Choosing endpoints in clinical studies and trials

BC Chauhan

Abstract

Currently, it is challenging to select endpoints for measuring disease progression in glaucoma clinical trials. The endpoints that are used are not biological, but statistical, events because there is no reference standard for progression. Progression rates depend on the criteria used to define change, and in the absence of a reference standard for progression, reasonable criteria can be determined by paying attention to specificity (ie false-positive rate). Past randomized-clinical trials (CITGS, OHTS, and CNGT) have not made a accurate estimate of specificity at the time of trial design. It is not meaningful to compare the sensitivities (or more accurately the 'hit rate') of different progression criteria unless their respective specificities are equivalent. The ideal approach to compare progression rates, therefore, would be to equalize the specificities with each approach. However, owing to the lack of an external standard, this is difficult to achieve with reasonable confidence. With the advancement of research in visual field and imaging, clinical endpoints will have to be reviewed to incorporate new criteria for progression detection and measurement.
Eye (2007) **21**, S34–S37; doi:10.1038/sj.eye.6702886

Keywords: glaucoma; visual fields; optic disc morphology; structural function; clinical trials; endpoints

Department of
Ophthalmology and Visual
Sciences, Dalhousie
University, Halifax, Canada

Correspondence:
BC Chauhan,
Department of
Ophthalmology and Visual
Sciences,
Dalhousie University,
Halifax, Canada
Tel: + 1 902 473 3202.
E-mail: bal@dal.ca

Introduction

It is currently challenging to select endpoints in glaucoma clinical studies. Endpoints are statistical, rather than biological, events as no reference standard for progression is available, because one has to rely on the very test(s) to make an estimate of efficacy as those that are used to define glaucoma and monitor its progression. Similarly, findings from one study cannot be easily compared with those of another as strategies for detecting progression vary widely.

Event- and trend-based analyses

Some of the data available are from 'cross-sectional' analysis of longitudinal data, which may be counterintuitive and inappropriate. Patients are followed up longitudinally, yet cross-sectional tools are used to detect progression, ie when a test yields an abnormal result from a previously normal one.

Most large clinical trials have used event-based analysis,[1–4] whereby progression is deemed to occur when a test reaches a predefined threshold for change.[5] Compared with the event-based approach, trend-based approaches are more appropriate for providing information on the rate of progression, which is important for clinical decisions. In trend analysis, comparison is made with the patient's own previous data. The tools that have been developed for longitudinal analysis include the glaucoma progression analysis (GPA), the topographical change analysis on the HRT[6,7] and linear regression techniques.

The trend-based approach has certain advantages, particularly when many (>8 or so) examinations are available as the visual field (VF) measurements over time can be used to determine the rate and magnitude of change over time. Furthermore, by using the patient's own VF data, a permutation analysis where the sequence of the data is reordered in a form of simulation analysis to generate a distribution of possible slopes can be determined. It is assumed that this represents the noise distribution on to which the actual trend can be compared. The advantages of this technique are the preservation of the variance structure.

Although event analysis may identify test locations that appear progressive, with as few as three test results, it is dependent on the degree of change exceeding test–retest variability, which is high for damaged locations. To maintain reasonable specificity, most investigators have found it necessary to have glaucoma change probability (GCP) points outside normal limits to be confirmed on two or more tests.

While longitudinal analysis is a more appropriate approach than cross-sectional analysis, it has only rarely been used in the recent glaucoma trials. The only such published study is the early manifest glaucoma trial (EMGT), which used the number of points that occur in a GPA-type analysis based on pattern deviation.[8] The Canadian Glaucoma Study[9] uses the same type of analysis but based on total deviation.

Progression rates are method-dependent

When so many different criteria have been used in studies, and in the absence of an accepted standard of glaucoma progression, it is not surprising that a review of outcomes from clinical trials shows great variation. For example, the two studies that have used an untreated arm are the Collaborative Normal-Tension Glaucoma Study (CNTG) and the EMGT study.[1,8] The progression rate in the treated patients in the CNTG is 12%, while the EMGT showed an almost four-fold (45%) difference. Progression rate, then, is arbitrary and dependent on the criteria chosen. Someone who wishes to advocate a particular therapy as very protective could choose an extremely conservative criterion, where nobody would progress. Alternatively, to show that a technique is very sensitive at picking up damage, a liberal criterion could be chosen. It is possible, in the same patients, to describe progression rates that vary from 10% progression to 60%, simply by varying the method used. Thus, the criterion used to define progression must always be borne in mind when considering study data.

