Ophthalmic statistics note 2: absence of evidence is not evidence of absence

SCENARIO 1

Patients undergoing vitrectomy surgery for idiopathic full-thickness macular holes used to be routinely advised to follow a strict regime of posturing face down for a variable period (up to 2 weeks) after surgery.[1–3] There was a scientific rationale for this —the tractional forces of gravity would force gases against the macula allowing it to heal more readily. Patients who postured were therefore believed to be less at risk of their macular hole reopening and of the need for repeat surgery to repair the hole. Medicine has clearly changed very significantly over time with a far greater emphasis on patient based outcomes and upon the need for an evidence base to justify practice.[4 5] A senior colleague tells me that he has run a large randomised controlled clinical trial on patients who have had vitrectomies for macular holes. He states that the trial shows there is no difference in failure rates between patients who spent a week posturing face down after surgery and those who did not. He considers that this trial means that it is now unethical to ask patients to posture—particularly because several patients who did posture fed back to him how uncomfortable they found posturing. I ask him for a little more information about the trial and learn that it was a randomised controlled clinical trial with larger numbers of patients than typically found in ophthalmic surgical studies of 200 patients in each arm. Of those who spent a week posturing face down, one required repeat surgery. Of those who did not, two required repeat surgery. There is a published p value from a Fisher's exact test that was used to compare failure rates in the two groups of 0.999 and what seems to me to be an entirely cogent argument that this demonstrates no need for posturing (see online supplementary appendix 1, table 1 for results of analysis). I have a persistent doubt however that something is not quite right with this argument and the issue leaves me pondering somewhat. I decide to go back to grass roots and search the internet for a definition of a p value.

Table 1 Application of flow diagram approach to Scenario 1

1. Research question	Is posturing advisable in patients undergoing vitrectomy surgery?
2. Null hypothesis	There is no difference in risk of failure between patients undergoing vitrectomy surgery in Group A (posturing) and Group B (no posturing).
3. Result	Odds in Group A=1/199 Odds in Group B=2/198 OR=2.02 95% CI=0.18 to 22.3 p=0.999
4. Interpretation	The best estimate of the OR is 2, that is, failure is twice as common in the non-posturing group as it is the posturing group. There is however a lot of uncertainty with this estimate. The trial data means that it is plausible that the odds are the same in the two groups; however it is still possible that the odds are actually more than 20 times in the non-posturing group or as little as a fifth as high. If we were to simply consider the p value none of this uncertainty would be apparent and we would be far more tempted to simply say that there is no evidence of a difference in failure rates.

> The p value is the probability of obtaining the observed data or data that were more extreme due to chance if the null hypothesis were true.

I am somewhat perplexed by the term null hypothesis and again resort to the internet.

> The null hypothesis is the situation you believe exists (in this scenario that the effect of interest is zero) and you perform a significance test to see whether there is sufficient evidence for you to reject the null hypothesis.

My interpretation of this in this scenario is that the null hypothesis is that the risk of failure with posturing following surgery is the same as the risk of failure with no posturing. Continuing in this vein, if there truly is no difference between the risks in the two groups the probability of observing the difference that I observed in this trial or something more extreme (two failures in the non-posturing group vs one failure in the posturing group) by chance alone is 0.999. I recall that p values must lie between 0 and 1, with a value of 0 meaning impossible and a value of 1 meaning absolute certainty. Here, a p value of 0.999 indicates that there is a very high chance that I would see a difference in proportions of 2/200 versus 1/200 due to chance alone and thus I have no evidence to reject the null hypothesis.

What does this mean? I have no evidence of a difference in failure rates and thus no evidence to support the use of posturing. Can I now simply advocate that it is safe for all patients not to posture? Patients have reported that they do not enjoy posturing, but the prospect of repeat surgery after an initial failure is also very daunting.

DISCUSSION

This scenario is given to illustrate challenges faced when interpreting statistical non-significance. Altman and Bland discuss this issue in a paper entitled 'Absence of evidence is not evidence of absence'.[6] Altman and Bland advocate that when presented with the statement 'there is no evidence that' consideration must be given as to whether absence of evidence means that there is no information at all. They suggest estimating the effect with a Confidence interval (CI) rather than simply looking at p values. Figure 1 illustrates a simple flow diagram approach based on this. Table 1 illustrates the application of the flow diagram approach to Scenario 1. In the scenario given, the odds of failure in posturing patients were 1/199, while those in the non-posturing patients were 2/198. (Odds are commonly seen in this context rather than risks, but for rare events, the OR and relative risk are approximately equal). Clearly the odds are slightly higher in the non-posturing group. The OR is 2 with a 95% CI of 0.18 to 22.3—see online supplementary appendix 1, table 2 for the computation of this. The odds of failure are estimated to be twice as common in the non-posturing patients as in the posturing patients but the CI

