

Comparison of Different Methods for Detecting Glaucomatous Visual Field Progression

Eija Vesti,^{1,2} Chris A. Johnson,¹ and Balwantray C. Chauhan³

PURPOSE. To compare the performance characteristics of seven methods for analyzing glaucomatous visual field progression, using a combination of real patient data and computer simulation techniques.

METHODS. The initial and final visual field results, separated by 7 years and measured with the full-threshold 30-2 program of the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA) of 76 patients with open-angle glaucoma were used. A computer simulation program generated 14 interim semiannual visual fields under conditions of high, moderate, and no variability. Progression was analyzed using the methods of the Advanced Glaucoma Intervention Study (AGIS), the Collaborative Initial Glaucoma Treatment Study (CIGTS), three criteria based on the Glaucoma Change Probability (GCP) analysis, and two criteria based on point-wise linear regression analysis (PLRA). Specificities were calculated by using the same visual field of each patient as both the initial and final field (no progression) under conditions of moderate and high variability.

RESULTS. Under the no-variability condition, progression rates were 18% for the AGIS, 36% for CIGTS, 47% to 62% for the three GCP methods, and 72% and 84% for the two PLRA methods. Progression rates increased with greater variability with the three GCP methods and decreased with all other methods. The time to detect confirmed progression was longest for the PLRA methods and shortest for the CIGTS and GCP methods. Under the moderate-variability condition, all methods yielded high specificity. The AGIS, CIGTS, and one of the GCP and PLRA methods were relatively resistant to high variability and maintained high specificities.

CONCLUSIONS. The AGIS and CIGTS methods had high specificity, but classified fewer cases of progression than the other methods. The GCP methods determined progression earliest; however, they were generally not as specific. Methods based on PLRA were specific but times to confirmed progression

were the longest. (*Invest Ophthalmol Vis Sci.* 2003;44:3873-3879) DOI:10.1167/iovs.02-1171

Detection of glaucomatous visual field loss and determination of subsequent visual field progression are among the most important aspects of glaucoma management. Progression is difficult to distinguish from variability, and a variety of statistical approaches have been described to assist in making such determinations.¹ Several of the clinical trials in glaucoma have used criteria for visual field progression²⁻⁶; however, these criteria are unique, making it difficult to compare the outcomes of these studies. Central to the issue of visual field progression is the lack of consensus on what constitutes a clinically significant change in the visual field.

The ability of a specific method to detect visual field progression is probably affected by the degree of the initial glaucomatous visual field loss and intratest and intertest threshold variation.⁷ Some methods may be sensitive for detecting progression of early losses but are less sensitive for detecting progression when the visual field loss is more advanced, and vice versa. There are some published reports that evaluate the performance of different methods for determining visual field progression using the same longitudinal data⁸⁻¹⁰; however, they did not investigate the range of methods currently available. Large and complete longitudinal sets of visual field data are usually difficult and time consuming to obtain. Computer simulation techniques^{11,12} offer an alternative method of acquiring and analyzing such data in addition to allowing control for the various levels of visual field damage and the different types and degrees of variability. The validation of this approach, using a combination of real and simulated data,¹³ and its use in other applications¹⁴ has been described previously.

The objective of this study was to compare the performance characteristics of seven different methods to detect glaucomatous visual field progression, under different conditions of variability. We compared the methods used by two large randomized clinical trials, the Advanced Glaucoma Intervention Study (AGIS)¹⁵ and the Collaborative Initial Glaucoma Treatment Study (CIGTS),⁵ three criteria based on the Glaucoma Change Probability (GCP) analysis available with the Statpac program of the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA),¹⁶ and two criteria based on point-wise linear regression analysis (PLRA).¹⁷⁻¹⁹ For each of the methods we determined the frequency, time to confirmed progression, and specificity, under different conditions of variability.

MATERIAL AND METHODS

Visual field data were selected from patients with open-angle glaucoma who were observed for 7 years or longer, after any presumed learning effects were accounted for, and who had at least eight or more visual field examinations with the full-threshold 30-2 program of the Humphrey Field Analyzer (Carl Zeiss Meditec). Because the data were obtained as part of a retrospective chart review with all identification material removed, the study was granted a Human Subjects Informed Consent exemption from the Legacy Institutional Review Board. The protection of the patients' privacy was in compliance with the Declaration of Helsinki.

From ¹Discoveries in Sight, Devers Eye Institute, Portland, Oregon; the ²Helsinki University Eye Hospital, Helsinki, Finland; and the ³Department of Ophthalmology, Dalhousie University, Halifax, Nova Scotia, Canada.

Supported in part by a grant from the Glaucoma Research Foundation (CAJ, BCC); Grant EY03424 from the National Eye Institute, Bethesda, Maryland (CAJ); Grant MT-11357 from the Canadian Institutes of Health Research (BCC); and grants from the Eye Foundation (Silmäsäätiö), Finland (EV); the Eye and Tissue Bank Foundation (Silmä- ja Kudospankkisäätiö); Finland (EV); and the Finnish Cultural Foundation, Finland (EV).

