Simulation of Longitudinal Threshold Visual Field Data

Paul G. D. Spry,1 Alex B. Bates,1 Chris A. Johnson,1 and Balwantray C. Chauhan2

PURPOSE. To describe and evaluate a computer model that simulates longitudinal visual field data.

METHODS. A computer model was designed using factors that influence thresholds of normal and glaucomatous visual fields. The simulation model was used to quantify the effects of fluctuation on the outcomes of pointwise linear regression by comparison with simulated gold standard data with no variability.

RESULTS. Serial sets of 10 stable and 10 progressive visual fields with different fluctuation levels were generated by simulation and were analyzed using pointwise linear regression. Regression outcome measures used were slopes of −1 dB/year or worse and slopes of −1 dB/year or worse that were also statistically significant. In stable visual fields, the number of locations with regression slopes worse than −1 dB/year increased with fluctuation and defect size and was inversely related to the number of fields. The number of locations with statistically significant slopes remained low and appeared unaffected by these variables. In progressive visual fields, analysis of a small number of visual field test results (<8) overestimated the number of locations with regression slopes worse than −1 dB/year and underestimated the number of locations with statistically significant slopes.

CONCLUSIONS. Computer simulation may be used to provide a gold standard outcome that permits evaluation of statistical tools for monitoring progressive glaucomatous visual field loss. (Invest Ophthalmol Vis Sci. 2000;41:2192–2200)

Detection of glaucomatous visual field progression remains one of the most difficult aspects of glaucoma management. The visual field is not a stable quantity, and it is therefore difficult to differentiate true change in visual field status (signal) from variability (noise).1 Measurement noise or variability is found for all visual field tests and other clinical psychophysical procedures. Variability is due partly to the probabilistic nature of psychophysical thresholds2,3 and partly by threshold estimation errors for thresholding strategies, such as staircases, used in clinical situations.4,5 Thresholds are not simple all-or-nothing responses and have zones of uncertainty representing intratest variability, or short-term fluctuation. In addition, a further source of variability is change in threshold between tests: intertest variability, or long-term fluctuation. Both types of fluctuation occur physiologically in normal eyes and have been found to be greater in eyes with glaucomatous visual loss.6 Unless true change in a glaucomatous visual field is larger than the combined short- and long-term fluctuation, it becomes statistically indistinguishable.

To date, methods for detection of glaucomatous visual field progression may be broadly grouped into three categories: subjective clinical criteria, event analyses, and trend analyses. Subjective clinical criteria represent scoring systems that stratify field loss by score and define progression as score change over time. An example of this is the Advanced Glaucoma Intervention Study (AGIS) visual field defect score,7 which has a range from 0 (no defect) to 20 (all test locations greatly depressed). However, the empiric basis for this scoring system is not well defined. Evaluation of AGIS scores demonstrates that 16% of individuals have a test–retest score change of four or more.8 Although quantification of visual field loss or reduction as a single number is easy to use and interpret, such a drastic reduction of data results in loss of spatial informational content.9 Subjective clinical scoring systems may therefore be unable to detect subtle visual field changes. In addition, there is no evidence that these scales are linear (i.e., a change in score from 2 to 6 may not represent the same change as from 12 to 16).

The second method is event analysis, which is been said to be sensitive to a single event of change relative to a reference examination. An example of this is glaucoma change probability (GCP),10 which calculates the difference in pointwise threshold deviation between a given field and reference mean threshold of a baseline test pair. Changes are compared with the test–retest difference distribution for stable glaucoma patients, and locations are highlighted as progressive or improved if the difference falls outside the upper or lower limits (5% and 95% probability levels, respectively) of the distribution. Although this method may identify test locations that appear progressive with as few as three test results, it is dependent on the degree of change exceeding test–retest variability, which is high for damaged locations. To maintain reasonable specificity, most investigators have found it necessary to have GCP points outside normal limits to be confirmed on one or more retests.11

