EDITORIAL

STATISTICS: UNDERSTANDING THE JARGON

*RAVI THOMAS, MD; FRANZCO**

Ophthalmology is a rapidly evolving specialty. "High tech" tests usually provide elaborate data printouts and "high tech" therapies promise better results. The practicing ophthalmologist is constantly under pressure to adopt such new techniques of diagnosis and treatment supported by articles showing a statistically significantly better result with the new (usually more expensive) technique. Regrettably, assessment of the relevant literature to make such decisions is usually limited to scanning for significant ''p'' values of the various statistical tests. Statistical significance, however, is not synonymous with clinical significance. What we clinicians need to know is whether the benefit of a treatment is *clinically significant*? While the all-important $p < 0.05$ or $<.01$ may be significant, assessment of clinical significance includes more: suitability in the type of practice in question, the learning curve, consideration of costs as well as risks of the new intervention. Considering the clinical implications, statistics has truly become too important to just leave to statisticians. This editorial will explain the jargon and introduce common sense concepts that should help the average ophthalmologist to better assess and apply management options. Space constraints restrict us to two clinically important concepts: the meaning of "p" values and confidence intervals, and the number needed to treat (NNT).

"p" (not a urological test) is the probability that the result obtained could have occurred by chance alone (1). "p" values serve a purpose but have major limitations, including dependence on sample size. Large samples can result in small (clinically insignificant) differences achieving statistical significance (low "p" value). Conversely with small samples, clinically significant differences may have a large, statistically insignificant "p" value. Importantly, "p" values do not distinguish between statistical and clinical significance. For example, the mean IOP measured in 3000 ophthalmologists working in North India (mean 16.3; Standard deviation 4 mmHg) was statistically significantly different ($p<0.05$) from the mean IOP of 3000 ophthalmologists working in South India (mean 16.1; Standard deviation 4 mmHg). This statistically significant result could prompt the "pees" amongst us to further study the role of climate or diet in IOP, or recommend relocation for therapeutic purposes. Or we could recognize that large numbers have shown a clinically insignificant difference. Small sample sizes have the opposite effect; the large "p" (statistically insignificant) value may hide a clinically significant result. It is also important to remember that the use of $p<.05$ as significant is just a convention, and lower ''p'' values (.01 or .001) only indicate the smaller role of chance in the result. It is more important to understand that these precise "p" values convey nothing about the actual magnitude of differences between study groups.

Confidence intervals on the other hand quantify the amount of imprecision in the study and should be our preferred measure. (1-3) Thus, instead of telling us that some degree of benefit is likely to occur by chance less than 5% ($p<0.05$) or 1% ($p<0.01$) of the time, the authors can indicate the range of benefit that is likely to occur 95% or 99% of the time. The 95% CI indicates the range within which, 95

.....

* *Queensland Eye Institute & Mater Hospital*
University of Queensland
Brisbane, Australia

out of 100 times, the true value will lie. In other words, we can be 95% certain that the truth lies somewhere inside the 95% CI. There is nothing sacred about 95%; we can opt for a 90% (with a narrower range) or a 99% (wider range) CI too. By examining the range of values included in the confidence interval, we can asses if an intervention seems clinically worth while, irrespective of the "p" value. As an example, in our unpublished prospective series of 60 patients with advanced glaucoma, one patient experienced a "wipe out", an incidence of 1.6%. This estimate (also called a "point estimate") might be considered rare and perhaps not worth discussing with the patient. However the 95% CI around the point estimate (0 to 5%) clearly showed that the true rate of "wipe out" could be as high as 5%, and allow us to better counsel our patients. The CI for various situations, including differences between treatment options can be easily calculated using simple formulae(1,6). Examination of the value obtained (point estimate) as well as the lower and upper limits of the CI allows us to make a judgment about clinical significance (6).

If no outcome of interest (complications or benefit) was encountered, the usual formula doesn't work. For example if we had no "wipe outs" in a series of 10 cases, our formula wouldn't work. In this case a general rule of thumb provided by Sackett is applied (3). If we did 10 cases without encountering a complication, the true complication rate could be as high as 26%. If the number of cases is 25 without complications, the true rate could be as high as 11%. With 50, 75 and 100 cases this becomes 6%, 4% and 3% respectively (3). The value of this rule of thumb becomes obvious in the interpretation of articles or presentations with small numbers but zero complications (or 100% success). We look at the upper end of the confidence interval, compare it to the reported literature, and decide if we consider this a clinically significant difference or not. Here's another, useful "ready reckoner". When considering a rare event, we must look at three times the number in the denominator to be 95% sure we'll come across at least one of those events (4). For example, we may wish to study whether a new technique of glaucoma surgery eliminates expulsive hemorrhage (compared to standard trabeculectomy). Lets say the incidence of significant expulsive hemorrhage with trabeculectomy is one per thousand (thousand is the denominator for this rare complication). To be 95% sure we'll see at least one case of expulsive hemorrhage, we'll need to perform at least 3000 trabeculectomies. To show a meaningful difference between groups we'll need a lot more cases than this in each group; such numbers are hard to come by. This rule of thumb allows determination of the usefulness of a report of other rare complication or side effect in a study that has enrolled the usual 100-200 cases in each group.