Submitted for publication November 15, 2002; revised December 18, 2002; accepted January 14, 2003.

Disclosure: E. Vesti, None; C.A. Johnson, None; B.C. Chauhan, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Balwantray C. Chauhan, Department of Ophthalmology, Dalhousie University, 2nd Floor Centennial Building, Queen Elizabeth II Health Sciences Centre, Halifax, N.S., Canada B3H 2Y9; bal@dal.ca.

For each patient, only two fields, 7 years apart, were selected from one randomly selected eye and used as input for the computer simulations as the initial and final visual fields, respectively. A previously described visual field simulation program¹¹ was used to generate the interim visual fields at 6-month intervals, and only these simulated fields were analyzed in this study. In the simulation program, both short-term (SF) and long-term (LF) threshold fluctuations can be specified. For this study, a simple linear interpolation of the point-wise thresholds based on the actual values in the initial and final fields were first computed for the interim 6-month visual fields. The simulated threshold values were then adjusted for various degrees of SF and LF, both of which are in the form of Gaussian distributions of underlying probability density functions, the widths of which were varied according to the level of fluctuation chosen. A random number generator was used to sample a threshold value from the probability density function described by SF and LF. This procedure was performed for all test locations. SF was also weighted according to the deviation of a test point threshold from the normative value (defect-related variability). The SD of the threshold distribution was increased by 0.08 dB/dB of deviation from the age-matched threshold. The LF component was homogeneous across all locations in the visual field, irrespective of threshold. Full details of the simulation procedures have been described elsewhere.¹¹

In this study, three sets of 14 visual fields at 6-month intervals were simulated for each patient for the follow-up. In the first set, 0 dB of SF and 0 dB of LF were used to describe progression in the ideal condition of no variability. In the second set, we used a moderate degree of variability (1 dB of SF, in addition to the defect-related variability, and 1 dB of LF). In the third set, we used a high degree of variability (2 dB of SF, in addition to the defect-related variability, and 1 dB of LF). Specificities of the seven methods were evaluated in another two sets of visual fields that were simulated using the moderate- and high-variability conditions. In these simulations, nonprogression was established by making the initial and final visual fields identical; hence, the thresholds were affected only by the two degrees of variability.

One part of the output of the simulation program is a total deviation map. For each test point, the deviation of a threshold is determined relative to age-corrected normative data. For each simulated field, the mean of these threshold deviations (mean deviation, MD) was calculated for program 24-2 test pattern, excluding one point above and below the blind spot. Because the AGIS and CIGTS methods are based on the program 24-2 test pattern of the Humphrey Field Analyzer, we analyzed these 52 locations for all seven methods.

The AGIS Scoring System

In the AGIS scoring system,¹⁵ the 24-2 area of the visual field is divided into three segments: the nasal area and the remaining superior and inferior hemifield areas. The amount of sensitivity loss (total deviation) necessary to be considered abnormal varies from 5 to 9 dB depending on the location. A nasal defect is considered to be a group of three or more contiguous depressed points that may cross the horizontal midline. A nasal step is 1 contiguous point or more in the superior or inferior nasal area without any depression in the opposite nasal area. Either of these defects results in a score of 1. If more than half (≥ 4 points) of the nasal test points have defect depths of 12 dB or more, the assigned score is 2. The superior and inferior hemifields are scored separately. The number of groups of three or more contiguous depressed points is identified and the total number of points within these groups is summed. A score of 1, 2, 3, or 4 is added respectively for a total of 3 to 5 points, 6 to 12 points, 13 to 20 points, and more than 20 points. Additional scores are given according to the defect depth. A score of 1, 2, 3, 4, or 5 is added respectively if at least half the points are depressed by ≥ 12 , ≥ 16 , ≥ 20 , ≥ 24 , or ≥ 28 decibels. A maximum score of 9 can be given to each hemifield, and a maximum score of 2 to the nasal area, resulting in a total score that can range from 0 (no field loss) to 20 (end stage). Progression is defined as an increase in score by 4 or more in three consecutive follow-up fields.²⁰ In this

study, changes in score were calculated for each follow-up field relative to the average score calculated from the initial two visual fields.

