Comparison of Methods to Predict Visual Field Progression in Glaucoma

Kouros Nouri-Mahdavi, MD, MSc; Douglas Hoffman, BA; Monica Ralli, MD; Joseph Caprioli, MD

Objective: To compare performance of pointwise linear regression, Glaucoma Change Probability Analysis (GCPA), and the Advanced Glaucoma Intervention Study (AGIS) method in predicting visual field progression in glaucoma.

Design: Longitudinal visual field data from AGIS. Proportion of progressing eyes and time to progression were the main outcome measures. One hundred fifty-six patients with 8 or more years of follow-up were included. Prediction of outcomes at 8 years was used to evaluate the performance of each method (pointwise linear regression, GCPA, and AGIS).

Results: Visual field progression at 8 years was detected in 35%, 31%, and 22% of patients by pointwise linear regression, GCPA, and the AGIS method, respectively. Baseline mean deviation was not different for nonprogressing vs progressing eyes for all methods ($P > .05$). Pointwise linear regression and GCPA had the highest pairwise concordance ($\kappa = 0.58$ [SD, 0.07]). The false prediction rates at 4 and 8 years varied between 1% and 3%. Glaucoma Change Probability Analysis predicted final outcomes better than pointwise linear regression at 4 years ($P = .001$).

Conclusions: All algorithms had low false prediction rates. Glaucoma Change Probability Analysis predicted outcomes better than pointwise linear regression early during follow-up. Algorithms did not perform differently as a function of baseline damage. Pointwise linear regression and GCPA did not agree well regarding spatial distribution of worsening test locations.

Arch Ophthalmol. 2007;125(9):1176-1181

The major problems in the detection of visual field worsening are long-term fluctuation\textsuperscript{1-3} and lack of a gold standard for validating any given technique. Methods that have been used for detection of visual field progression include Glaucoma Change Probability Analysis (GCPA),\textsuperscript{4,6} univariate linear regression analysis on individual test locations (pointwise),\textsuperscript{7-12} clusters of test locations\textsuperscript{11,13} or visual field indices,\textsuperscript{14} multivariate regression analyses,\textsuperscript{11,15} and arbitrary criteria designed for specific clinical trials, such as the scoring criteria for the Advanced Glaucoma Intervention Study (AGIS)\textsuperscript{16} and the Collaborative Initial Glaucoma Treatment Study.\textsuperscript{17} However, a review of the literature demonstrates that even pairwise agreement between methods is fair at best.\textsuperscript{18,19}

Given the lack of an external standard, evaluating methods by comparing their ability to predict visual field outcomes may be helpful. In this way, the visual field outcome defined with the same method at a subsequent point in time is used as a surrogate for validation. This concept is based on the assumption that if a technique performs poorly when data gathered during early follow-up are used to predict future outcomes by the same criteria, it may not be an appropriate technique for detection of progression. Similarly, the time to detection of the first solid evidence of progression is also important.

The specific aims of this study are (1) to compare the performance of 3 algorithms—pointwise linear regression analysis, GCPA, and the AGIS scoring system—for detection of visual field progression when the status at 8 years is used as the standard for comparison and (2) to compare the performance or predictive power of each algorithm for forecasting the visual field status at 8 years from data gathered during the first 4 years.

