VISUAL FIELD PROGRESSION IN GLAUCOMA. A REVIEW.

by

Alfredo Benjumeda MD, PhD

UNIVERSIDAD DE SEVILLA-FACULTAD DE MEDICINA
September 2006
CONTENTS

Introduction ................................................................. 3
The Nature of Progression ............................................... 4
Neuronal damage .......................................................... 4
Progression rate ............................................................. 7
Signal and Noises .......................................................... 10
Spatial filtering of the noise .............................................. 11
Measuring the fluctuation ............................................... 12
Variability vs. true change ................................................. 14
Methods for detecting progression ................................. 15
Choice of the best algorithm .......................................... 23
The Prediction value ...................................................... 24
Progression or stability ................................................. 24
Guidelines for progression tests ..................................... 27
Simulation of progression .............................................. 28
Conclusions ................................................................. 31
References ................................................................. 33
Introduction

Perimetry is used in Glaucoma management for various purposes including diagnosis, prognosis, risk prediction and disease screening. Once vision loss has occurred, it is irreversible; although the decline can be slowed or even halted by appropriate treatment in most cases. If untreated, the disease will normally progress, resulting in the vision loss in affected areas worsening and/or spreading to other areas; eventually this may result in complete loss of the vision in that eye.

Long-term follow-up of patients with open-angle glaucoma requires determination of either stabilization or progression of visual field loss and structural damage. Nevertheless determining what constitutes stabilization/progression of glaucoma damage is one of the most difficult tasks facing both clinician and researcher. The critical appraisal of methods employed in monitoring the amount of progression of the glaucoma visual fields, is the subject of this review.

Although structural damage can usually be detected in simple glaucoma before functional damage, evaluation of a longitudinal series of visual fields remains one of the most frequently used methods to detect early evidence of glaucoma and to observe patients. Functional deterioration probably becomes apparent on the test after further neuronal loss has occurred although visual field progression is often the only clinical sign that a change in the patient’s vision has taken place. The perimetry methods are used to chart the course of disease in patients with glaucoma, wherein progressive visual field loss is taken to reflect the worsening of the disease and is therefore usually regarded as a strong indication that the treatment regimen should be intensified (therapeutic impact). If the defect is stable, then unnecessary treatment can be both painful and costly; many treatments for
glaucoma have associated side-effects. Nevertheless it is necessary to keep in mind that the function stability could be in connection with previous treatments. Multiple medication changes and polypharmacy which may result from a false-positive diagnosis of progression may itself cause poor medication compliance and subsequent disease progression.

Visual function testing is the only thing that can assess effectiveness in the living human eye. Even if measures of optic nerve structure show no progression, this does not guarantee that function is spared or has recovered.

By both its diagnostic and therapeutic impacts computerized visual field testing remains after many years the benchmark for monitoring glaucoma.

The nature of progression
When a change is detected in a visual field series, the following three factors are likely to be responsible:
(a) Neuronal damage
(b) Noise.
(c) Spurious change due to learning and fatigue.

Neuronal damage
There are two feasible scenarios for ganglion cell alterations during glaucomatous progression. A dead-or-alive scenario: loss on a cell-by-cell basis is sudden, for example caused by previously healthy ganglion cells dying and causing a reduction in sensitivity.

But it is also possible that some of the recent psychophysical techniques are responding to dysfunction in cells rather than the total loss of function. In a ganglion cell dysfunction scenario each cell deteriorates from a functional perspective with ganglion cells, for example, deteriorating gradually over some time period \( t \) (deterioration parameter) from 100% functionality to eventual complete loss of functionality. If retinal ganglion cells deteriorate with gradual loss...
of function this would explain the gradual loss of visual field sensitivity seen in patients. If different cells deteriorate at different rates, this could then explain the increase in variability as sensitivity decreases. That is exposed in fig. 1 in the Divergent Dysfunction Model. 

Figure 1. Simplified version of the principle behind the Divergent Dysfunction model. The outcome of the Divergent Dysfunction model is that as glaucoma progresses (i.e., as t increases), the sensitivities will be spread over a wider range, since each is deteriorating by an amount different. In the figure as the three cells deteriorate over time at different rates, the resulting variability caused by sampling randomly from one of the three increases. The black symbols represent the average sensitivities of each of the three cells, with 90% confidence intervals shown by the error bars above and below; the gray box shows the overall 90% confidence interval for the sensitivity when any one cell is picked at random.