The CIGTS Scoring System

The CIGTS scoring system⁵ differs from the AGIS scoring system in that the probability levels of the total deviations, rather than the dB values themselves, of the 24-2 area are analyzed. A deviating test point is defined as one that has a total deviation with a probability value of 5% or less. If a deviating test point has two or more neighboring deviating test points, then a score from 1 to 4 can be assigned to it. The two most depressed points of all deviating neighboring points are identified, and the scoring is based on the three contiguous points (the deviating point and two most depressed neighboring points), whether they all have probability values of $\leq 5\%$ (score of 1), $\leq 2\%$ (score of 2), $\leq 1\%$ (score of 3), or $\leq 0.5\%$ (score of 4). These 3 deviating points have to be located within the same hemifield. Each of the 52 points is given a score (ranging from 0 to 4) and all the scores are summed (maximum score of 208). The sum is divided by 10.4 to scale the total score between 0 (no defect) and 20 (end stage). Progression is defined as an increase in score by 3 or more in three consecutive follow-up fields.⁵ The change in score at each follow-up field was calculated relative to the average score determined from the initial two visual fields.

The Glaucoma Change Probability Analyses

In the glaucoma change probability (GCP) analyses, 52 locations in the 24-2 visual field area were analyzed in a manner similar to that described by Heijl et al.¹⁶ For the GCP calculations, we used our own data set for test-retest variability stratified by defect depth, as reported in a previous publication.²¹ The change at each follow-up field was calculated relative to the average deviation at the same test point calculated from the simulated baseline and first follow-up fields. Locations with changes lower than the 5th percentile of the appropriate test-retest variability value were flagged with black triangles. We compared the performance of the GCP analysis from our software to that of the commercial Statpac program (Carl Zeiss Meditec) in sets of real visual field analyses from 17 randomly selected patients (who had a total of 98 examinations) whose visual field damage ranged from mild to severe. The total number of points with differences in deviations at the less than 5% level in the Statpac analyses was 629, whereas the corresponding number in our software was 556. The two methods agreed in the location of 453 points, and disagreement between the two techniques occurred mostly when the significance of the changes in total deviation were close to the 5% change probability cutoff. These results indicate good agreement between the commercial Statpac GCP analysis and the one derived from our dataset.

The three GCP methods analyzed were: (1) GCP (2×4): presence of four or more overlapping points at the less than 5% level (as indicated by black triangles in the GCP printout) occurring in two of three consecutive fields²²; (2) GCP (8, 2×4): presence of eight or more points at the less than the 5% level, of which four or more points had to be confirmed in one of the next two consecutive fields; and (3) GCP (3×4): presence of four or more overlapping points at the less than 5% level in three consecutive fields. For each of these analyses, the points did not have to be spatially contiguous.

Point-wise Linear Regression Analysis

Point-wise least-squares linear regression analysis (PLRA) was performed separately for each of the 52 test points of the 24-2 field area. The analysis was performed at each visit, starting from the third follow-up visit when the first four fields were included in the calculations. Then, the first five fields were included and so forth until at the last follow-up visit all 14 fields were included in the calculations to determine progression status at each test point. Test points with a significant regression slope ($P < 0.01$) showing 1.0 dB/y or more of sensitivity loss were identified for each follow-up visit. A point was considered to have confirmed progression if a significant slope was

TABLE 1. Criteria for Confirmed Visual Field Progression

Method	Criteria
AGIS	Increase in score by ≥ 4 in 3 consecutive follow-up fields.
CIGTS	Increase in score by ≥ 3 in 3 consecutive follow-up fields.
GCP (2 × 4)	Threshold deviation outside test-retest variability limits for the baseline threshold deviation in ≥ 4 overlapping points occurring in 2 of 3 consecutive fields.
GCP (8, 2 × 4)	Threshold deviation outside test-retest variability limits for the baseline threshold deviation in 8 or more test locations and confirmed in one of the two following consecutive fields with ≥ 4 overlapping points.
GCP (3 × 4)	Threshold deviation outside test-retest variability limits for the baseline threshold deviation in ≥ 4 overlapping points in 3 consecutive fields.
PLRA2	Significant ($P < 0.01$) regression slope of -1.0 dB/y or less in the same 2 test points in 3 of 4 consecutive fields.
PLRA3	Significant ($P < 0.01$) regression slope of -1.0 dB/y or less in the same 3 test points in 3 of 4 consecutive fields.

detected in the same point in three of four consecutive fields. The criteria for visual field progression were: (1) confirmed progression in two or more test points (referred to as the PLRA2 criterion), and (2) confirmed progression in three or more test points (referred to as PLRA3 criterion).

The specific methods analyzed in this study are summarized in Table 1. To ensure proper analysis of progression with the AGIS and CIGTS scoring systems, only eyes with baseline AGIS scores from 1 to 16 and baseline CIGTS scores from 1 to 17 were included. These inclusion criteria ensured that at baseline all fields had a glaucomatous visual field defect that was not too advanced for the AGIS and CIGTS methods to detect progression (necessity for increase in score by 4 and 3 points, respectively). The time to confirmed progression was determined for each of the seven methods.