METHODS

The AGIS design and methods have been described in detail elsewhere.\textsuperscript{20} The AGIS protocol adhered to the tenets of the Declaration of Helsinki and was approved by all participating institutional review boards. All patients provided informed consent to enroll in AGIS. The institutional review board at the University of California, Los Angeles, also approved our study.
Visual field tests were conducted with a Humphrey Visual Field Analyzer I (Humphrey Instruments Inc, San Leandro, California) with the central 24-2 threshold test, size III white stimulus, full-threshold strategy, and the foveal threshold test turned on. Visual field defects were scored according to a system specifically designed for the study, which assigned a defect score ranging from 0 (no defect) to 20 (end-stage damage). Study measurements were made at baseline, at 3 months after initial intervention, and at each 6-month follow-up examination. Patients were included if they had 8 or more years of follow-up, an AGIS baseline visual field score of 16 or lower, and 6 or more reliable visual fields (AGIS reliability score ≤ 2) during the first 4 years of follow-up and 6 or more additional reliable visual fields between 4 and 8 years of follow-up. A minimum of 6 visual field examinations during each half of the follow-up were required so that an adequate number of visual fields would be available for applying pointwise linear regression at 4 and 8 years. Data after 8 years were truncated to achieve similar follow-up for all eyes. If both eyes of a patient were eligible, 1 eye was randomly selected. The visual field status at 8 years was evaluated with pointwise linear regression, GCPA, and the AGIS criteria for progression. For each technique, the outcome at 8-year follow-up was considered to be the reference. The performance of each method at each visit after baseline was compared with the reference.

POINTWISE LINEAR REGRESSION ANALYSIS

SPSS statistical software, version 11.5 (SPSS Inc, Chicago, Illinois), was used to perform pointwise linear regression. Sequential pointwise linear regression was performed beginning with 4 available visual fields, because regression analysis is not likely to detect a trend when fewer data are available. Estimated regression slopes were considered significant if less than or equal to −1.0 dB per year with a P ≤ .01 for testing the hypothesis (regression slope = 0).

For evaluation of visual field series during follow-up and at 8 years, we used criteria similar to those of GCPA so a comparison of the 2 techniques would be more meaningful: presence of 3 or more test locations, not necessarily contiguous, that demonstrated worsening, confirmed on at least 3 consecutive visual fields. When progression was observed before the end of follow-up, consistent evidence of worsening through the end of follow-up was required. Time to visual field progression was defined as the time from initial visual field to first evidence of visual field progression. The number and location of worsening points were noted.

GLAUCOMA CHANGE PROBABILITY ANALYSIS

All original AGIS visual fields were imported into a Humphrey Field Analyzer II, model 750 (Humphrey Instruments Inc). Visual fields from selected series of eyes were printed in GCPA format. Unreliable visual fields were excluded. The baseline for GCPA consisted of the preintervention visual field and the first examination performed within the 6 months after initial surgery. A progressing test location was defined as a location flagged by a black triangle (equivalent to P < .05) on the GCPA printout. The following criteria defined visual field progression with GCPA at 8 years (adopted from the Early Manifest Glaucoma Trial13): presence of 3 or more test locations, not necessarily contiguous, that demonstrated worsening on the last 3 examinations. The time to first detection of visual field progression was defined as the first visit at which at least 3 test locations progressing consistently until the end of follow-up were detected. The number and location of worsening points were noted.

AGIS CRITERIA

Customized software calculated AGIS scores for individual visual fields. Progression at 8 years, according to the AGIS scoring system, was defined as an increment of 4 or more points greater than the baseline value on the last 3 examinations. Time to visual field progression was defined as the first visit at which consistent worsening of the visual field through the end of follow-up (a minimum of 3 visual fields) was detected.

COMPARISON OF METHODS AT 4 AND 8 YEARS

For comparison of methods at 4 and 8 years, the same criteria and a set of less stringent criteria were applied to data available at 4 years. Less stringent criteria consisted of evidence of progression only on the last 2 examinations, including the 4-year examination for all 3 methods (1 confirmation). Four-year outcomes were compared with 8-year outcomes for each algorithm and prediction rates were calculated.

STATISTICAL METHODS

Statistics were used to estimate agreement of the methods at 8 years. We used the paired t test, repeated-measures analysis of variance, or the Mann-Whitney test for comparing numerical data, as required. Statistical significance was P < .05. Multiple comparisons performed would potentially increase chances of a type I error, but we felt this would be acceptable given the exploratory nature of this study.

RESULTS

A total of 156 eyes from 156 patients were included (Table 1). The median number of available visual fields at 4 and 8 years was 10 and 18, respectively. At baseline, the mean AGIS visual field score was 7.7 (range, 0 to 16) and the average (SD) mean deviation was −10.4 dB (5.6). Sixty-seven eyes (43%) underwent cataract surgery during the 8 years of follow-up.

