Methodological Variations in Estimating Apparent Progressive Visual Field Loss in Clinical Trials of Glaucoma Treatment

Joanne Katz, ScD; Nathan Congdon, MD, MPH; David S. Friedman, MD, MPH

Objectives: To compare methods to estimate the incidence of visual field progression used by 3 large randomized trials of glaucoma treatment by applying these methods to a common data set of annually obtained visual field measurements of patients with glaucoma followed up for an average of 6 years.

Methods: The methods used by the Advanced Glaucoma Intervention Study (AGIS), the Collaborative Initial Glaucoma Treatment Study (CIGTS), and the Early Manifest Glaucoma Treatment study (EMGT) were applied to 67 eyes of 56 patients with glaucoma enrolled in a 10-year natural history study of glaucoma using Program 30-2 of the Humphrey Field Analyzer (Humphrey Instruments, San Leandro, Calif). The incidence of apparent visual field progression was estimated for each method. Extent of agreement between the methods was calculated, and time to apparent progression was compared.

Results: The proportion of patients progressing was 11%, 22%, and 23% with AGIS, CIGTS, and EMGT methods, respectively. Clinical assessment identified 23% of patients who progressed, but only half of these were also identified by CIGTS or EMGT methods. The CIGTS and the EMGT had comparable incidence rates, but only half of those identified by 1 method were also identified by the other.

Conclusions: The EMGT and CIGTS methods produced rates of apparent progression that were twice those of the AGIS method. Although EMGT, CIGTS, and clinical assessment rates were comparable, they did not identify the same patients as having had field progression.

Arch Ophthalmol. 1999;117:1137-1142

AUTOMATED static threshold perimetry is a widely used means of monitoring visual field loss in patients with glaucoma. Several analytic methods for identifying patients with visual field progression and estimating the rate of progression have been proposed, but there is no general agreement about which of these approaches is most sensitive and specific for detection of visual field loss.1-14 The problem of how to select an optimal method of visual field analysis is exacerbated by the lack of a criterion standard for what constitutes visual field progression. Detection of visual field progression is important not only for clinical practice but also for randomized trials of glaucoma treatment, in which a major outcome is visual field progression. Several ongoing clinical trials use different analytic strategies for detecting progression. Because the eligibility criteria and the type of treatment comparisons vary across trials, it will be difficult to compare the rates of progression from one trial to another. In this study, we apply the analytic methods used by each of 3 large randomized clinical trials of glaucoma treatment to a common data set to better identify differences (or similarities) between these methods that might provide insight into how to compare the results of the trials regarding their visual field outcomes.

The Advanced Glaucoma Intervention Study (AGIS) uses a scoring system to grade the severity of field loss at any one point in time. The score is based on the extent of depression at different locations in the Program C-24 (Humphrey Instruments, San Leandro, Calif) field test found in the total deviation plot.15-18 The score ranges from 0 to 20 on an integer scale, with 0 being no field loss and 20 being end-stage disease. A “reliability” score is also used to assess the quality of the visual field test. Information about the number of presentations, the catch trial responses, and short-term fluctuation is used to calculate a score that ranges from 0 to 7, with 7 being extremely unreliable. In
MATERIALS AND METHODS

Using a common data set of annual visual field measurements of patients with glaucoma, the incidence of visual field progression was calculated for each approach taken by 3 National Eye Institute–sponsored clinical trials. The common data set was obtained from patients enrolled in the Glaucoma Screening Study, a natural history study of risk factors for glaucoma at the Wilmer Institute, Johns Hopkins School of Medicine, Baltimore, Md, from 1981 through 1993, sponsored by the National Eye Institute. A diagnosis of primary open-angle glaucoma was made by 1 of 4 study ophthalmologists during a comprehensive eye examination performed according to study protocol. All patients also had a history of elevated intraocular pressure (>21 mm Hg) documented on at least 2 occasions before treatment. Treatment during the trial was determined by patients’ ophthalmologists rather than by study investigators. Patients underwent annual automated perimetric testing using Program 30-2, stimulus size III, and the full threshold strategy of the Humphrey Field Analyzer administered by trained technicians using the study protocol. Stereo- scopic color photographs of the optic disc were taken annually, and vertical and horizontal cup-disc ratios were read by 2 ophthalmologists with reasonable agreement. To be eligible for this visual field analysis, the eyes of study patients had to have their first 2 study field values outside normal limits on the Glaucoma Hemifield Test, a minimum of 5 years of follow-up, and at least 6 field tests. If a patient had a visual acuity worse than 20/100 in either eye, that eye was excluded. A total of 67 eyes of 56 patients met these criteria.