Statistical Methods

The Kolmogorov-Smirnov test was used to test for normality of distributions. Pair-wise comparisons were performed using either a paired *t*-test or a Wilcoxon signed rank test with a Bonferroni correction to account for multiple statistical comparisons. Comparison of follow-up times to confirmed progression between different methods was performed by either one-way analysis of variance (ANOVA) with the Tukey test for post hoc multiple pair-wise comparisons or the Kruskal-Wallis one-way ANOVA on ranks with the Dunn test for post hoc multiple pair-wise comparisons. In the post hoc analyses, $P < 0.05$ was chosen for the significance level. Confidence intervals for the specific estimates were made by using exact binomial methods.²³

RESULTS

Seventy-six patients qualified for the study. The mean age \pm SD of the patients was 63.4 ± 11.1 years. The mean \pm SD AGIS and CIGTS scores of the real fields that were used as initial input fields in all the simulations were 7.6 ± 3.5 and 8.6 ± 4.1 , respectively. The mean \pm SD MD of these fields was -8.5 ± 4.7 dB with a range from -0.3 to -20.4 dB. The mean change in MD \pm SD from the first to the last (14th) simulated field was -2.3 ± 3.1 , -2.3 ± 3.2 , and -2.2 ± 3.0 dB, respectively, for the simulations with no, moderate, and high variability.

AGIS and CIGTS Methods

The AGIS method classified 14 (18%) of the simulated visual fields with no variability as progressing (Fig. 1). The number of progressing fields with moderate and high variability was 7 (9%) and 6 (8%), respectively (Fig. 1). The corresponding mean follow-up times \pm SD to confirmed progression under the three variability conditions were 5.1 ± 1.0 , 5.3 ± 1.1 , and 5.1 ± 1.7 years, respectively.

For corresponding variability conditions, the CIGTS method classified 2 to 3.5 times as many progressing fields as the AGIS method (Fig. 1). The mean follow-up times to confirmed progression with the CIGTS method were approximately 1 year shorter than the corresponding follow-up times with the AGIS method (Fig 1).

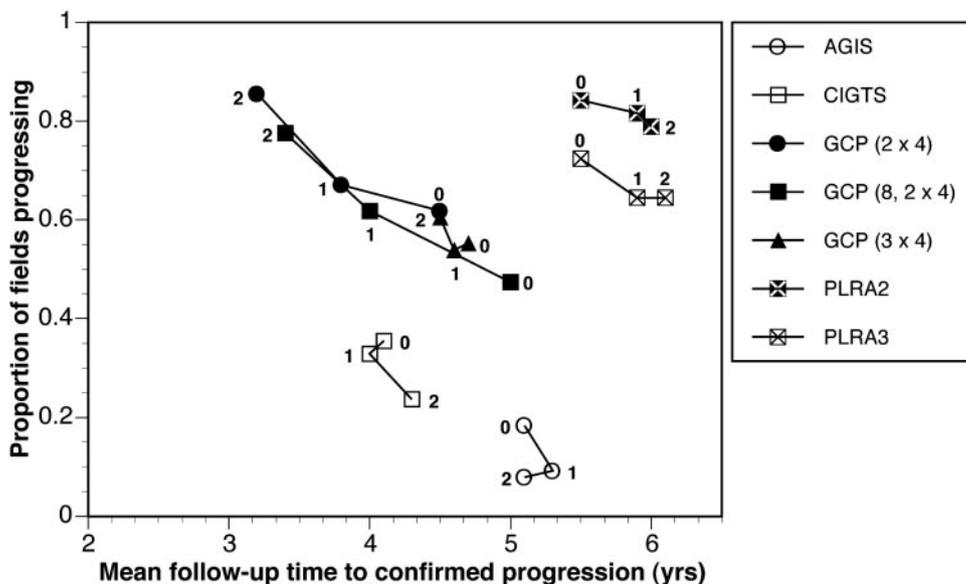


FIGURE 1. Proportion of progressing cases as a function of the time to confirmed progression for the seven methods using simulations with no threshold variability (0), moderate threshold variability (1), and high threshold variability (2).