Visual field progression at 8 years was detected in 54 (35%), 49 (31%), and 34 (22%) eyes according to pointwise linear regression, GCPA, and the AGIS technique, respectively. Pointwise linear regression analysis and GCPA detected significantly more progressing eyes compared with the AGIS method (McNemar test, pointwise linear regression vs AGIS, P = .001; GCPA vs AGIS, P = .02), while no difference was observed between pointwise linear regression and GCPA (McNemar test, P = .46). All 3 techniques agreed on progression in 22 of 156 eyes (14%) at 8 years (Figure 1). The highest pairwise agreement at 8 years was observed between pointwise linear regression and GCPA with κ = 0.58 (SD, 0.07). Pairwise agreements of the AGIS method with pointwise linear regression and GCPA was 0.47 (SD, 0.07) and 0.40 (SD, 0.08), respectively.

Figure 2 demonstrates the cumulative proportion of progressing eyes according to the 3 techniques. Median time to first detection of progression according to pointwise linear regression, GCPA, and the AGIS technique was 5.1 (range, 1.5-7.4), 4.7 (range, 1.0-7.4), and 3.8 years (range, 0.6-7.6), respectively. When considering eyes that were identified as progressing by all methods (22 eyes), there was no difference in the time to first detection of progression (repeated-measures analysis of variance, df = 2,

©2007 American Medical Association. All rights reserved.
Mean (SD) time to first evidence of progression in this subgroup was 4.3 (1.2), 3.9 (1.5), and 4.0 (2.3) years for pointwise linear regression, GCPA, and the AGIS technique, respectively.

We also investigated whether baseline severity of glaucoma had any influence on the performance of the methods. Average (SD) mean deviation in nonprogressing eyes was −9.8 dB (6.8), −10.4 dB (6.8), and −9.9 dB (6.6) according to pointwise linear regression, GCPA, and the AGIS technique, respectively, and −10.3 dB (5.6), −9.1 dB (5.3), and −10.1 dB (5.4) in the worsening eyes. The mean deviation was not different between nonprogressing and progressing eyes for all methods (unpaired t test, P = .7). Proportions of progressing eyes were compared according to baseline glaucoma severity and were not different (P > .05). We compared the mean deviation in eyes that progressed by all methods and in eyes that were not detected by any of the 3 methods. The average mean deviation in these 2 groups was −9.1 dB and −10.6 dB, respectively (Mann-Whitney test, P = .27).

### COMPARISON OF POINTWISE LINEAR REGRESSION AND GCPA

Pointwise linear regression analysis and GCPA showed similar results on visual field progression in 37 of 156 eyes (24%). The median number of progressing test locations at the time of first detection was similar for both pointwise linear regression and GCPA (4 test locations) with a range of 3 to 20 for pointwise linear regression and 3 to 22 for GCPA. However, only 58 progressing test locations (33% of 175 with pointwise linear regression and 35% of 165 with GCPA) were simultaneously detected by both methods. For eyes detected by both methods, the mean (SD) time to first confirmed progression was 4.7 (1.3) and 4.6 (1.6) years for pointwise linear regression and GCPA, respectively (paired t test, P = .81).

The number of progressing test locations according to pointwise linear regression was compared in eyes that were detected by all methods (22 eyes) vs eyes that progressed by pointwise linear regression only (32 eyes). The median number of progressing test locations was 4 (range, 3-20) and 3 (range, 3-10) in the 2 groups, respectively (Mann-Whitney test, P = .30). When the number of test locations was compared in eyes that were detected by all methods (22 eyes) vs eyes that progressed by GCPA only