Two fellowship-trained glaucoma specialists (N.C. and D.S.F.) masked to results of the statistical analysis of visual field progression reviewed the visual fields of all patients in their temporal sequences. Clinicians were not provided with any other clinical information about the patients. Each sequence was classified as “definite progression,” “possible progression,” “stable,” “improved,” or “too unreliable to assess.” The standard Humphrey printouts for each visual field test were provided but the Glaucoma Change Probability printouts were not. After independent review, the differences in ratings between clinicians were adjudicated to produce a “clinical” assessment of visual field progression. Kappa statistics were calculated to assess agreement between clinicians. For comparison with the statistical methods of progression, definite progression was defined as clinical progression. All other categories were classified as “no progression.” For clinical assessment of “improvement,” both clinicians had to agree that the field test results had improved.

The 3 clinical trial methods for identifying eyes (and patients) that demonstrate progressive visual field loss were applied to the data set of patients from the Glaucoma Screening Study. In this study, field testing was performed annually so that confirmation of progression used tests done 1 and 2 years after initial progression was detected. The reliability criteria were applied to each method as described in the different trial manuals and publications. If a field measurement was considered “unreliable,” the next available reliable test result was substituted. Because all methods required at least 2 field tests to determine eligibility and baseline, the number of field test results available for follow-up was the same for all methods. The only exception to this was if there were unreliable field tests that would be discarded by one method that were not discarded by another method.

The incidence of visual field progression was estimated for each method using all 67 eyes of 56 patients, and the overlap between the 3 methods was identified with a Venn diagram. The incidence rates for the 3 methods were also compared with that of the adjudicated clinical assessment and with changes in the vertical cup-disc ratio obtained from color photographs. An additional estimate of progression excluded patients who would not have been eligible for each of the trials. Analyses were done by eye and by patient (progression in either eye defined the patient as having progression). The prevalence of improvement was also estimated by applying the same definitions as those used for deterioration but in the opposite direction. For example, improvement in the AGIS would mean that the score decreased by at least 4 from baseline, as confirmed on 2 subsequent consecutive tests.

Written informed consent for participation in the Glaucoma Screening Study was obtained from each patient. Ethical approval for that study was provided by the Joint Committee on Clinical Investigations of Johns Hopkins School of Medicine. For this study, only secondary analysis was done with computerized records that did not include personal identifiers. For this reason, the Committee on Human Research of Johns Hopkins School of Hygiene and Public Health determined that the study was exempt from further ethical review.

the AGIS, 2 preintervention tests are done no more than 60 days apart—the first to determine eligibility (the score must be between 1 and 16, and the reliability score must be <3, or the test must be repeated) and the second as a baseline for subsequent field tests without the eligibility constraints of the first test (except for reliability). Progression of visual field loss is defined as an increase from the baseline reference score of 4 or more confirmed by 2 consecutive, reliable tests. The score change of 4 or more was derived from baseline data from the AGIS, in which only 5% of participants had a score deterioration of 4 or more between 2 tests taken 1 to 6 weeks apart. The Collaborative Initial Glaucoma Treatment Study (CIGTS) also uses a scoring system but is based on P values obtained from the total deviation plot rather than on the actual decibel deviations from age-matched control subjects used by the AGIS. A score from 0 through 4 is assigned to each location based on how small the P value is at that location and on how small the P values of its neighbors are. Based on 52 locations in the Program C-24 test (excluding the 2 locations above and below the blind spot), the sum of the scores ranges from 0 through 208. The score is then standardized to range from 0 through 20 by dividing it by 10.4. The CIGTS also has a reliability score that is the same as that for the AGIS except that it does not include the number of presentations and hence has a maximal score of 6 rather than 7. A test is considered reliable if the score is less than 4. All unreliable tests must be repeated. For eligibility, 2 baseline visual field measurements must be reliable and out-
RESULTS