The final method is trend analysis, which follows test parameters sequentially over time to determine the magnitude and significance of patterns within the data. Negative trends...
should exceed expected physiologic age-related loss to be labeled progressive.\textsuperscript{12} Trend analysis (linear regression) is of value, because it may provide the ability to extract small amounts of loss or signal from variability or noise.\textsuperscript{13} The time required to detect progression is influenced by factors including underlying rate and type of progression, degree of variability, frequency of examinations, and position of the visual fields within the time series.\textsuperscript{14–16} Trend analyses have been performed on individual test locations (pointwise), glaucoma hemifield test zones, and global indices. It has been suggested that regression analysis of any global index may diminish information from local defects and therefore may not be clinically reliable.\textsuperscript{17} Studies confirm that pointwise regression detects more cases of progression than global indices, suggesting that it has greater sensitivity, whereas global indices have greater specificity. Use of glaucoma hemifield test zones has been suggested as a compromise.\textsuperscript{15}

It is evident from the literature that there is no consensus on which method of detection of progression is best for differentiating stable defects from progressive loss. This is in part because there is no independent gold standard.\textsuperscript{18} It is therefore difficult to quantify the success of any tool that may be used for the detection of visual field change.

This article describes a new computer model that simulates longitudinal glaucomatous visual field testing. This approach permits generation of simulated visual field series with chosen levels of fluctuation and progression, allowing comparison of outcomes of statistical analysis from simulated visual field series with no variability with those exhibiting typical glaucomatous variability. We attempt to validate use of simulated data by comparing simulation-based evaluation of pointwise linear regression with data from published clinical evaluations.

**METHODS**

**Computer Simulation Design**

A C++ program simulates a set of serial visual fields between initial and final tests. Simulated visual field data are presented in the program 30-2 format of the Humphrey Field Analyzer (Humphrey Instruments, Dublin, CA) and are output as text files. A schematic representation of the model is shown in Figure 1. The following parameters are varied:

1. **Data Type**: Initial and final empiric input data are entered as raw threshold values or threshold (total) deviations from the age-matched normal data. Simulation output data are in the same form.
2. **Mode of Progression**: Simulated visual fields are produced by interpolation between the initial and final visual fields. The simulation provides for linear, episodic, and bilinear decay of data values. By making initial and final visual fields the same, no progression can also be simulated.
3. **Patient Age**: The initial and final ages determine the number of years over which visual fields are simulated.
4. **Number of Simulated Fields per Year**: This quantity can be one or more.
5. **Short-Term Fluctuation**: This quantity represents the fluctuation occurring during a single test session. It is equivalent to within- or intratest variability of threshold estimates made using the conventional full-threshold algorithm. The model uses the interpolated threshold and a short-term fluctuation value as the mean and SD, respectively, of a normal distribution. The range of the distribution described by these two parameters represents the extent of thresholds likely to occur for a particular location during a single full-threshold test session. The area beneath the normal distribution is then standardized, and a random number generator is used to select a short-term-fluctuation–adjusted simulated threshold. This is performed independently for each test location and sequentially for all fields within the series. The amount is specified by entering a value for the SD of an assumed normal distribution of threshold values around the interpolated mean.
6. **Long-Term Fluctuation**: This quantity represents the amount of homogenous fluctuation that occurs between one test and the next. It is considered equivalent to between- or intertest variability of threshold estimates made using the full-threshold algorithm.\textsuperscript{1} Adjustment for long-term fluctuation is achieved in a manner similar to that for short-term fluctuation, but a single number is generated for the entire visual field, and the whole visual field is adjusted by this amount.
7. **Eccentricity-Related Fluctuation**: When selected, this function produces an increase in short-term fluctuation with increasing eccentricity. The increase is based on reported change in short-term fluctuation measured with

![Figure 1. Schematic representation of computer simulation design.](Image 321x388 to 561x726)
the full-threshold algorithm. An eccentricity-weighting factor calculates a quantity that is added to the SD of the short-term fluctuation, thereby increasing the range of possible simulated threshold values.