Let's move on. A well designed study has been brilliantly executed by respected investigators; appropriate statistics have been used and are statistically highly significant (with tight confidence intervals). Now we are under pressure by patients, the media, the industry as well as our peers to use the treatment recommended by the study. Our question is, "Is the statistically significant claim clinically significant"? To answer that I suggest using the Number Needed to Treat (NNT).

The NNT is an intuitive measure of clinical significance (3,5). NNT tells us the number of patients we need to treat (with the particular drug or procedure) in order to achieve one benefit or prevent one complication; it can be compared to NNT's of alternative treatments. Let us consider a situation where the patient is at risk for a particular outcome or complication. If we do not intervene, there is a chance (risk) that the patient will have the complication - the absolute risk with no intervention. For example, untreated ocular hypertensives in the OHTS study had an absolute risk of progression of 9.5%. If we intervene, there may be a reduced risk of encountering the outcome - the absolute risk of having the complication despite intervention. The absolute risk of progression in treated ocular hypertensives in the OHTS study was 4.5%. The difference between the risk of progression without intervention and that progression despite intervention gives us the "absolute risk reduction" (ARR). The ARR for the OHTS study was 9.5% minus 4.5% = 5%. This, we all agree, is a little difficult to understand. However, the inverse of the ARR, yields

a number - the number of patients we need to treat with our intervention in order to achieve one benefit, or one less complication. For the OHTS study, the inverse of 5%, or 100/5 = 20. We need to treat 20 ocular hypertensives for 5 years in order to prevent one of them from progressing to early glaucoma. (The number needed to harm (NNH) can be calculated in a similar manner.)
In practice, we examine the three elements that make NNT useful. First, we compare the risk of doing nothing at all with the benefits of the recommended procedure. Next, we examine the potential to cause harm - side effects, toxicity, complications and so on, arising out of the intervention. Finally, we try to identify high-risk or high-response sub-groups of patients who have the most to gain from the intervention in question.

Good treatments have a (comparatively) low NNT: all else being equal, treating 10 patients for one benefit is better than treating 100 patients to obtain the same one benefit. As we just calculated, the NNT for treating all ocular hypertensives to prevent a change in the optic disc is 20; (the NNT to prevent an early field defect is 40). When we compare this to an NNT of 7 for treating high risk ocular hypertension (or 5 for preventing progression of primary open angle glaucoma; or 2 for surgery of advanced glaucoma), clinical options and decisions become clearer and easier to explain to patients and colleagues. NNT makes us focus on the high risk groups that will yield a lower NNT. Another good thing about the NNT approach is that it is applied downstream, at the very end - after the study has been done (hopefully in an appropriate manner), and the confidence intervals (not just the p value) have been looked at. An added advantage is that NNT allows us to formally incorporate patient preferences in the choice of treatment (4).
There is a lot to confidence intervals and NNT but space constraints only permit me to provide a couple of ''easy reading'' references including my clinical common sense Bible and that for assessing the literature (2,3). To conclude:

1. **Examine confidence intervals as well as "p" values.**
2. **If you want to use any new therapy or surgery, be sure to check out the number needed to treat (and the number needed to harm).**

REFERENCES

(1) GARDNER M.J., ALTMAN D.G. – Statistics with Confidence. London: British Medical Journal 1989: 6-19.
(2) RIEGELMAN R.K.,HIRSCH R.P. – Studying a study and testing a test 2nd edition. Toronto: Little, Brown and Co. 1989
(3) SACKETT D.L., HAYNES R.B., GUYATT G.H., TUGWELL P. – Clinical Epidemiology: A Basic Science for Clinical Medicine 2nd ed. Toronto: Little, Brown and Co. 1991.
(4) STRAUSS S.E., RICHARDSON S.W., GLASZIOU P., HAYNES B.R. – Evidence - Based medicine: How to practice and teach EBM 3rd edition. 2005. Elsevier Churchill Livingston 2005.
(5) THOMAS R., BRAGANZA A., MULIYIL J.P. – Assessment of clinical significance: The number needed to treat. Ind J Ophthalmol 1996;44(2):113-115.
(6) THOMAS R., BRAGANZA A., MULIYIL J.P. – Confidence with confidence intervals. Indian Journal of Ophthalmology Reviews 1996-1998 . All India Ophthalmological Society 2001.

Professor Ravi Thomas was funded in part by the Prevent Blindness Foundation through Viertels Vision.