinical benefit of a reduction in blood pressure by 13 mmHg.

IV. Conclusions

The p-value is used to determine whether a null hypothesis is to be accepted or rejected and it enables the recognition of statistically significant findings. The confidence interval provides an estimate of the precision with which a statistic estimates a population value. This enables conclusions to be made about the statistical significance and clinical importance of the study findings. The presentation of both the *p*-value and CI is desirable since they provide supplementary information. However, if only one is to be reported, CI must be given preference over the *p*-value for the sake of clinical significance or relevance.

References

[1]. Machin, D., Campbell, M. J. and Walters, S. J. (2007). Medical statistics:A Textbook for the Health Sciences. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.

[2]. Harari, G.1. (2014). Confidence interval or p-value--similarities and differences between two important methods of statistical inference of quantitative studies. Harefuah, 153: 48-51.

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

[4]. Jain, S., Sharma, N. and Jain, D. (2015). Basic fundamentals of designing a quality research. J. of Adv. Med. Dent. Sci. Res., 3: 88-95.
[5]. Last, J. M. (1988). A dictionary of epidemiology. Oxford: International Journal of Epidemiology.
[6]. Nguyen, H.T. (2019). Editorial: How to test without p-values? Thai. Stat.; 17(2): i-x.
[7]. Sandeep K. G. (2012). The relevance of confidence interval and *P*-value in inferential statistics.Indian J.Pharmacol. 44(1): 143–144. DOI: 10.4103/0253-7613.91895.
[8]. Boos D.D., Stefanski L.A. (2011). P-Value precision and reproducibility. Am Stat, 65:213-21. 10.1198/tas.2011.10129.
[9]. Matthew S.T., Brenden R.and Ulrike O. (2016).P-value interpretations and considerations.J.Thorac Dis.; 8(9): E928–E931.DOI: 10.21037/jtd.2016.08.16.
[10]. Nuzzo R.(2014). Scientific method: statistical errors. Nature;506:150–152.
[11]. van Rijn M.H. C., Bech A., Bouyer J., van den Brand Jan A. J .G.(2017).Statistical significance versus clinicalrelevance. Nephrol Dial Transplant. 2017 Jan 07; DOI: 10.1093/ndt/gfw385.
[12]. Rosenthal, J.S. (2018). Statistics using just one formula. International Journal for Teachers; 40 (1), 7-11.
[13]. Altman D.G. (2002). Confidence intervals in practice. In: Altman D.G., Machin D., Bryant T.N., Gardner M.J. BMJ; 6–9.
[14]. Jean-Baptist du Prel, Gerhard Hommel, Bernd Röhrig, Maria Blettner (2009). Dtsch Arztebl Int., 106(19): 335–9 DOI: 10.3238/arztebl.2009.0335.
[15]. Harris M. and Taylor G. (2003). Medical statistics made easy. Martin Dunitz, Taylor & Francis Group, United Kingdom.

KazeemAdewaleOsuolale, et al. "Confidence Interval as a Better Alternative to P-Value for Clinic Significance." *IOSR Journal of Mathematics (IOSR-JM)*, 16(3), (2020): pp. 47-51.