The mean ± SD age of the 56 patients was 62.0 ± 10.4 years, and 45% were black. Follow-up ranged from 5 to 9 years (median, 6.3 years). The number of field tests per eye ranged from 6 to 14 (median, 7). At baseline, 85% of patients were taking medication for glaucoma and 27% had undergone previous laser or filtering surgery. Thirty-four percent of patients underwent glaucoma surgery during follow-up, 9% underwent combined cataract and filtering procedures, and 6% underwent cataract surgery only. The average vertical cup-disc ratio of these patients was 0.62 at the time of the baseline field test.

The prevalence of unreliable field tests was 4% and 1% using the AGIS and CIGTS definitions, respectively. One percent of field measurements were not used in the EMGT analysis because they had abnormally high sensitivity. The prevalence of catch trial responses of 50% or more was 4% for fixation losses, 1% for false-positive responses, and 2% for false-negative responses. The average mean deviation at baseline was −7.43 dB, and the average corrected pattern SD was 7.09 dB. Based on the classification of Hodapp et al,23 28% had early defects, 30% had moderate defects, and 42% had severe defects.

Agreement between the 2 clinicians using all categories (definite, possible, stable, improved, and unreliable) was substantial, with a weighted κ statistic of 0.69 (95% confidence interval, 0.53-0.85) and 81% agreement. Agreement using only 2 categories—definite and no progression —was almost perfect, with a κ of 0.87 (95% confidence interval, 0.63-0.99) and 96% agreement. Clinicians agreed that 5 eyes (8%) had a sequence of field tests that were too unreliable to classify progression. These were included in the no progression category for purposes of estimating the incidence of progression.

The incidence of progression among patients was 11%, 22%, and 23% using the AGIS, CIGTS, and EMGT methods, respectively (Table 1). Clinical assessment produced identical incidence rates to that of the EMGT method. The incidence by eye was slightly lower than by patient for each method. The incidence of improvement was 4%, 0%, and 11% using the AGIS, CIGTS, and EMGT methods, respectively. The incidence of improvement was 11% based on clinical assessment. Five eyes (8%) of 5 patients (9%) were identified by all 3 statistical methods as having visual field progression (Figure). Clinical assessment also identified these 5 patients as having significant progression.

The 5 patients who were considered to have progressed by all 3 methods were first identified as having progressed by the AGIS and CIGTS methods at the same clinic visit; the EMGT method identified these same 5 patients an average of 14 months earlier. Each patient was identified by the EMGT method at least 1 year earlier than by the other 2 methods.

Based on AGIS eligibility criteria, 6 eyes (9%) of 6 patients (11%) who met Glaucoma Screening Study eligibility criteria would have been excluded. Two patients were excluded because they had baseline scores of 18 and the remainder because they had scores of zero. Eleven eyes (16%) of 10 patients (18%) would have been ineligible for the CIGTS. These patients all had 2
field values outside normal limits on the Glaucoma Hemifield Test, but they failed eligibility because they did not have at least 3 contiguous locations with P values less than .02 on the total deviation plots. All patients would have been eligible for the EMGT because of our criteria for selection. The exception to this was the requirement that the abnormality on the Glaucoma Hemifield Test be in the same sector on 2 occasions. Because the Humphrey printouts do not provide this information, we attempted to infer it from the available data in the probability maps. This process did not yield any exclusions.

