MAKING SENSE OF STATISTICS FOR FAMILY PRACTITIONERS:

"p" ing with confidence

Ogunbanjo GA, Durrheim DN

It is common in the medical literature to be inundated with "p" values in publications. But what do they mean? To interpret p values, it is important to understand the "null hypothesis". The majority of statistical analyses involve comparisons between groups of study participants, and the comparison of interest is often called the "effect". In general, the null hypothesis states that the results observed in the particular study are no different from what might have been expected as a result of chance alone. It is often the opposite of the research hypothesis that leads to the study. Having set up the null hypothesis, we then evaluate the probability that the observed data could have resulted from chance alone. It is this probability that is called the "p value". In other words, the p value gives probability that the observed difference could have occurred by chance alone, assuming that in reality there is no difference between the populations. The smaller the p value, the more untenable is the null hypothesis. That is to say, a small p value of 0.0025 means that there is only a 25 in 10 000 probability (0.25%) that the observed difference is due to chance alone, assuming that in reality there is no difference.

Interpretation of the p value has proven problematic for many family practitioners, who have desired to interpret the findings of a study and implement these findings in their practice. If a clinical trial comparing two drugs found a result with a p value of, for example 0.3, does this mean that the data could have occurred when the null hypothesis is in fact true and that we cannot rule out the possibility that the null hypothesis is true i.e. the two drugs are equally effective? If the p value is small, for example 0.002, does it mean that the null hypothesis appears implausible because the data could hardly ever arise purely by chance when the null hypothesis is true and that one drug is superior? To answer these questions, the test of the null hypothesis has to be related to whether or not the p value is below a chosen cut-off point and importantly the confidence limits around the measure of effect. Statisticians often arbitrarily set the cut-off point at 0.05. This means that when the p value is <0.05, the result is said to be "statistically significant" and when > or equal to 0.05, it is "not significant". However a statistical test with a p value <0.05 does not automatically imply medical significance. Separate judgement of the medical relevance of an observed difference is necessary before a statistical test is performed. The entrenched use of the "magical" p <0.05 in medical literature has led to the perception that achieving "statistical significance" is a prerequisite for judging a study as successful and documented "non significance" is often perceived as a failure on the part of the researcher(s). It must be emphasized that minor changes in the data, by for example increasing the size of a study, can often result in a shift of the p value to either side of the cut-off point of 0.05. That is why "confidence intervals" are so valuable.

The "confidence interval" shows the uncertainty or lack of precision in the estimate of interest, and thus conveys more useful information than the p value. A confidence interval gives the range of parameter values considered plausible for the population, based on the sample data. The width of the confidence interval provides a reflection of the precision of the measure of effect i.e. the narrower the confidence interval, the more precise the assertion is about where the true value lies. The wider the confidence interval, the less precise is the assertion. A reduction of data variability and an increase in the sample size result in narrower confidence intervals.

The presentation of both the actual p value (not just "p = significant" or "not significant") and confidence interval is desirable for correct interpretation of results. The p value will be less than 0.05 i.e. statistically significant only when the 95% confidence interval does not include zero i.e. the value specified in the null hypothesis. The reason for this relationship is that both methods are based on similar aspects of the theoretical distribution of the test statistic.

In conclusion, it is important to understand the true meaning of the p value, and to realize that statistical significance and clinical importance are not the same thing and it is good practice when presenting results to show both the actual p values and confidence intervals.

Prof. Gboyega A. Ogunbanjo
MBBS, MFGP (SA), M Fam Med (MEDUNSA), FACRRM
Department of Family Medicine & Primary Health Care, Medunsa South Africa
& School of Public Health, James Cook University, Townsville Australia.

Prof. David N. Durrheim
MBChB, DTM&H, DCH, MPH &TM, FACTM
School of Public Health, James Cook University, Townsville Australia.