Eccentricity-weighting factor = \[ \frac{\sqrt{(4.5 - i)^2 + (4.5 - j)^2}}{6.5} \]

where \( i \) and \( j \) are the coordinates of each location in a 10 × 10 grid, where the upper left hand corner is \( i = 0, j = 0 \).

8. **Defect-Related Fluctuation**: When selected, this function increases short-term fluctuation at locations with greater threshold deviations. For every 5 dB of deviation from age-matched threshold, 0.4 dB is added to the SD of short-term fluctuation, extending the range of probable simulated thresholds. This quantity is based on previous reports.

### Pointwise Linear Regression of Simulation Output and Statistical Analysis

Each simulated visual field set comprised 10 serial visual field test results. Two progressive and two nonprogressive visual field defect conditions were evaluated (Fig. 2 and also see Fig. 5). For each condition, the simulation produced 20 visual field sets containing 10 serial visual fields (200 visual fields in total). Short- and long-term fluctuations used were based on previously reported values obtained using the full-threshold strategy. Combinations of short-term fluctuation levels of 1, 2, and 4 dB and long-term fluctuation levels of 0, 1, and 2 dB were investigated for linear visual field progression. To adhere closely to the described behavior of glaucomatous visual fields, eccentricity-related and defect-related fluctuations were included.

Pointwise least-squares regression was performed for each of the 74 non–blind-spot locations within each 30-2 test series. This was performed for the first 4 through 10 test results within the series to assess the change in outcome with each additional field used in the regression. Two progression outcome measures were used: mean number of locations per field with slopes of −1 dB/year or worse and mean number of locations with a statistically significant slope of −1 dB/year or worse \( (P < 0.05) \). The selection of this −1 dB/year slope criterion was based on its use in previous clinical studies of progressive glaucomatous visual field loss. This criterion also represents the value used in commercially available software to calculate pointwise univariate linear regression. Additionally, the mean number of locations with slopes of +1 dB/year or better (improvement) were examined.

### RESULTS

#### Stable Glaucomatous Visual Field Defects

For simulation of stable but damaged visual field test results, the initial and final fields were the same, and therefore no progression was present. If zero short- and long-term fluctuation was specified, the regression slope at each location was found to be zero because each simulated visual field was identical. When fluctuation was added, a temporal trend resulted at each location. With long-term fluctuation alone, the resultant trend was the same for all locations. If short-term fluctuation was used, the trend at each location differed. Because random number generation independently specified short-term fluctuation at each location, similar numbers of locations with improving and declining temporal trends occurred for stable defect conditions.

The effect of different levels of fluctuation on the two stable defects (Fig. 2) studied are shown in Figures 3A, 3B, 4A, and 4B. Figure 3 depicts the effects of varying amounts of short- and long-term fluctuation on a small nasal step defect of mean deviation (MD) −0.27 dB, and Figure 4 shows similar effects for a moderate defect of MD −9.35 dB. The number of locations demonstrating negative regression line slopes of −1 dB/year were compared.
dB/year or worse was strongly affected by any degree of fluctuation (Figs. 3A, 4A). The numbers of such slopes increased with short- and long-term fluctuation and defect size and decreased with the length of follow-up. Short-term fluctuation exerted a greater influence over the number of correctly identified slopes than long-term fluctuation. The number of statistically significant slopes of $-1$ dB/year or worse was low compared with the number of nonsignificant slopes (Figs. 3B, 4B) and never exceeded two locations. The number of such slopes was not affected by fluctuation or defect size within the conditions studied.

Some locations exhibited positive regression line slopes in a manner similar but opposite to the negative slopes. These positive slopes were derived from locations where fluctuation produced sensitivity increases. Figures 3C and 4C demonstrate nonsignificant positive slopes of $+1$ dB/year or better in both simulated defects. The number of locations with such positive slopes was similar to those found with negative slopes, increasing with both short- and long-term fluctuation and defect size and decreasing with length of follow-up. No locations exhibited significant ($P < 0.05$) $+1$ dB/year slopes.