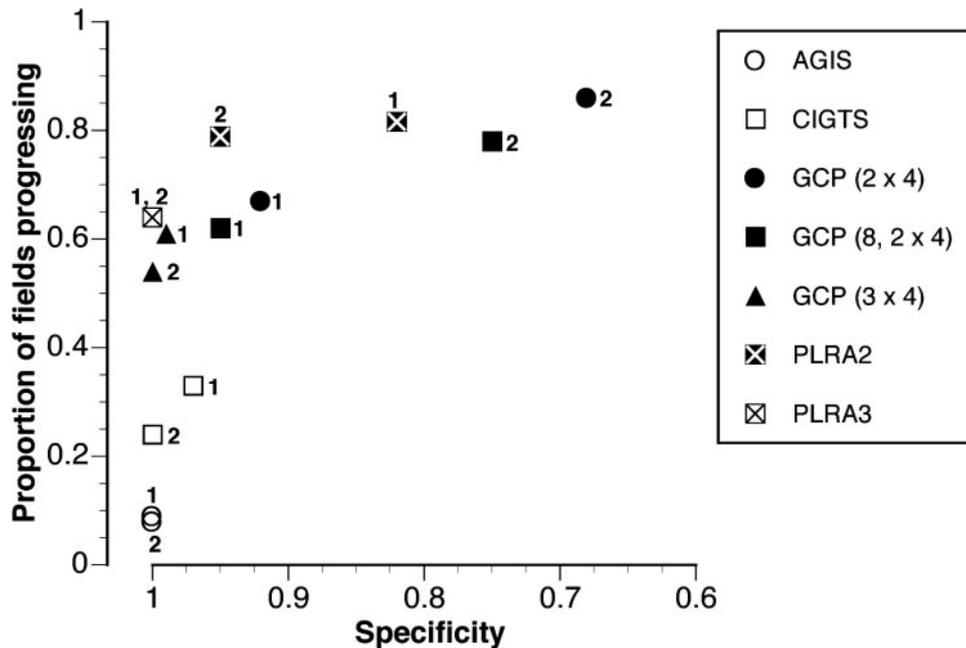


FIGURE 2. Proportion of progressing cases as a function of specificity for the seven methods using simulations with moderate threshold variability (1) and high threshold variability (2).

The CIGTS method identified 12 (86%) of the fields progressing with the AGIS method under the no-variability condition and all AGIS progressing fields under the moderate- and high-variability conditions. In both AGIS and CIGTS progressing fields, the CIGTS method detected progression significantly earlier than the AGIS method under the no-variability (by 1.5 ± 1.2 years, $P < 0.001$, paired *t*-test) and moderate-variability (by 2.3 ± 0.8 years, $P < 0.001$) conditions. The CIGTS method also detected progression earlier in the high-variability condition but the difference failed to reach statistical significance after Bonferroni correction (2.0 ± 1.5 years, $P = 0.02$, paired *t*-test).

GCP Methods

Under the no-variability condition, the GCP methods classified 36 (47%) to 47 (62%) fields as progressing, whereas the respective figures for the moderate- and high-variability conditions were 41 (54%) to 51 (67%) and 46 (61%) to 65 (86%), respectively (Fig. 1). For each of these conditions, the follow-up time to confirmed progression varied significantly between the GCP methods ($P < 0.001$, Kruskal-Wallis ANOVA, in the no-variability and high-variability conditions, and $P = 0.003$ in the moderate-variability condition). The Dunn post hoc test of the simulations with no variability showed that the follow-up time to confirmed progression with the GCP (8, 2×4) method was significantly longer than with the GCP (2×4) method (Fig. 1). Under the moderate- and high-variability conditions, the follow-up time to confirmed progression was longest for the GCP (3×4) method. Unlike the other two GCP methods, however, it appeared to be the most resistant to variability, both in terms of the number of progressing fields and time to confirmed progression.

Under all variability conditions, the GCP methods classified more fields as progressing than either the AGIS or CIGTS methods (Fig. 1). The mean follow-up time to confirmed progression was shorter with the GCP methods compared with the AGIS methods, but was more comparable with the CIGTS method. Contrary to the AGIS and CIGTS methods, the number of fields classified as progressing with the GCP methods increased with increasing variability.

PLRA Methods

In the simulations with no variability, the PLRA2 and PLRA3 methods respectively classified 64 (84%) and 55 (72%) fields as progressing, with the number decreasing with increasing variability (Fig. 1). The time to confirmed progression increased with increasing variability.

Under all variability conditions, the two PLRA methods classified more eyes as progressing than either the AGIS or CIGTS methods. They also classified more eyes as progressing than any of the GCP methods in the simulations with no variability, but with moderate or high variability, the number of progressing cases was more similar. Also under all variability conditions, the mean follow-up time to confirmed progression was longer for the PLRA methods than for any other method with neither of the PLRA methods able to confirm progression before a mean follow-up of 5.5 years (Fig. 1).

Specificity

Under the moderate-variability condition, all methods attained a specificity of 82% (CI: 71%–89%) or higher (Fig. 2). Under the high-variability condition, the specificities for only the GCP (2×4) and GCP (8, 2×4) methods decreased respectively to 68% (confidence interval [CI]: 57%–78%) and 75% (CI: 64%–83%), but remained high for the other methods (95% [CI: 87%–98%] or higher; Fig. 2).

Agreement between Progression Methods

The rates of agreement between the CIGTS, GCP (3×4), and PLRA3 methods under the three variability conditions was evaluated. These methods were chosen as the best (balance between the number of progressing cases and specificity) among the three classes of methods to evaluate progression.