After excluding patients who did not meet the specific eligibility criteria for each study, the incidence of progression was 12%, 26%, and 23% and the incidence of improvement was 2%, 0%, and 11% using the AGIS, CIGTS, and EMGT methods, respectively. These rates are comparable to those calculated without regard to eligibility. The EMGT method of identifying visual field progression was in fair agreement (κ = 0.30; 95% confidence interval, 0.06-0.54) with an increase in vertical cup-disc ratio of 0.1 or more from baseline to final examination (Table 3). No agreement was seen between AGIS and CIGTS methods and cup-disc ratio changes. There was also no agreement between the clinical assessment and change in cup-disc ratio (κ = −0.07). Mean changes in cup-disc ratios among eyes that progressed were 0.002, −0.007, and 0.043 for the AGIS, CIGTS, and EMGT, respectively. The mean change among eyes considered to have progressed by clinical assessment was 0.014.

The 3 methods of classifying visual field progression are based on different principles. The AGIS uses the absolute deviations from age-matched control subjects to score the severity of the field, whereas the CIGTS uses P values rather than absolute differences and smooths the score by averaging P values surrounding each location. This should lead to lower intraperson variability of the scores, but this was not the case. However, in this analysis, the apparent rate of progression was higher in the CIGTS than in the AGIS, while producing a lower rate of improvement, which would suggest that the misclassification is lower with this method. The EMGT method uses P values to compare individual locations to see if they have deteriorated and requires deterioration in the same locations on at least 3 occasions. This seems strict, but it produced the highest rate of progression of the 3 methods. If the rule of any 3 locations deteriorating were replaced by 3 contiguous locations, the rate would have been 12% of patients progressing (12% of eyes), so a contiguity requirement would halve the incidence of pro-

Table 2. Agreement Between Statistical Methods and Adjudicated Clinical Assessment in the Detection of Visual Field Progression in 67 Eyes

<table>
<thead>
<tr>
<th>Visual Field Progression</th>
<th>Clinical Progression</th>
<th>κ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGIS</td>
<td>No</td>
<td>0.39 (0.17-0.61)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.49 (0.25-0.73)</td>
</tr>
<tr>
<td>CIGTS</td>
<td>No</td>
<td>0.45 (0.25-0.73)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.37 (0.13-0.61)</td>
</tr>
<tr>
<td>EMGT</td>
<td>No</td>
<td>0.39 (0.17-0.61)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.49 (0.25-0.73)</td>
</tr>
</tbody>
</table>

*AGIS indicates Advanced Glaucoma Intervention Study; CIGTS, Collaborative Initial Glaucoma Treatment Study; EMGT, Early Manifest Glaucoma Treatment study; and CI, confidence interval.

Table 3. Association Between Visual Field Progression and Increase in Vertical Cup-Disc Ratio of 0.1 or More During Study Follow-up in 65 Eyes

<table>
<thead>
<tr>
<th>Visual Field Progression</th>
<th>Increase in Cup-Disc Ratio</th>
<th>κ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGIS</td>
<td>No</td>
<td>0.07 (−0.16 to 0.30)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>−0.06 (−0.31 to 0.19)</td>
</tr>
<tr>
<td>CIGTS</td>
<td>No</td>
<td>0.30 (0.06 to 0.54)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.30 (0.06 to 0.54)</td>
</tr>
</tbody>
</table>

*Abbreviations are explained in the footnote to Table 2.
progression, making this rate comparable to that of the AGIS method.

The advantage of these comparisons is that the methods are applied to the same patient mix (distribution of disease severity). This eliminates the problem of not knowing whether differing progression rates across trials are caused more by disease severity in the population than by the analytic method of estimating progression. The disadvantage is that differences in eligibility criteria can only be taken into account by removing different patients from the full data set, reducing the advantage of using the common set of patients to compare methods. However, an analysis in which visual field eligibility criteria were used produced comparable progression rates to those found using the full data set. This study did not attempt to estimate what the incidence of field loss might be in the individual trials because the patient mix may be different because the trials have other entry criteria that are unrelated to visual field characteristics. For example, the AGIS study has a 5-year incidence rate that is somewhat higher than what we estimated in our patient population.

Another aspect of this analysis that does not directly mimic the trials is that the field tests used here were spaced 1 year apart, and because there were no “confirming” field tests, the next annual field test had to serve that purpose. This should not affect the comparison of progression rates across study methods. However, the most likely effect this would have would be to increase the time to progression for each of the 3 methods equally because they all require 2 confirming field tests after initial identification of progression. Hence, the incidence of and time to progression seen in this study may be an underestimate of what the trials would observe with the same patient mix in the same time period.