### Progressive Glaucomatous Visual Field Defects

Figures 5A and 5B present the initial and final visual fields for moderate and small degrees of progression, respectively. Figures 6A and 6B and 7A and 7B demonstrate the effect of different levels of fluctuation on two progressive visual field...
defects with worsening MD from $-0.28$ to $-9.35$ dB (MD change of $-9.08$ dB, Fig. 5A) and from $-4.13$ to $-10.06$ dB (MD change of $-5.93$ dB, Fig. 5B), respectively. The solid horizontal line indicates the number of locations undergoing true progressive change—that is, meeting the slope and significance criteria when no fluctuation was introduced. The criterion of slopes of $-1$ dB/year or worse overestimated the true number of progressive locations, whereas the criterion of statistically significant slopes of $-1$ dB/year yielded an underestimation. As observed for nonprogressive defects, the ability of both outcome measures to correctly identify progressive locations was related to the number of test results and decreased with degree of fluctuation. Of the two outcome measures, the number of significant slopes approached the true number of progressive locations more rapidly than the number of nonsignificant slopes.

Evaluation of positive regression line slopes revealed that in visual field series with underlying progression, some locations exhibited nonsignificant positive slopes of $+1$ dB/year or better when any degree of fluctuation was present (Figs. 6C, 7C). The number of such locations decreased with the degree of progression. This effect made some locations demonstrate an apparent improvement in sensitivity, although the visual field defect was progressing. The number of such nonsignificant improving locations decreased with length of follow-up and remained remarkable for typical levels of glaucomatous fluctuation ($1$–$4$ dB of short-term fluctuation and $1$–$2$ dB of long-term fluctuation) until at least eight examinations were performed.

**FIGURE 4.** The effect of different amounts of short-term fluctuation (STF) and long-term fluctuation (LTF) on the number of test locations (A) with regression line slopes of $-1$ dB/year or worse, (B) with statistically significant regression line slopes worse than $-1$ dB/year, and (C) with slopes of $+1$ dB/year or greater, for a stable defect (MD = $-9.35$ dB).
performed. No such effect was noted for significant ($P < 0.05$) slopes.

**DISCUSSION**

Evaluating the performance of analytic tools for detection of visual field progression requires an independent gold standard. This problem is widely recognized in the published literature, and alternative methods of defining glaucomatous progression have been attempted. One example of this is clinical assessment of optic nerve head characteristics. The disadvantage of optic nerve head assessments is that they are subjective and exhibit high variability, although quantitative imaging techniques may be helpful in this regard. Furthermore, it may be inappropriate to use a structural measure as a gold standard for evaluating progressive functional loss, because their relationship is not completely understood. A second example uses a series of multiple visual fields and uses final visual field in relation to the initial visual fields to define progression and nonprogression. This has the disadvantage of requiring a large longitudinal data set and depends on the correlation of the final visual field test result with prior tests.

An alternative approach to producing a gold standard is to replace real patient data with visual field data generated by a computer model. This technique is advantageous, because it permits assessment of progression analysis methods without requiring longitudinal patient data. Additional advantages of this technique include the ability to define the magnitude and type of progression and variability. This model may be constructed to generate a large, clean data set for rigorous statistical analysis using a design that emulates the behavior of a progressive glaucomatous visual field. The major disadvantage of using simulated data is that a poorly designed model may not properly simulate results obtained clinically.

Computer simulation has been used in perimetry to evaluate many threshold strategies and to study the effects of changing staircase properties on accuracy and efficiency of threshold estimates. Spenceley and Henson used simulated data to study the effects of increased levels of short-term fluctuation on perimetric threshold. These simulation experiments were able to provide information defining optimal visual field test strategies.

We have designed a computer model for studying visual field progression and analytic tools to detect progression. This procedure models physiological and pathophysiological visual field behavior by taking into account most factors reported to affect threshold variability and by emulating empiric data gathered with conventional full-threshold algorithms. Our model permits control over conditions of progression, and provides information that complements real patient data because simulated longitudinal visual field data can be generated without variability. When any given tool for detection of progression is applied to these data, analysis of simulated data without variability creates a gold standard. This information can be used as the yardstick for comparison with analysis of visual fields simulated from the same input and output data, but with variability added. Assuming that data simulated by the model are representative of empiric findings, the conclusions may then be generalized to patient data.