Under the no-, moderate-, and high-variability conditions, all three methods agreed on no progression in 21 (27.6%), 25 (32.9%), and 21 (27.6%) patients, respectively, whereas under the same variability conditions, progression was detected by all three methods in 27 (35.5%), 23 (30.3%), and 17 (22.4%) patients (Fig. 3). Hence, there was overall agreement between the methods in between 38 (50%) and 48 (63.2%) cases, depending on the variability condition. Progression was infre-

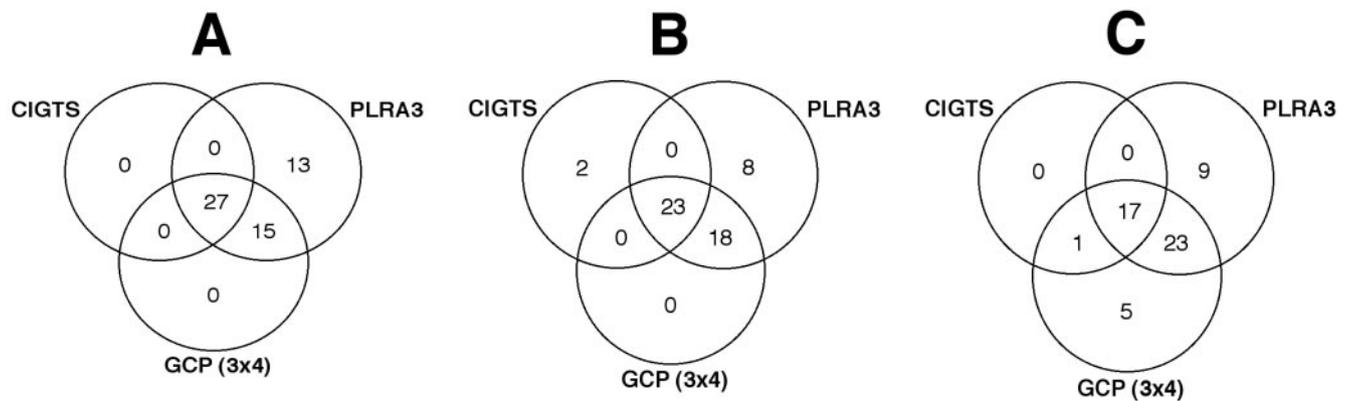


FIGURE 3. Agreement rates between the CIGTS, GCP (3×4), and PLRA3 methods for progressing cases using simulations with no (A), moderate (B), and high (C) threshold variability.

quently detected by the CIGTS or GCP (3×4) methods alone, whereas a sizeable number of cases were identified as progressing by the PLRA3 method alone. There was poor agreement between the CIGTS and PLRA3 methods (Fig. 3). Although the agreement between the CIGTS and GCP (3×4) methods was better, it deteriorated markedly with increasing variability. The agreement between the GCP (3×4) and PLRA3 methods was the highest (Fig. 3).

Relationship between Visual Field Severity and Progression

There were no statistically significant differences in visual field severity (as measured by baseline MD) between progressing and nonprogressing eyes using either of the seven methods under the no-variability condition ($P > 0.08$). With moderate variability, the baseline MD in nonprogressing cases was significantly worse ($P < 0.04$) compared with progressing cases for only the three GCP methods (-10.8 and -8.6 dB, respectively, for the GCP [2×4] method, -10.9 dB and -8.4 dB, respectively, for the GCP [$8, 2 \times 4$] method, and -10.6 and -8.3 dB, respectively, for the GCP [3×4] method). With high variability, the MD difference was significant for the GCP ($8, 2 \times 4$) method (-11.7 dB for nonprogressing and -8.7 dB for progressing cases, $P = 0.012$).

Visual field severity was also not related to the time to detection of progression with any of the methods under any of the variability conditions ($P > 0.06$).

DISCUSSION

In this study, we evaluated three different classes of methods for determining visual field progression in glaucoma. The AGIS and CIGTS methods are ordinal scoring methods based respectively on point-wise total deviation and their probabilities. Although the GCP analysis is commercially available with the Statpac program of the Humphrey Field Analyzer (Carl Zeiss Meditech) and is widely used, there is no consensus on the magnitude of the change required to constitute progression. Similarly, although PLRA methods have been used for many years now, there are no accepted criteria on either the number of test locations that show significant decline in sensitivity, the significance level, or the magnitude of change (decibels per year) at each location.

As a result, the performance of the GCP and PLRA methods depends critically on the criteria used and the number of repeated tests required to confirm progression. This is an inherent limitation of studies such as these that seek to evaluate the relative performance of these techniques. It can be

argued for example that selecting a more liberal criterion with PLRA, such as a significant decline in sensitivity at two locations, or at $P < 0.05$, with no minimum requirement for slope change or number of occasions that significance has to be confirmed would have led to both a higher number of progressing fields and a shorter time to detection. It is also likely that there would have been a larger number of falsely progressing cases with such liberal criteria. Because we also evaluated specificity, we were positioned to identify those criteria that were clearly too liberal.