At t = 0, a theoretical stimulus has an equal probability of hitting any one of three receptive fields, each of whose sensitivities varies slightly over time, as indicated by the error bars above and below the symbol representing the sensitivity of each mechanism (in fact a Size III stimulus could cover as many as 200 receptive fields, but such numbers would result in too complicated a figure). The response will then vary within the gray-boxed range, giving roughly a 90% confidence interval for the sensitivity measurement. Over time, these three mechanisms deteriorate at different rates; by t = 40, the response will vary within the much larger range shown by the gray box on the right-hand side. In the actual model, the sensitivity is given by the maximum of the sensitivities of many cells chosen out of a larger population, instead of (as here) just one cell chosen out of three. (Gardiner, S.K., Demirel, S and Johnson, C.A, 15).
In quantifying the rate of progression of structural damage (retinal ganglion cell loss), it be borne in mind that at different locations and at different stages of disease there are different relationships between ganglion cell number and visual function. Observing how visual fields change over time can tell us about how neuronal damage of glaucoma progresses. So four distinct groups has been described: linear (49%), curvilinear (20%), episodic (7%), and non-progressing (24%) (fig. 2).

![Figure 2. Relationship between ganglion cells loss and threshold sensitivity with conventional automated perimetry. Mean, standard deviation, and fitted curve are shown (Girkin C.A., 16).](image)

Such a progression may be considered occurs in three phases: occult, threshold and critical:

a) Initially, there is a prolonged *occult period* during which sub clinical disease is present while white-on-white visual fields remain normal. Following psychometric and statistical convention, this true but unobservable state of illness can be called a “latent class”.

b) Visual field defects at the *threshold of detectability* have been suggested to be intermittent. It is well known that increased threshold variability can be an early indicator of pathologic changes in visual function. These may even disappear following treatment, only to reappear later; intermittent abnormalities can take up to
5 years to become consistently present. Visual field defects apparently appearing abruptly may simply represent those at threshold arriving at the end of their transient phase. It has been postulated that factors responsible for increased variability in this glaucoma stage could include greater background neural noise, for example atypical firing characteristics of damaged retinal ganglion cells, or reduced pooling of ganglion cell response signal caused by lower ganglion cell densities resulting from glaucomatous cell death. 

c) Once past threshold, the disease enters a critical phase, carrying significant risk of blindness. Disease is no longer early, and despite treatment, some eyes will show visual deterioration. Less than one-third of treated patients manifested statistically significant progression. Significant visual impairment may be present within a decade and absolute visual field loss within 30 years.

It has long been thought that, as glaucoma progresses, visual-field worsening tends to appear focalized as deepening or expansion of existing defects. Scotomata frequently evolve by deepening, although enlargement, frequently in Bjerrum’s area, and the development of new defects are also common. The hypothesis that glaucomatous visual field losses are exclusively focalized has often been challenged. Several studies have shown that focal visual field loss in glaucoma is usually associated with a diffuse component. With advancing glaucomatous damage, more test locations become involved (i.e., more diffuse losses).

**Progression rate**

As it occurs commonly in survival studies of diseases that often progress without symptoms and require periodical clinical evaluation, each glaucoma patient’s disease status is evaluated at regular intervals. The time required to detect progression and the pattern of the progression are 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. The increase in statistical power gained by frequent examination may be particularly important for persons with high long-term variability or for those with...
critical field loss in whom early detection of progression is essential. Progression rate is influenced by numerous additional factors possibly including both visual field measurement-dependent factors, such as variations in threshold accuracy and precision and measurement-independent factors, such as differential structural vulnerability to damage and current treatment, type and effectiveness: a great deal has been written about the effect of lowering of intraocular pressure on the disease evolution. The most important predictors for visual field progression seems to be the older age at the time of first glaucoma intervention, greater intraocular pressure fluctuation and higher mean IOP. Conversely, it could not be found a clear intraocular pressure level that reliably distinguished between glaucoma patients with stable and progressive fields. Likewise is true that many patients can experience further progression of the field despite extensive lowering of the intraocular pressure. Progression of visual field loss is dependent on the stage of the disease loss but baseline damage is not a reliable predictor. The progression of visual field loss starts rapidly when there is only little to moderate field loss and may be due to other factors besides IOP. With relatively moderate loss of visual field thresholds, there is a more rapid rate of loss of visual field than with more visual field loss. Because of ethical constraints, no prospective data are available on how quickly glaucomatous visual field defects would progress if left untreated. Between-eye correlation of visual field progression in patients with chronic glaucoma was founded statistically and clinically significant.