Evaluation of criteria for measuring VF progression

Computer simulation is helpful in understanding and evaluating the different criteria used. Real data are input into the computer and then progression pathways modelled in, using algorithms, where parameters such as direction and variability may be modulated.[10] Vesti et al[11] have developed a computer model for studying VF progression, which has paths for low, moderate and high variability, by taking into account most factors reported to affect threshold variability. Specificities were calculated by using the same VF of each patient as both the initial and final field (no progression) under conditions of moderate and high-variability. This model
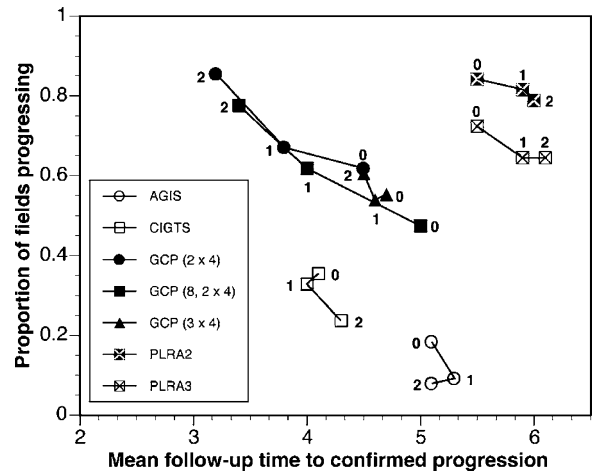

Figure 1 Proportion of progressing cases as a function of the time to confirmed progression for seven methods to analyse glaucomatous visual field progression, using simulations with no threshold variability (0), moderate-threshold variability (1), and high-threshold variability (2). The simulations are based on two examinations per year. (Reproduced from Vesti et al[11] with permission of Invest Ophthalmol Vis Sci via Copyright Clearance Center.)

permits control over conditions of progression, and provides information that complements real patient data because simulated longitudinal VF data can be generated without variability.

Sensitivity of methods for analysing VF progression

The computer model has been used to compare the different progression algorithms used in glaucoma. The findings show that the AGIS approach takes a relatively long time to detect progression, and it does so in a low number of subjects. CITGS detects about twice as many progressors at about 6–12 months earlier. According to the GCP types of analysis, more patients progress than are detected using AGIS and CITGS. Depending on the variability criteria, the GCP-type criteria may pick up progression earlier. Linear regression techniques (pointwise linear regression analysis (PLRA)) result in a high number of progressing fields but take the longest to pick up progression (Figure 1).

Specificity of methods for analysing VF progression

Specificity is an important consideration as it supports the significance of the findings. The criteria of AGIS are very highly specific, which may cause the test to miss some patients who are progressing (Figure 2). The CIGTS and GCP criteria are less specific than those of AGIS but are still highly specific and have a higher sensitivity (Figure 2). The PLRA techniques also result in high specificity.
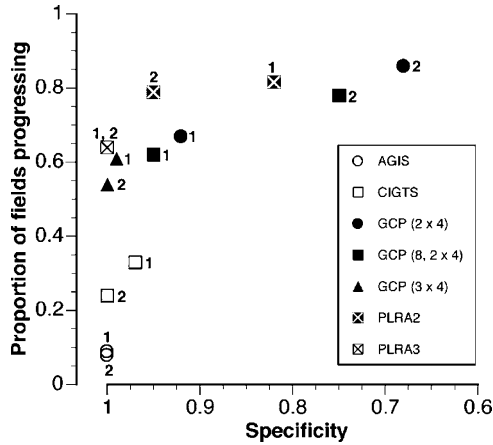
Figure 2 Proportion of progressing cases as a function of specificity for seven methods to analyse glaucomatous visual field progression, using simulations with no threshold variability (0), moderate-threshold variability (1), and high-threshold variability (2). The simulations are based on two examinations per year. (Reproduced from Vesti *et al*[11] with permission of *Invest Ophthalmol Vis Sci* via Copyright Clearance Center.)

Selection of criteria for VF progression in the 'real world'

It is not meaningful to compare sensitivities between different tests unless their respective specificities are equal. A test that has lower specificity (ie falsely flags progression where no true change has taken place) is likely to be more sensitive to change compared to a test with high specificity. The ideal approach to compare progression rates, therefore, would be to equalize the specificities with each test. However, owing to the lack of an external standard, this is difficult to achieve with reasonable confidence. Owing to sampling variation, the limits of test–retest variability derived in one sample of patients may not be identical to those that would have been derived from another sample, yet even small differences in specificity are likely to have a large effect on a test's sensitivity to change. A meaningful comparison of progression rates, based on empirical data, is therefore very difficult to make.