Computer simulation techniques have been used extensively to evaluate the performance of thresholding algorithms²⁴⁻²⁶ in addition to methods for the evaluation of progression.^{11,12,14} The advantages of these approaches are numerous, including control of variability parameters, the frequency of tests, and the length of follow-up. By using the same initial and final visual fields, we were able to evaluate the specificity of the methods as they related to the number of progressing cases under the same variability conditions (Fig. 2). An evaluation of specificity from real longitudinal data may not be meaningful, as visual field progression cannot be ruled out, even if other clinical markers of disease progression appear to be stable. We have shown that the type of computer simulation used in this study shows close agreement to real visual field data.¹³ The advantage of this approach is that we did not impose a definition of what constitutes visual field progression, which would invariably result in an evaluation that is criterion dependent. The disadvantage, however, is that cases identified as progressing likely contain both true- and false-positive results. Hence although we were able to estimate specificity, because of the lack of an external standard for visual field progression, we could not evaluate sensitivity. By examining the changes in specificity and the number of fields identified as progressing under different variability conditions (Fig. 2), one can evaluate the effect of variability on performance.

The AGIS method was the most conservative and identified the least number of eyes as progressing. Like the AGIS method, the CIGTS method also had high specificity under moderate- and high-variability conditions, but the CIGTS method classified twice as many eyes as progressing as did the AGIS method and detected confirmed progression 1.4 to 2.3 years earlier than the AGIS method. Similar differences in progression rates between these two methods have been reported by Katz et al.,⁹; however, no differences were found in the time to progression. The fact that for similar specificity, the CIGTS method detected both more cases of progression and earlier time to progression than the AGIS method suggests that the latter may

be overly conservative. In comparison to the GCP and PLRA methods, however, the CIGTS method detected progression in substantially fewer eyes. The follow-up time to confirmed progression with the CIGTS method was on average approximately 1.5 years less than with the two PLRA methods, but was comparable to that of the GCP methods.

Because of the higher progression rates with the GCP and PLRA methods, it is likely they detect smaller changes than either the AGIS or CIGTS methods. The GCP (2×4) and GCP ($8, 2 \times 4$) methods had the highest variation in the number of progressing cases, the time to detect progression and specificity suggesting a considerable vulnerability of these methods to threshold variability. Under the high-variability condition, the GCP (2×4) and GCP ($8, 2 \times 4$) methods resulted in the highest number of progressing cases, but also lowest specificity. These findings suggest that these methods may result in an excessive frequency of progression in patients with moderately advanced field losses that exhibit the largest degree of measurable variability.^{21,27,28} In comparison, under the low-variability condition, the GCP (2×4) and GCP ($8, 2 \times 4$) methods have considerably higher specificity at a cost of a modest reduction in the number of progressing cases. It is interesting to note that although the GCP (3×4) method on average detected fewer cases of progression and at a later time compared with the other GCP methods, its performance was remarkably resistant to variability. These findings indicate that it is not the GCP analysis itself that results in performance variation, but probably the number confirmatory results that are necessary to declare progression.

The PLRA methods generally had high specificity. The PLRA3 criterion produced 100% specificity under both variability conditions. The PLRA methods also declared a relatively large number of fields as progressing. The drawback of these methods is that they required the longest time to detect progression (5.5–6.5 years). The criterion for progression at a single location was decay of 1.0 dB/y or more with $P < 0.01$. The same criterion has recently been used by Membrey et al.,²⁹ but $P < 0.05$ and 0.10 have also been reported.⁸ The probability level directly affects the number of fields needed before progression can be confirmed. Using our criteria, the earliest confirmed progression was detectable at only at the 5.5-year visit (i.e., at the 12th semiannual visual field). These findings are in accord with the results of a clinical study by Katz et al.³⁰ who showed that seven fields over a 6-year period are insufficient for PLRA to detect sensitivity decay of less than 1.0 dB/y. Furthermore in another study, detection of 1.0-dB/y sensitivity loss in an individual test point with $P < 0.001$ has been calculated to require, depending on the degree of long-term variability, from 10 to 14.5 years of follow-up, with semiannual testing of visual fields.³¹ Finally, in a computer simulation study, a minimum of eight annual tests have been shown to be required before a 1.0-dB/y point-wise sensitivity loss at $P < 0.05$ and specificity $\geq 75\%$ can be detected.¹¹ Such long follow-up times before confirmed progression can be detected may be unacceptable for both clinical and research purposes. However, when number of fields available and the follow-up increases, PLRA may be valuable particularly because the method can be highly specific. Furthermore, PLRA allows measurement of the rate of change of threshold or threshold deviation, which may be important in clinical management.