A major eye disease, as glaucoma is, exhibits strong but far from perfect correlations between eyes in their occurrence and this leads to substantial numbers of tied times of disease progression in the setting of interval-censored data. These tied times of progression can complicate estimation of the relationship between disease progression in right and left eyes.

**Signal and noises**

Validity refers to the ability of a test to measure accurately and reliably what it is intended to measure. Reproducibility is the ability of any measuring device to
detect a true difference among several measurements. In the case of perimetry, reproducibility is confounded by the fact that the object being measured, the sensibility threshold, seems to be constantly shifting like a shaky jelly mass (fluctuation, variability). 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). The vast majority of eyes that seem to have a worsening in the visual field on a single test will subsequently have that worsening disappear\(^37\).

**Normal noise**

Biological normal systems possess considerable variability, however. It is well recognized that the visual field of a normal individual often fluctuates on repeated testing\(^9\). The amount of fluctuation depends on many factors, including neural noise\(^18-39\), the subject’s ability to understand and do the test (cognitive function), the one’s criteria for deciding whether a light stimulus is present (signal detection criterion)\(^38\) and thresholding strategy.

*Short-term variability* is the major component of the variability. The variation in the threshold estimate within a visual field examination at a given stimulus location is known as the short-term fluctuation (SF). Probably a minor component is *long-term* fluctuation (LF), which means that the sensitivity of the entire field varies homogenously between different testing sessions over hours or days. *Long-term fluctuation* can be defined as the physiological fluctuation in retinal sensitivity, and is the net variability not explained by short-term fluctuation. These longer term sensitivity modulations have been attributed to reversible ocular and neural sensitivity fluctuations and are referred to as inter-test (between-test) variability. Both noises classes are expected to have a component based on the spatial configuration of the visual field (neuron cells density dependant). On a normal field peripheral test points tend to fluctuate more than points in the centre. Variability is incrementally greater with test point eccentricity: higher nasally than temporally and higher superiorly than inferiorly\(^10-24\).

Test duration likely plays an important role in determining the level of overall fluctuation whereby the longer the test, the higher the threshold variability (this increase in variability is commonly attributed to fatigue)\(^18-68\).
A phenomenon known as the “learning effect” has been identified whereby the patients baseline visual field is worse than subsequent tests. But significant learning effects are unlikely when it was required that subjects have reproducibly normal and reliable visual fields at baseline.

**Abnormal noise**

But it has now been established that larger amounts of variability are found in the visual fields of glaucoma suspect patients and with glaucoma than in normal subjects. This fluctuation can originate either in the stimulus or in the observer perception system.

Both short-term and long-term fluctuations are influenced by the severity of the disease. Test points with lower sensitivity (i.e., within a defect) tend to fluctuate more than test points within an area of higher sensitivity. For functional progression it would be more accurate, from a practical point of view, to evaluate such instability in searching closely the areas of existing scotomas.

Variability in patients with glaucoma can be reduced employing larger stimulus size, at the cost of loss of spatial resolution.

Whenever looked at, if field defects are inevitably accompanied by fluctuation of thresholds, this must seriously affect the reliability of the recordings and make the decision of a change highly hazardous. A difference between examinations done at different times must be greater than the expected physiologic fluctuation to be recognized as a genuine alteration. The magnitude of the variability is such that has been suggested that clinicians should not make judgments about progression of disease without at least five to six automated visual fields in hand.

In the presence of large amounts of variability, the separation of true change in the visual field from fluctuation becomes an important problem. Because the rate of progression of adequately treated chronic open-angle glaucoma is slow, one could make the assumption that changes in measurements between a baseline visit and a visit 6 months later represent variability, rather than true change. Lending support to this assumption is the observation that at the 6-month visit the mean...
values were stable, and roughly an equal number of eyes had “improved” as had “worsened,” suggesting that we were observing variation rather than progression.