There are three ways in which false-positive rate can be assessed. Test–retest data are often used to derive limits for variability but are probably not entirely independent and may not be the best choice to use. Parallel cohorts of control subjects can be tested to see how frequently they progress and if this provides an indication of false-positives. This gives an independent measure of validation, but it assumes that the eyes of the control subjects do not change, which is not likely in reality. The third approach is to look for parallel improvement rates in patients; this assumes that all improvement is 'noise'.
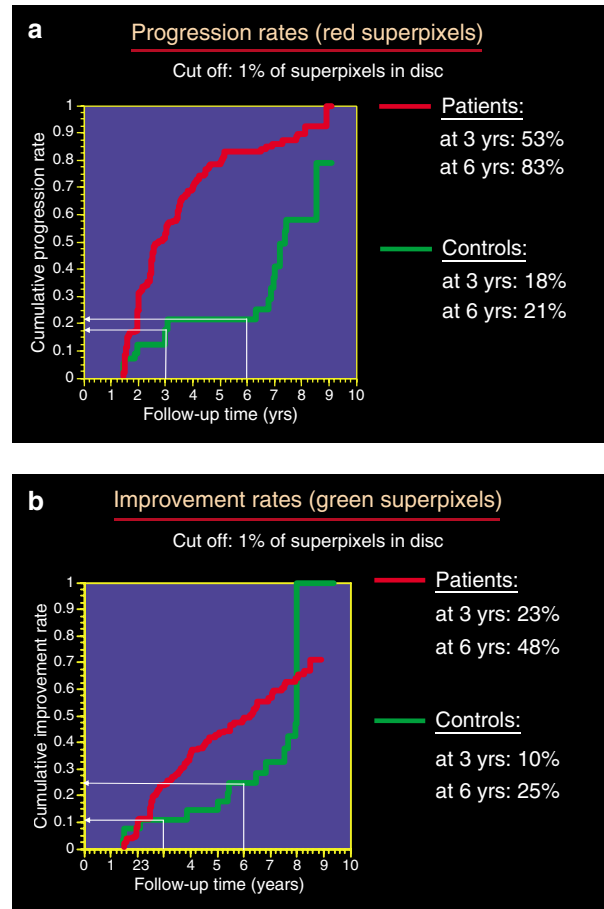

Figure 3 (a) Progression rates in topographical change analyses: superpixels with statistically significant negative height change ($P < 0.05$) in three images, compared with baseline, are marked in red. (Reproduced with permission from Artes PH and Chauhan BC, private communication.) (b) Improvement rates in topographical change analysis: superpixels with statistically significant positive height change ($P < 0.05$) in three images, compared with baseline, are marked in green. (Reproduced with permission Artes PH and Chauhan BC, private communication.)

Progression rates as determined by imaging techniques

A statistical technique for detecting serial topographic changes in the optic nerve head (ONH) has been developed.[6,7] A new version of the HRT3 software was used to assess progression by determining the statistical significance of surface height changes from baseline. Progression was expressed by the number of superpixels with statistically significant negative height change (red superpixels), and improvement was expressed by the number of superpixels with statistically significant positive height change (green superpixels), in three consecutive images taken at baseline and at follow-up (Figure 3a and b).

The criteria for change were described in terms of the largest cluster of red superpixels within the ONH

boundary. To account for the large physiological variation in ONH, the size of this cluster was expressed as a percentage relative to the total number of superpixels within the contour line. With a cut-off point for progression set at only 1% of the superpixels within the contour line, the progression rate was very high—in fact, at 6 years, 83% of patients progressed. However, a similar high rate of progression was seen in controls, and so this criterion did not have high specificity. Improvement rates were also fairly high, and so the criteria needed to be modulated until there was a trade-off between improvement rates and progression rates that was meaningful.[9]

Conclusion

Selecting endpoints in glaucoma trials is very challenging. In the absence of a reference standard for progression, reasonable criteria can only be derived if special attention is paid to specificity, or by determining the false-positive rate. Current knowledge of VF and ONH has come a long way since studies such as CITGS, OHTS and CNGT were designed. It is time to apply this knowledge to devising clinical trial endpoints.

References

1 Collaborative Normal-Tension Glaucoma (CNTG) Study Group. Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. *Am J Ophthalmol* 1998; **126**: 487–497.

2 AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): The relationship between control of intra-ocular pressure and visual field deterioration. *Am J Ophthalmol* 2000; **130**: 429–440.

3 Lichter PR, Musch DC, Gillespie BW, Guire KE, Janz NK, Wren PA, CIGTS Study Group. Interim clinical outcomes in the Collaborative Initial Glaucoma Treatment Study comparing initial treatment randomized to medications or surgery. *Ophthalmology* 2001; **108**: 1943–1953.

4 Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP *et al.* The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open angle glaucoma. *Arch Ophthalmol* 2002; **120**: 701–713.

5 Hitchings R. Update on the natural history of glaucoma. 2007; **21**(Suppl 1): S2–S5.

6 Chauhan BC, Blanchard JW, Hamilton DC, LeBlanc RP. Technique for detecting serial topographic changes in the optic disc and peripapillary retina using scanning laser tomography. *Invest Ophthalmol Vis Sci* 2000; **41**: 775–782.

7 Chauhan BC, McCormick TA, Nicolela MT, LeBlanc RP. Optic disc and visual field changes in a prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography with conventional perimetry and optic disc photography. *Arch Ophthalmol* 2001; **119**: 1492–1499.

8 Heijl A, Leske MC, Bengtsson B, Bengtsson B, Hussein M, Early Manifest Glaucoma Trial Group. Measuring visual field progression in the early manifest glaucoma trial. *Acta Ophthalmol Scand* 2003; **81**: 286–293.

9 Canadian Glaucoma Study Group. Canadian Glaucoma Study: 1. Study design, baseline characteristics, and preliminary analyses. *Can J Ophthalmol* 2006; **41**: 566–575.

10 Spry PG, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci* 2000; **41**: 2192–2200.

11 Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci* 2003; **44**: 3873–3879.