Use of simulated data to assess the specificity of an analytic tool for detection of progression provides a rigorous standard: Stable glaucomatous visual field data are created from identical baseline and final input data. However, assessment of sensitivity may be influenced by lack of an external standard when simulating progressive conditions, because the simulation model assumes that change between baseline and final fields represents real progression. Because the model uses initial and final empiric input data to simulate a longitudinal visual field series, this assumption is not entirely valid. Although marked defect changes between initial and final empiric data may represent true progression, small differences may be caused by short-term fluctuation or intratest variability present during data collection.
We have attempted to validate our computer model by evaluating pointwise linear regression as a tool for the detection of progression. Our assessment of pointwise linear regression is reinforced by evidence from clinical studies using longitudinal patient data. Our model showed that ability to detect progression by pointwise linear regression depends on the number of test results and on degree of variability, as has been concluded by others. For nonprogressive defects, use of slopes worse than or equal to $\frac{2}{1}$ dB/year as an outcome measure without requiring statistical significance makes stable simulated visual fields appear progressive, because many nonprogressive locations are misclassified. This occurs even when fluctuation is conservatively estimated (2 dB short-term fluctuation and 1 dB long-term fluctuation) and 10 simulated annual visual field results are evaluated. Use of significant slopes of $-1$ dB/year or worse for evaluation of the same stable defects misclassifies some locations as progressive, although their number is small and independent of fluctuation and number of examinations. For simulated progressive defects, locations with a nonsignificant slope overcall and locations with significant slopes undercall the true number of progressing locations. The locations with significant slopes approach the correct number of progressing locations more quickly than nonsignificant $-1$ dB/year slopes. Of these two outcome measures, analysis of simulated data indicates that a significant slope of worse than or equal to $\frac{2}{1}$ dB/year is the parameter of choice for discrimination of stability from progression. This outcome measure has been used in previous visual field investigations that support our simulation findings.

Analysis of simulated data with pointwise linear regression has demonstrated that an inadequate number of examination
results may cause misclassification of individuals with progression of visual defects as stable or vice versa, depending on the outcome measure used. We simulated and analyzed data from 20 iterations of two different progressive defects, one with an MD change from \(-0.28\) dB to \(-9.35\) dB and the other from \(-4.13\) dB to \(-10.06\) dB. Assuming that the amounts of short- and long-term fluctuation present in glaucoma are \(2\) dB and \(1\) dB, respectively, use of simulated data has shown that to correctly identify 75% of all progressing test locations, at least eight annual visual field evaluations are required. Lower accuracy is obtained if fewer test results are used within regression analysis. This is clinically important because clinicians may falsely believe that pointwise linear regression can be used to verify progression with fewer test results. If the real amount of glaucomatous fluctuation is higher than our conservative estimate, more than eight visual field test results are required before this 75% level of accuracy is reached. This finding is supported by investigations of empiric visual field data published by several research groups. For example, Katz et al.\(^1\) have shown that use of seven visual field test results performed over a 6-year period could not detect mean sensitivity changes of less than \(1\) dB/year. Other investigators have estimated that a minimum of 5 or 6 years of annual follow-up is required for pointwise linear regression to reliably detect glaucomatous visual field defect progression.\(^2\)\(^,\)\(^3\)

We have demonstrated use of a computer model to simulate longitudinal visual field data. Analysis of simulated data with pointwise linear regression produced outcomes that were

![Figure 7](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/932908/)
The results of this simulation study and prior empiric investigations indicate that pointwise linear regression is able to detect progressive field losses of 1 dB/year. However, to maintain high specificity, the slope of the regression line must be significantly different from zero, and a minimum of at least seven to eight annual visual field test results is needed. We intend to use this simulation approach further, to compare the sensitivity and specificity of different approaches to detection of glaucomatous visual field defect progression for different fluctuation conditions and amounts of progression, and to develop new, more robust methods of analyzing visual field changes over time.

References