All the three possibilities of visual fields’ sensibility change: for better, stability or deterioration can become obscured by variability effect.

Knowledge of the biological variability of the measurements in a healthy population allows the clinician to determine the likelihood that the visual field or the optic disc shows signs of damage. The LF derived from patients with stable glaucoma provides a statistical indicator for the determination of progressive glaucomatous visual field loss. So, at present, to identify glaucomatous visual-field progression, the clinician must identify changes that exceed the level of long-term fluctuation (LTF) for that individual (normal+abnormal fluctuations), then identify that the change is consistent with glaucoma and, ideally, correlate this defect with optic-disc appearance.

**Spatial filtering of the noise**

True change in a glaucomatous visual field has to be larger than the noise before it becomes statistically distinguishable. Improved methods of data acquisition have tended to focus on reducing the time of perimetry examination rather than making the measurement more accurate. We can also process the data in such a way that the noise would be reduced, without any additional testing time. One way of doing this relies on exploiting the relationships between the actual sensitivities of different points; in essence, if one point has a reduced sensitivity, then its neighbors are more likely to also have reduced sensitivities. This principle of *spatial filtering* is a widely employed image-processing technique used to improve the quality of digital information and points towards spatial filtering of the data as a possible solution (fig 3). Computer simulation has been effective to test not only whether the noise was being reduced by filtering, but also whether true localized defects were being blurred out.
The visual field test locations, unlike say pixel values in digital images, are not physiologically linked in a “grid-like” fashion as indicated by the matrix of values on the visual field chart and the Gaussian filter takes no account of the actual anatomical structure of the optic disc.

![Visual Field Matrix](image.png)

Figure 3. How the new sensitivity is calculated by the Gaussian filter. In this method, the raw sensitivity value is replaced by one derived from a linear combination of the sensitivities at the nine points in a square centered on the point of interest. This is repeated for each point in the field in turn, each time looking at the points in a square surrounding the point of interest (Gardiner, SK, Crabb, DP, Fitzke FV, Hitchings, RA, 14).

In that sense Gonzalez de la Rosa has obviated the problem developing a spatial filter taking into account the structure of the retinal nerve fiber layer [based on coefficients of determination (r²) between all the points of the central visual field previously published 19] that will reduce the noise without obscuring true defects 20.

**Measuring the fluctuation**

Techniques for measuring the fluctuation of the visual field over time have been developed by several workers. Varying perimetry strategies used for threshold estimation introduce different amounts of measurement error and, therefore, influence measurement of variability made in this manner. In general, calculations of fluctuation in automated perimetry have been based globally, upon the variance of repeated determinations of threshold. Using this method based on variance, the
total fluctuation of the entire visual field for each subject is defined as the square root of the mean variance of all subject's test locations, excluding test locations for which the threshold was always zero dB. The SLV and PSD global indexes have three components: the short fluctuation, the local defect and the errors induced by the strategy.

Some authors, however, have based variability calculations on the range. If only a few visual fields are available on a patient, the range or difference between the first and last fields may be more useful for detecting progressive change in the field than a variance calculated from a small number of values. Range variability for a single test location is defined as the absolute difference between the highest and lowest sensitivity readings for that test location. With this method, total fluctuation of the entire visual field for a single subject was defined as the mean of the range variability of all test locations, excluding test locations where the sensitivity was always zero dB.

The most widely known measured parameter is short-term fluctuation (STF), which a simple global index is derived from multiple threshold estimates made during a single test. STF may be considered as the clinical equivalent of intra-test variability, although it is influenced by the thresholding strategy. STF was originally defined as the scatter occurring during one visual field test and may be thought of simplistically as the average standard deviation of the differences between two threshold estimations at a number of predefined test locations. STF is available in the statistical packages on both the HFA (PSD) and Octopus perimeters (sLV) and is usually between 1 and 2 dB in normal individuals.

