P - VALUE AND CONFIDENCE INTERVALS – FACTS AND FARCES
Adedokun B.O.

Correspondence:
Adedokun B. O.
Lecturer
Department of Epidemiology and Medical Statistics,
College of Medicine,
University of Ibadan.

P values and confidence intervals are reported in almost all scientific writings and are used in interpreting results of statistical analysis. It is usual for medical researchers and other investigators to ask questions such as 'Is the result significant?' or 'what is the p value?' Many clinicians worry when they carry out statistical analysis and there are no significant results. This article describes some facts about the p value and confidence intervals.

The reporting of p values and confidence intervals usually follows hypothesis testing or significance testing. Most scientific investigations involve the testing of hypotheses. These are formal procedures for testing whether findings from the investigations are compatible with a so called null hypothesis. 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. The null hypothesis is a statement of no association between variables or no difference in means of groups while the alternative hypothesis states that there's a difference or an association. The interests of medical researchers are varied and research questions result in statement of hypotheses. Examples of such questions are: Is there a significant difference in proportion of low birth weight babies delivered to mothers with single and multiple pregnancies? ; Is there a difference in effects of three antiretroviral drugs on reduction in viral load? ; Is there a correlation between body mass index and systolic blood pressure; or Is there a difference in reduction in blood sugar between a standard hypoglycemic and a new drug. The null hypothesis for the last study objective will be 'There is no correlation between body mass index and systolic blood pressure'. The use and interpretation of p values and confidence intervals will now be discussed.

There are different definitions of the 'p value'. Perhaps the most popular definition is 'The probability of obtaining a value as extreme or more extreme as found in the study if the null hypothesis were true'[1]. Put more simply, it can be defined as the probability that the observed result is due to chance alone[2]. An important point to note in these definitions is the use of phrases 'found in the study' and 'observed result'. The p value only tells us whether what we have observed – which is usually obvious- is statistically significant. This is an important point to note. For example in a study which examined the difference in prevalence of low birth weight deliveries between singleton and multiple

pregnancies, the figures for the prevalence could have been 12.5% for multiple pregnancies and 3.6% for singletons. All the statistical jargon about p value and confidence intervals do not negate the fact that the proportion of low birth babies delivered to mothers with multiple pregnancies is higher than the proportion for singletons. Hence from the initial descriptive statistics used to summarize variables such as proportions and means we have an idea of the results of our study but the statistical significance is what the p value helps to 'endorse'. The interpretation of p values 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. Hence p values less than this number are significant while those above are not.

The confidence interval gives the range of values within which we are reasonably confident that the population parameter lies[2]. The parameter here could be difference in means or proportions of two groups or it could be a measure of association between two variables. The most commonly reported interval is the 95% confidence interval. A way to think about the concept of confidence intervals is to imagine that the study was repeated about a thousand 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. 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 the earlier example on difference in mean reduction in blood sugar between a standard hypoglycemic and a new drug, the null hypothesis is that there is no difference in the mean reduction in blood sugar by the two drugs or that the difference in the means between the two groups being compared is zero. The null value here is zero and any interval computed for the difference in the means which includes zero is not significant. Another set of study designs involves investigation of relationships whereby two variables typically an exposure (or risk factor) e.g. alcohol intake and an outcome such as liver disease are being related. The appropriate measure of association between these variables is the odds ratio and the null value- that is when there is no relationship between alcohol intake and liver disease- is 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 maternal body mass index (in kg/m^2) and babies' birth weight (in kg) where the measure of association here is the correlation coefficient. The null value here is zero and any interval for the correlation coefficient between body mass index and birth weight 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 or both negative - depending on direction - then the difference is significant. Also for odds ratios when the upper and lower limits are both decimals or both whole numbers then we have a significant result.

It is worthy of note that the p value and confidence interval 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. The width of the confidence interval and the size of the p value are related, the narrower the interval, the smaller the p value. 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. In conclusion confidence intervals should be the preferred means of interpreting results of statistical analysis, because in addition to evaluating the role of chance it reflects the degree of variability and the sample size.

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