### Table 1. Study Sample Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. of patient eyes</td>
<td>156</td>
</tr>
<tr>
<td>No. of eyes</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>76 (48.7)</td>
</tr>
<tr>
<td>Left</td>
<td>80 (51.3)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>78 (50)</td>
</tr>
<tr>
<td>F</td>
<td>78 (50)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>79 (50.6)</td>
</tr>
<tr>
<td>White</td>
<td>76 (48.7)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Age at baseline, y</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>64.4 (9.4)</td>
</tr>
<tr>
<td>Range</td>
<td>41 to 81</td>
</tr>
<tr>
<td>Intervention sequence</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>79 (50.6)</td>
</tr>
<tr>
<td>TAT</td>
<td>77 (49.4)</td>
</tr>
<tr>
<td>Cataract surgery during follow-up</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>89 (57.1)</td>
</tr>
<tr>
<td>Yes</td>
<td>67 (42.9)</td>
</tr>
<tr>
<td>Median No. of visual field examinations per eye (range)</td>
<td></td>
</tr>
<tr>
<td>At 4 y</td>
<td>10 (6 to 11)</td>
</tr>
<tr>
<td>At 8 y</td>
<td>18 (13 to 19)</td>
</tr>
<tr>
<td>AGIS visual field score at baseline</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>7.7</td>
</tr>
<tr>
<td>SD</td>
<td>4.5</td>
</tr>
<tr>
<td>Range</td>
<td>0 to 16</td>
</tr>
<tr>
<td>Mean deviation at baseline (dB)</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>−10.4</td>
</tr>
<tr>
<td>SD</td>
<td>5.6</td>
</tr>
<tr>
<td>Range</td>
<td>−23.7 to 1.6</td>
</tr>
</tbody>
</table>

Abbreviations: AGIS, Advanced Glaucoma Intervention Study; ATT, argon laser trabeculoplasty, trabeculectomy, trabeculectomy; TAT, trabeculectomy, argon laser trabeculoplasty, trabeculectomy.

*Values are number (percentage) unless otherwise indicated.*

---

**Figure 1.** Venn diagram showing agreement of the pointwise linear regression analysis, Glaucoma Change Probability Analysis, and the Advanced Glaucoma Intervention Study criteria to detect progression at 8 years.

**Figure 2.** The cumulative proportion of eyes progressing according to the pointwise linear regression (PLR) analysis, Glaucoma Change Probability Analysis (GCPA), and Advanced Glaucoma Intervention Study (AGIS) criteria.
(27 eyes), the difference was statistically significant (Mann-Whitney test, \( P = .03 \)). The median number of progressing test locations was 4 (range, 3-22) and 3 (range, 3-6), respectively, in the 2 groups.

**COMPARISON OF METHODS AT 4 AND 8 YEARS**

We compared the performance of the 3 methods at 4 and 8 years to estimate the false prediction rates for detection of visual field progression (Table 2). The false prediction rate varied from 1% (pointwise linear regression) to 3% (GCPA) if 2 confirmations were required at 4 years for definition of progression, while it ranged from 4% to 11% if only 1 confirmation was required. Glaucoma Change Probability Analysis predicted 8-year outcomes significantly better at 4 years than pointwise linear regression (\( \chi^2 \) test, \( P = .001 \), for both 1 and 2 confirmations).

To explore the potential causes for the discrepancy of the pointwise linear regression results at 4 vs 8 years, we looked at the scatterplots of the threshold sensitivity at test locations for all 4 eyes demonstrating change at 4 years but not at 8 years (Figure 3). All 4 eyes had cataract surgery during follow-up. In addition to long-term fluctuation and cataract removal, possible causes of disagreement between pointwise linear regression results at 4 and 8 years included a variety of factors, such as fluctuation in threshold sensitivity, presence of cataract, and the use of pointwise linear regression analysis.

### Table 2. Performance of 3 Methods at 4 Years Compared With Visual Field Outcomes at 8 Years

<table>
<thead>
<tr>
<th>Method</th>
<th>8-y Results, %</th>
<th>2 Confirmations, %</th>
<th>1 Confirmation, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Progressing</td>
<td>Nonprogressing</td>
<td>Sensitivity (SE)</td>
</tr>
<tr>
<td>PLR</td>
<td>54</td>
<td>102</td>
<td>19 (5)</td>
</tr>
<tr>
<td>GCPA</td>
<td>49</td>
<td>107</td>
<td>45 (7)</td>
</tr>
<tr>
<td>AGIS</td>
<td>34</td>
<td>122</td>
<td>35 (8)</td>
</tr>
</tbody>
</table>

**Abbreviations:** AGIS, Advanced Glaucoma Intervention Study; GCPA, Glaucoma Change Probability Analysis; PLR, pointwise linear regression analysis.