The rate at which patients “converted” was 2% for the AGIS, 4% for the CIGTS, and 4% for the EMGT per year. If one uses an increase in AGIS score of 3 or more (as used by the CIGTS), the rate of progression would be 3% per year, comparable to the CIGTS rate, with 69% agreement. This implies that a major difference between these 2 methods is the increase in score required to define progression. Using the same data set, progression based on statistically significant linear regression of any 1 of the corrected pattern SD, Glaucoma Hemifield Test sectors, or individual locations resulted in 33% of patients classified as having visual field deterioration over 6 years.10

If one assumes that improvement in the visual field did not occur over an average of 6 years, one can use the improvement rates as a measure of misclassification of the method if these are examined relative to the “deterioration” rates. Although it is possible to have a learning effect that might explain the improvement, this is less likely with these data because eligibility was based on 2 abnormal visual field test results, and this has been shown to produce subsequent field tests with low rates of improvement over time.27 The AGIS method has relatively low rates of progression and improvement, whereas the EMGT method had higher rates of progression and improvement. The CIGTS method has a rate of progression similar to that of EMGT, but no eyes were classified as improving by the CIGTS method. This implies that the EMGT method is likely more variable than the AGIS and CIGTS methods with regard to classification of progression. With the EMGT method, eyes can be classified as having both deterioration and improvement in different areas of the field. This was the case with 2 of 6 eyes classified as having improved by the EMGT method.

The earlier time to progression and the higher rate of improvement seen with the EMGT method implies that this method is detecting smaller changes earlier than the other 2 approaches. Given the nature of the EMGT, in which patients with mild visual field damage are randomized to treatment or no treatment, a method that is more likely to detect small changes (even if it is at the expense of some misclassification) is important in ensuring that patients who truly do progress are identified and treated in a timely manner.

In the absence of a criterion standard, it is difficult to assess which of these methods is “better” in terms of classifying progression. Werner et al showed that there was poor agreement between clinicians and that agreement between different statistical analyses was much better than between clinicians. We found excellent agreement between clinicians (trained in glaucoma treatment at different institutions), but only fair to moderate agreement between the statistical methods and the clinical assessment, with no particular method correlating better with the clinical assessment than the other methods. The EMGT method of identifying visual field progression was in fair agreement with increases in vertical cup-disc ratios, whereas there was no agreement with the other 2 methods. If one uses the clinical assessment as the criterion standard, then sensitivity was 36%, 50%, and 57% and specificity was 96%, 91%, and 87% for the AGIS, EMGT, and CIGTS, respectively. These findings make it difficult to identify one method as having clear validity over the others. These data suggest that other measures of progression of glaucomatous optic nerve damage in clinical trials will be important in evaluating the trial results.

Accepted for publication April 23, 1999.

Supported by grants EY11592, EY09130, EY03605, and RR04060 from the National Eye Institute, National Institutes of Health, Bethesda, Md.

Corresponding author: Joanne Katz, ScD, Johns Hopkins School of Hygiene and Public Health, Room W5513, 615 N Wolfe St, Baltimore, MD 21205-2103.

REFERENCES


A look at the past . . .

Dr. A. Jacobi presented a case of amaurotic family idiocy (Sachs) with symmetrical changes at the macula. The girl is at present one year old, the fourth child of healthy, not neurotic, not syphilitic, and not alcoholic parents, who emigrated from Russia, both belonging to the Semitic race. The other children are in good health. . . .

In discussing Dr. Jacobi’s case, Dr. B. Sachs said that it was interesting to see the way observations of the ophthalmologists (among whom Dr. Knapp was one of the first) and those of the neurologists had finally been fused to constitute one common type of disease. The neurological aspect of amaurotic family idiocy was quite as distinct as the ophthalmological aspect is said to be. Idiocy, blindness, and a uniform progress to a fatal issue are the symptoms of this new clinical condition.