The ideal method for analyzing visual field change should be sensitive, detect progression with few examinations, maintain high specificity, and be resistant to fluctuation. Our results show that none of the methods investigated could achieve all these attributes. Methods that yielded a high number of progressing cases were often less specific and were influenced by fluctuation. In contrast, those methods that yielded high spec-

ificity often required very long follow-up times. Furthermore, in agreement with a previous study,⁹ there was poor concordance among the methods with respect to the patients identified as progressing. Although we have investigated with computer simulation techniques a range of methods to analyze progression, there are many possible variations in criteria for progression within a class of method. For example, the degree of decibel loss per year and probability level with the PLRA methods and other combinations of progressing points and confirmations with the GCP methods, may yield more favorable results. Our study confirms that detecting visual field progression reliably requires a considerable number of examinations and length of follow-up. For many patients, these conditions may not be acceptable to the clinician. It is not clear whether these findings are due to the nature of conventional perimetry itself or the methods of analysis. Future research should focus on both alternative methods of data analysis and other modalities of detecting glaucomatous progression.

Acknowledgments

The authors thank Alex Bates and Allen Butler for providing computer programming expertise and Paul Artes for constructive comments.

References

1. Spry PG, Johnson CA. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol*. 2002;47:158–173.
2. The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials*. 1994;15:299–325.
3. Collaborative Normal-Tension Glaucoma 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.
4. Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol*. 1999;117:573–583.
5. Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. 1999;106:653–662.
6. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology*. 1999;106:2144–2153.
7. Flammer J. Fluctuations in the visual field. In: Drance SM, Anderson DR, eds. *Automatic Perimetry in Glaucoma A Practical Guide*. Orlando, FL: Grune and Stratton, Inc.; 1985:161–173.
8. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Visual field progression: comparison of Humphrey Statpac2 and point-wise linear regression analysis. *Graefes Arch Clin Exp Ophthalmol*. 1996;234:411–418.
9. Katz J, Congdon N, Friedman DS. Methodological variations in estimating apparent progressive visual field loss in clinical trials of glaucoma treatment. *Arch Ophthalmol*. 1999;117:1137–1142.
10. Lee AC, Sample PA, Blumenthal EZ, Berry C, Zangwill L, Weinreb RN. Infrequent confirmation of visual field progression. *Ophthalmology*. 2002;109:1059–1065.
11. Spry PG, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci*. 2000;41:2192–2200.
12. Gardiner SK, Crabb DP. Frequency of testing for detecting visual field progression. *Br J Ophthalmol*. 2002;86:560–564.
13. Vesti E, Spry PG, Chauhan BC, Johnson CA. Sensitivity differences between real-patient and computer-stimulated visual fields. *J Glaucoma*. 2002;11:35–45.
14. Spry PG, Johnson CA, Bates AB, Turpin A, Chauhan BC. Spatial and temporal processing of threshold data for detection of progressive glaucomatous visual field loss. *Arch Ophthalmol*. 2002;120:173–180.

15. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;101:1445-1455.
16. Heijl A, Lindgren G, Lindgren A, et al. Extended empirical statistical package for evaluation of single and multiple fields: Statpac 2. In: Mills RP, Heijl A, eds. *Perimetry Update 1990/1*. New York: Kugler and Ghedini; 1991:303-315.
17. Holmin C, Krakau CE. Regression analysis of the central visual field in chronic glaucoma cases: a follow-up study using automatic perimetry. *Acta Ophthalmol (Copenh)*. 1982;60:267-274.
18. Fitzke FW, Hitchings RA, Poinosawmy D, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80:40-48.
19. Wild JM, Hutchings N, Hussey MK, Flanagan JG, Trope GE. Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology*. 1997;104:808-815.
20. The Advanced Glaucoma Intervention Study (AGIS): 4. Comparison of treatment outcomes within race: seven-year results. *Ophthalmology*. 1998;105:1146-1164.
21. Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci*. 1999;40:648-656.
22. Chauhan BC, House PH, McCormick TA, LeBlanc RP. Comparison of conventional and high-pass resolution perimetry in a prospective study of patients with glaucoma and healthy controls. *Arch Ophthalmol*. 1999;117:24-33.
23. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 1998;17:857-872.
24. Bebie H, Fankhauser F, Spahr J. Static perimetry: strategies. *Acta Ophthalmol (Copenh)*. 1976;54:325-338.
25. Johnson CA, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci*. 1992;33:2966-2974.
26. Chauhan BC, House PH. Estimating thresholds in conventional and high-pass resolution perimetry using computer simulation. *J Glaucoma*. 1994;3:132-139.
27. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130-135.
28. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654-2659.
29. Membrey WL, Poinosawmy DP, Bunce C, Fitzke FW, Hitchings RA. Comparison of visual field progression in patients with normal pressure glaucoma between eyes with and without visual field loss that threatens fixation. *Br J Ophthalmol*. 2000;84:1154-1158.
30. Katz J, Gilbert D, Quigley HA, Sommer A. Estimating progression of visual field loss in glaucoma. *Ophthalmology*. 1997;104:1017-1025.
31. Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci*. 1996;37:1419-1428.