*For pairwise comparison of methods with 2 confirmations: PLR vs GCPA, \( P = .001 \); AGIS vs PLR, \( P = .08 \); AGIS vs GCPA, \( P = .22 \). For pairwise comparison of methods with 1 confirmation: PLR vs GCPA, \( P = .001 \); AGIS vs PLR, \( P = .14 \); AGIS vs GCPA, \( P = .14 \). (All based on \( \chi^2 \)).

**Figure 3.** Scatterplot examples of individual test location belonging to eyes that progressed at 4 years according to pointwise linear regression analysis but did not show progression at 8 years. A, Early progression during the first 4 years followed by stability between 4 and 8 years apparently caused by cataract extraction. B, Although one of the criteria for progression (decline in threshold sensitivity \( \geq 1 \text{ dB per year} \)) is met at 8 years, progression is not detected at 8 years because the high long-term fluctuation results in a \( P \) value \( > .01 \). A similar scenario could happen if the threshold starts at \( \leq 10 \text{ dB} \) and declines to 0 during early follow-up. C, Early apparent progression due to falsely high threshold sensitivity at baseline (arrowhead). Pointwise linear regression is highly influenced by outlying points at either end of follow-up. D, Reversal of progression due to higher fluctuation along with fair stability of the threshold sensitivities during the latter part of follow-up. Arrows indicate time of cataract extraction.
method is conservative.12,13,18,19,22 An interesting finding progression and possibly a longer time to progression. led to a more conservative estimate of the rate of pro-
vision from prior studies in that we required evidence of
overcalls progression or identifies the wrong eyes as pro-
tion rate during early follow-up is important. If a method
each technique at the final follow-up. The false predic-
visual field progression was defined by the findings of
consensus of clinicians or simulation of visual field data has been previously reported.19,22 We compared point-
wise linear regression, GCPA, and AGIS criteria with a
different approach in this investigation. The standard for
field progression was defined by the findings of
each technique at the final follow-up. The false predic-
rate during early follow-up is important. If a method
overcalls progression or identifies the wrong eyes as pro-
gressing early during follow-up, it may not be a desir-
able method for clinical use. The current study is differ-
ent from prior studies in that we required evidence of
progression to be consistently present from the time of
detection through the end of follow-up. This may have
led to a more conservative estimate of the rate of pro-
progression and possibly a longer time to progression.
Our findings confirm previous reports that the AGIS
method was conservative.12,13,18,19,22 An interesting finding of our study was that the conservative nature of AGIS
criteria was evident only after about 4 to 5 years of follow-
up. Also, the mean time for the first detection of visual
field progression was similar in the subset of eyes that progressed with all 3 techniques. This is contrary to the
findings of Katz and colleagues.22 However, there were
only 5 eyes in their report on which all the methods
agreed. A caveat of our study is the censoring of a small
number of visual fields owing to unreliability. Such cen-
soring could lead to a systematic bias in measuring the
time to the first visual field worsening; ie, the time to the
first confirmed progression could be overestimated. How-
ever, this is not an issue when various methods are com-
pared with each other within the same database.
We found that the detection rate of visual field pro-
gression was linear with all 3 methods (Figure 2). There
was no evidence of a late sudden surge of the progres-
sion rate. This indicates that we were not misdiagnos-
ing progressing eyes early on, only to detect them to-
ward the end of follow-up. The different performance of
the methods could be a function of differential field ef-
effects, since the average mean deviation in our study sample was −10.4 dB, and therefore some eyes had advanced glau-
coma. The gradual linear increase in the proportion of
progressing eyes over time for all 3 methods is also not
consistent with a significant late floor effect.
The results of this investigation may help us deter-
mine which algorithm would more accurately detect glau-
coma progression. Any single method might be more ap-
propriate at different times during follow-up. For example,