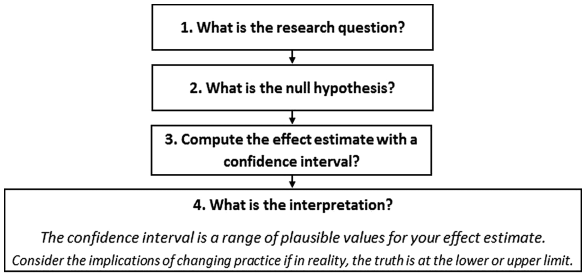


Figure 1 Flow diagram when presented with non-significance.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2013-304811 on 3 March 2014. Downloaded from http://bjo.bmj.com/ on March 29, 2024 by guest. Protected by copyright.

Table 2 Application of flow diagram to Scenario 2

1. Research question	For patients treated for age-related macular degeneration, are SAEs more common on ranibizumab or bevacizumab? A meta-analysis of the CATT and IVAN trials
2. Null hypothesis	There is no difference in SAE risk between patients in Group A (ranibizumab) and Group B (bevacizumab).
3. Result	Odds of 1 or more SAE in bevacizumab=314/568 Odds of 1 or more SAE in ranibizumab=271/642 OR=0.76 95% CI=0.63 to 0.93 p=0.003
4. Interpretation	Individual trial results: OR for IVAN trial=0.94 (95% CI 0.65 to 1.35) and OR for CATT trial=0.70 (95% CI 0.55 to 0.89). The results of the meta-analysis indicate the need for patients to be informed of the disparity in reported rates of SAEs between the two drugs.

CATT, comparison of age-related macular degeneration treatments trial; IVAN, inhibition of VEGF in age-related choroidal neovascularisation; SAE, serious adverse event.

indicates considerable uncertainty in this estimate. The data are indeed consistent with there being no difference between the two trial arms in that the CI includes an OR of 1 (no difference) but the data are also consistent with the odds of failure in the posturing arm being as much as 22 times the odds of failure in the non-posturing arm (the upper limit of the CI) or indeed as little as a fifth as high. I feel much less confident now in stating that there is no difference between the treatment arms and am not sure that I entirely agree with my colleague about it being unethical to ask patients to posture when there is so much uncertainty in my estimate.

By computing a CI uncertainty is revealed which wasn't apparent when simply looking at a p value. So should patients be posturing or not? The answer is currently unclear. What hopefully is clear is that absence of evidence is not evidence of absence and to assume that it is the case is unwise.

Most randomised trials wish to determine whether a treatment is superior to the current standard treatment. However non-inferiority and equivalence trials are becoming more common in the medical literature.[7] A non-inferiority trial seeks to determine whether a new treatment is not worse than the standard treatment by more than an acceptable amount (known as the non-inferiority margin). An equivalence trial seeks to determine whether a new treatment is therapeutically similar to a standard treatment, that is, whether a new treatment differs from the standard treatment by no more than the non-inferiority margin. It is important to note that the term equivalence has in the past been used in error to report negative results of superiority studies—such trials often lacked statistical power to rule out important differences.[8 9] The trial conducted by my colleague has not demonstrated equivalence as most clinicians and patients would consider an OR of 22.3 (which lies within the CI) as an unacceptable difference, although defining the non-inferiority margin can present a real challenge to researchers.

SCENARIO 2

The issue is of particular relevance when considering adverse events. These may be rare yet catastrophic for the individuals affected and their families. For treatments that are in widespread use even small differences in risk can equate to sizeable numbers of people and very large studies are needed to demonstrate differences. The recent controversy regarding the use of bevacizumab (Avastin) or ranibizumab (Lucentis) for the treatment of age-related macular degeneration (AMD), the leading cause of certifiable sight loss in the UK, very much centres around the absence of evidence issue.[10] Ranibizumab was licensed for ocular use but costs substantially more than bevacizumab which does not have a marketing authorisation in this indication. Prior to licensing for AMD treatment, many people chose to have treatment with bevacizumab since without any treatment they faced rapid blindness and they preferred to accept the possibility of increased side effects with the unlicensed product. A large body of evidence built up as a result of off license use, which suggested little evidence of harm however this evidence was mostly from case series rather than Level 1 evidence. The ABC study demonstrated that bevacizumab was better than standard National Health Service (NHS) care (prior to licensing of ranibizumab) and that it appeared to offer similar benefits to ranibizumab while not appearing to increase harms.[11] The study was not designed to have adequate power to examine safety concerns and so failure to detect a difference should not equate to evidence of safety. The harms under consideration were not trivial and included arteriothrombotic events and heart failure. Two large studies, inhibition of VEGF in age-related choroidal neovascularisation (IVAN) and comparison of age-related macular degeneration treatments trial (CATT), have recently been reported, both of which suggest that the drugs are indeed very similar with respect to harms and safety, and calls to license bevacizumab for use in AMD in the NHS have been made.[12–14] IVAN and CATT were conducted in different parts of the world, yet the methodology was sufficiently similar to enable results from the two studies to be validly combined using a technique called meta (Greek for after) analysis.[15] Recent pooling of results from the two studies has suggested that a higher proportion of patients who receive bevacizumab experience one or more serious adverse events, although numbers are small and so the jury is still out. Table 2 illustrates the application of the flow diagram approach to Scenario 2.

