

HOW CONFIDENT IS THE CONFIDENCE INTERVAL

K.I. Egbuchulem

Division of Paediatric Surgery, Department of Surgery, University College Hospital, Ibadan

The classroom teaching of statistics is the pivot for answering these research questions – How confident is the confidence interval and how confident do you want to be as a researcher that sample estimates collected are as accurate as you wish? These are the focus of this editorial, to highlight some of these facts, myths about test of significance as it relates to confidence intervals and the question how confident the confidence interval is needs to be interrogated.

Most biomedical research testing the level of significance report Confidence Intervals (C.Is), and these are useful in interpreting results of statistical analysis. Literally, it should give the researcher some degree of confidence about the research output in terms of reliability, accuracy, and precision. It is usual for biomedical researchers and other investigators to ask questions such as ‘Is the result statistically significant?’, and this is a source of serious concern among researchers., Some tend to disregard or downplay a finding just because it was not significant while some are worried just because the outcomes were not statistically significant.

The reporting of confidence intervals usually follows hypothesis testing or significance testing. Hypotheses refer to statements concerning the situation being investigated which are usually stated as two mutually exclusive options, a null hypothesis, and an alternative hypothesis. These can be stated as two tail which is usually favored or a one tail hypotheses. The null hypothesis is a statement of no association between variables or no difference in means while the alternative hypothesis states that there’s a difference or an association beyond what is attributable to chance.¹ Each time a null hypothesis is rejected, there is always an alternative hypothesis for possible acceptance. The interests of medical researchers are varied, and research questions result in statement of hypotheses.

Examples of such questions are: In the article on Burden of Erectile dysfunction among chronic heart failure patients in Ibadan: A pilot study, one may want to find out if there is a significant difference in the International index of erectile function between chronic heart failure patients and patients without cardiac failure.

The probability that the observed result is due to chance alone is what is referred to as P – value.²

The P- value only tells us whether the observed clinical difference is statistically significant or not. The confidence interval reflects the precision of the sample values in terms of their standard deviation and the sample size.³

P- value has a positive correlation with C.I, that is both are usually significant or not significant depending on the outcome of the study. Usually, the initial descriptive statistics used to summarize variables such as proportions, frequencies and means gives an idea of the results of our study, but the statistical significance is what the p value helps to ‘endorse’ and the C.I confirm it thereafter with either a 95% or 97% affirmation order as it is commonly used.

The interpretation of p – values (alpha) is based on reference to a particular cut off for the probability or the so-called level of significance which is conventionally set at 0.05 for a 95 % C.I and 0.01 for a 99% C.I. Hence p-values less than this number are significant while those above are not statistically significant. The confidence interval gives the range of values within which we are reasonably confident that the population parameter lies within.³ The parameter here could be difference in means, or proportions of two groups or it could be a measure of association between two variables such as odds ratio. The most reported interval is the 95% confidence interval at alpha value of 0.05. When the study is repeated several times, about 95% of the different possible results obtained will lie in this interval. Alternatively, we can say that we are 95% confident that the true population value of what we are estimating in our study lies within the interval. Confidence intervals makes it far easier to determine whether a finding has any substantive (e.g., clinical) importance, as opposed to statistical significance. While statistically significant tests are vulnerable to type I error, C.I is not. Confidence level is the complement of the Type 1 error (1- α).

The higher the confidence interval e.g., 95% versus 90%, the lesser the chances of error and the more precise and accurate the study is. In same vein C.Is have upper and lower boundaries, the narrower the interval the more precise the analytes are, and a wider boundary or limit is more likely to be error prone.⁴ The width of the confidence interval and the size of the p value are related, the narrower the interval, the

smaller the p value. The criterion for judging an interval as significant or not depends on the presence of a null value. The null value refers to the value of the test statistic when the null hypothesis is true.

In using serum biochemical markers as predictors of enterocolitis in children with colorectal anomalies, the null hypothesis is that there is no difference in the cut off for predicting enterocolitis in these children using either Calprotectin or C-reactive protein.

The null value here is zero and any interval computed for the difference in the mean serum levels of these analytes which includes zero is not significant.

Another set of study design involves investigation of relationships between two variables typically risk factors e.g., the article Relationship between Heart Rate variability and hypotension with bradycardia following spinal anesthesia in patients undergoing elective surgery, the null hypothesis is that there is no difference in the mean heart rate variability and hypotension with bradycardia. The appropriate measure of association between these variables is the odds ratio and the null value; and when there is no relationship between heart rate variability and hypotension with bradycardia, the null value becomes 1. Hence a confidence interval including 1 will not be a significant interval.

A third scenario is if the variables being investigated are both numeric, say the relationship between the serum levels of biomarkers, Calprotectin and C-reactive protein in children with colorectal anomalies post-surgery, where the measure of association here is the correlation coefficient. The null value here is zero and any interval for the correlation coefficient between them including zero will not be significant. As a guide to interpreting confidence intervals for

difference in means, when the lower and upper limits are both positive and both negative, depending on the direction then the difference is significant. Also, for odds ratios when the upper and lower limits are both decimals and both whole numbers then we have a significant result.

It is worthy of note that the confidence interval in addition to p-value are two equivalent methods of interpreting results of a statistical analysis. Whether or not we have a significant result can be determined from

the p value based on whether it is less than 5% or not; or the confidence interval based on whether the null value lies within the interval. However, the confidence interval gives valuable information about the likely magnitude of the effect being investigated and the reliability of the estimate. Larger sample sizes will give narrower and hence more reliable intervals.

Finally, the use of C.I promotes cumulative knowledge development by obligating researchers to think meta-analytically about estimation, replication and comparing intervals where in lies the mean, odds ratio, and relative risk etc. across studies.⁵ It also minimizes the sampling error.

In conclusion:

Everyone should promote and encourage the use of confidence intervals around sample statistics as both the confidence of the researcher and the validity of a research finding is illuminated.

C.I should be the preferred means of interpreting results from biomedical research, because in addition to evaluating the role of chance it reflects the degree of precision and accuracy. This core duty lies in the hands of biostatistics teachers, medical journal editors like ours – *Annals of Ibadan Postgraduate Medicine*, reviewers, and many other granting agencies.

REFERENCES

1. **Bamgboye EA** 2002. *A Companion of Medical Statistics*, 1st edition. Folbam publishers, Ibadan.
2. **Kirkwood BR**, Sterne JAC 2003 *Essential medical statistics*, 2nd edition Blackwell Publishing Ltd, London.
3. **Eugene KH**. On P value and confidence intervals (Why can't we P with more confidence). Editorial. *Chin Chem* 1993; 39(6):927 – 928.
4. **Adedokun BO**. P – value and Confidence intervals – Facts and Farces. *Ann. of Ib. Postgrad. Med.* 2008; 6 (1): 33.
5. **Thomas B**. What future quantitative social science research could look like: confidence interval for effect sizes. *Educ Research* 2002; 31:25-32.