the AGIS method, despite being conservative in the long
run, seemed to detect more progression than pointwise
linear regression and GCPA during the first few years of
follow-up. This is especially true in comparison with
pointwise linear regression. Regression analyses are quite
sensitive to the number of available data points and would
not be expected to perform well during early follow-up
of patients with glaucoma when only a few visual fields
are available. In such situations, GCPA or the AGIS
method seem more appropriate. In contrast to the AGIS
criteria, GCPA is sensitive to long-term fluctuation, be-
cause the information derived from the remainder of the
field series since each follow-up visual field is compared
only with the baseline fields. Hence, the need for con-
firmation of progression is greatest with GCPA. Con-
versely, as pointwise linear regression uses the informa-
tion derived from all the available visual fields, it is less
prone to false positives, and when a large number of vi-
sual fields are available, it is highly specific as recently
reported by Vesti et al.19
The requirement of 1 or 2 confirmations for defining
visual field progression has been used to reduce the false-
positive rate of detection.4,6,16,17 It has been shown that
as the number of confirmations increases, the number of visual field series detected as progressing decreases pro-
portionally. We required 2 confirmations (evidence of
progression on 3 consecutive examinations) to keep the
false prediction rate as low as possible. Because of the
design of our study, we cannot measure true specificity.
However, using the 8-year outcomes as a standard for
the performance of each of the methods at 4 years, we found
that all 3 methods correctly predicted stability (97%-
99%) when_Corrections were required to detect pro-
gression at 4 years. Even when one confirmation was re-
quired, the false prediction rate remained acceptable
(<11%). This has clinical implications, because multiple
confirmations are not always practical in routine
clinical care.
We were interested in determining whether point-
wise linear regression and GCPA agreed on the spatial
distribution of worsening test locations. We used simi-
lar criteria for the definition of visual field progression
with pointwise linear regression and GCPA to make them
as comparable as possible.4,6 However, the 2 techniques
agreed on only about one-third of the test locations found
to be progressing by each method separately. The 2 tech-
niques use inherently different approaches. Pointwise lin-
ear regression is a trend analysis, whereas GCPA uses an
event analysis approach for detection of progressing test
location.23 Therefore, pointwise linear regression may be
better suited to detect linear trends of progression, while
GCPA might be potentially more sensitive for detection
of episodic progression.
We evaluated and compared the performance of 3 com-
monly_used algorithms for the detection of visual field
progression. The more conservative nature of the AGIS
technique appeared after a few years of follow-up. Point-
wise linear regression and GCPA had the highest pair-
wise concordance but did not agree well on the spatial
distribution of test locations, demonstrating evidence of
progression. Pointwise linear regression was the least sen-
sitive technique during early follow-up of patients with glaucoma. Therefore, either GCPA or AGIS algorithms may be preferable in this setting, though they have a slightly higher false prediction rate. We found no evidence that any of the 3 techniques performed differently as a function of baseline severity of glaucomatous damage.

Submitted for Publication: September 5, 2006; final revision received January 25, 2007; accepted February 19, 2007.

Correspondence: Joseph Caprioli, MD, Glaucoma Division, Jules Stein Eye Institute, 100 Stein Plaza, Los Angeles, CA 90095 (caprioli@jsei.ucla.edu).

Financial Disclosure: None reported.

Funding/Support: This study was supported by a Grant-in-Aid from Fight for Sight (Dr Nouri-Mahdavi), an unrestricted grant from Research to Prevent Blindness, and grant R01 EY12738 from the National Institutes of Health (Dr Caprioli).

Previous Presentation: This study was presented at the Annual Meeting of the American Glaucoma Society; March 3, 2005; Snowbird, Utah.

Additional Contributions: Elliot M. Landaw, MD, PhD, Chairman of the Department of Biomathematics, David Geffen School of Medicine at UCLA provided statistical and editorial assistance.

REFERENCES