Catey Bunce,[1] Krishna V Patel,[1] Wen Xing,[1] Nick Freemantle,[2] Caroline J Doré,[3] On behalf of the Ophthalmic Statistics Group

[1]NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK
[2]Department of Primary Care and Population Health, University College London, London, UK
[3]UCL Clinical Trials Unit, University College London, London, UK

Correspondence to Dr Catey Bunce, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, City Road, London, EC1V 2PD, UK; c.bunce@ucl.ac.uk

Lesson learnt Absence of evidence ≠ Evidence of absence.

Collaborators The following additional members of the Ophthalmic Statistics Group were given the opportunity to view the final paper prior to submission and provide suggestions and comments: Jonathan Cook, David Crabb, Phillippa Cumberland, Gabriela Czanner, Paul Donachie, Andrew Elders, Marta Garcia-Fiñana, Rachel Nash, Neil O'Leary, Toby Prevost, Chris Rogers, Luke Saunders, Selvaraj Sivasubramanium, Irene Stratton, Joana Vasconcelos and Haogang Zhu.

Contributors CB drafted the paper. CB, KVP and WX reviewed and revised the paper. CJD and NF conducted an internal peer review of the paper.

Competing interests The posts of CB, KVP and WX are partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Provenance and peer review Not commissioned; externally peer reviewed.

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bjophthalmol-2013-304811).

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2013-304811 on 3 March 2014. Downloaded from http://bjo.bmj.com/ on March 29, 2024 by guest. Protected by copyright.



OPEN ACCESS



Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/



To cite Bunce C, Patel KV, Xing W, *et al*. *Br J Ophthalmol* 2014;**98**:703–705.

Published Online First 3 March 2014

Br J Ophthalmol 2014;**98**:703–705. doi:10.1136/bjophthalmol-2013-304811

REFERENCES

1 Wendel RT, Patel AC, Kelly NE, *et al*. Vitreous surgery for macular holes. *Ophthalmology* 1993;100:1671–6.
2 Thompson JT, Glaser BM, Sjaarda RN, *et al*. Effects of intraocular bubble duration in the treatment of macular holes by vitrectomy and transforming growth factor-beta2. *Ophthalmology* 1994;101:1195–200.
3 Thompson JT, Smiddy WE, Glaser BM, *et al*. Intraocular tamponade duration and success of macular hole surgery. *Retina* 1996;16:373–82.
4 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
5 Sacket DL, Rosenberg WM, Gray JA, *et al*. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
6 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
7 Piaggio G, Elbourne DR, Pocock SJ, *et al*. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–604.
8 Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. *Biom J* 2005;47:12–27.
9 Eyawo O, Lee CW, Rachlis B, *et al*. Reporting of noninferiority and equivalence randomized trials for major prostaglandins: a systematic survey of the ophthalmology literature. *Trials* 2008;9:69.
10 Bunce C, Xing W, Wormald R. Causes of blindness and partial sight certifications in England and Wales: April 2007–March 2008. *Eye* 2010;24:1692–9.
11 Tufail A, Patel PJ, Egan C, *et al*. ABC Trial investigators Bevacizumab for neovascular age related macular degeneration (ABC Trial): multicentre randomised double masked study. *BMJ* 2010;340:c2459.
12 CATT Research Group. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med* 2011;364:1897–908.
13 Chakravarthy U, Harding SP, Rogers CA, *et al*. Ivan Study Investigators Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration: one-year findings from the IVAN randomized trial. *Ophthalmology* 2012;119:1399–411.
14 Chakravarthy U, Harding SP, Rogers CA, *et al*. Ivan Study Investigators Alternative treatments to inhibit VEGF in age-related choroidal neovascularisation: 2-year findings of the IVAN randomised controlled trial. *Lancet* 2013;382:1258–67.
15 Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2013-304811 on 3 March 2014. Downloaded from http://bjo.bmj.com/ on March 29, 2024 by guest. Protected by copyright.