It is also possible to measure the clinical equivalent of inter-test variability, referred to as long-term fluctuation (LTF), which can be calculated mathematically from multiple visual field examinations, although this index is not available on commercial instrumentation. As previously said, empirical observations and simulated experiments have found that short-term fluctuation is a far more important factor for determining total variability than long-term fluctuation.

An alternative method for quantifying (LTF+STF) variability using data gathered with clinical visual field instrumentation is test-retest analysis that should be
regarded as a compound measure of within and between-test variability (fig 6). This technique has been used in a number of reports, and it quantifies the scatter of measures made on two or more examinations at different test sessions. Like the other variability quantification techniques described, this method also demonstrates increased scatter of threshold estimations when sensitivity is reduced due to glaucoma. Test–retest analysis is advantageous as it can be generalized to clinical environments and easily demonstrates the profound effect of variability in glaucoma. In damaged areas of the visual field, considerable test–retest variability occurs; for example, test locations with sensitivity deficits of - 6 dB can exhibit variable sensitivity deficits of between -1 dB to -16 dB on 90% of occasions when retested. This means that the test–retest variability for such initial defects takes in nearly half of the measurement range of the instrument. Furthermore, in locations with initial deficits between -8 dB and -18 dB, the variability extends to encompass almost the entire measurement range of the instrument. Clearly, this presents a significant challenge for attempting to determine whether a change from one examination to the next represents progression of a pathologic process or merely variability between examinations.

**Variability vs. true change**

So the goal of methods designed to detect progression in automated visual fields is to distinguish between fluctuation and a true decline in visual function. But the magnitude of inter-test variability is high and tends to be greater in patients with glaucoma than in normal subjects, rendering this task more difficult. This variability can both mask and mimic glaucomatous change, and several statistical methods have been devised to distinguish fake from true progression. The lack of gold standard against which glaucomatous change can be measured has led sometimes to the development of both arbitrary and empirical criteria for change. Criteria developed for clinical use have been “Arbitrary” (x points changing by y dB represents significant change) or ‘Empiric’ (derived from the study of a population followed over time) 49.
Less has been written about assessing progression through crossed examination with the structural correlates of glaucoma damage, namely, optic disc configuration, nerve fiber layer parameters, and macular thickness measurements. Variability in imaging also occurs, albeit without the component of the subjective patient response, as in visual field testing. It is generally accepted that progression will require a change in parameters that exceeds the variability seen in persons whose conditions are stable.

A given parameter may change dramatically only at early stages of the disease, thus being useful for early progression, while having a poor sensitivity to change when the whole range of disease is considered.

**Methods for detecting progression**

The concept of “progression” is in itself a soft endpoint. "Hard" endpoints on a clinical trial are well-defined in the study protocol, definitive with respect to disease process, and require no subjectivity. "Soft" endpoints are those that do not relate strongly to the disease process or require subjective assessments by investigators and/or patients. Progression is a “modification inside the illness” rather than a magnitude. It’s the key.

The reliability indices of the visual field test should be regarded as a primary consideration when assessing a visual field for progression.

How do we assess the probability of an uncertain event or the value of an uncertain quantity? Visual field and optic disc changes have been measured using a variety of methods that range from a subjective clinical judgment based on only two observations to complex statistical analyses of many observations in time. Because we lack an independent measure of glaucomatous progression that in itself does not rely upon a measure of function or structure, it may be prudent to clearly distinguish between visual field and optic disc changes, rather than referring to “glaucomatous progression”.

A) **Subjective means**. Progression of visual field loss in many studies has been evaluated by experts; that is, the observer determined by himself whether
progression had occurred with non-automated or automated visual fields simply looking through charts. The experts rely on a limited number of heuristic principles that reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations.

Figure 4. The figure shows a possible event model for clinical evaluation of a series of 6 visual fields. First, a learning effect field (L) is performed that can be excluded from analysis. Next, baseline visual fields are collected (B1, B2). If any future field worse than the worst baseline field (B2) is considered outside normal variability. Thus, follow-up fields that are equal to or better than the worst baseline field are considered stable whereas a field (F1) that shows more defects than the worst baseline field is flagged as suspected progression. If progression is suspected, then one or more confirmatory fields are performed (F2) (F3). If the confirmatory fields are also worse than the worst baseline field, then progression is diagnosed (Jansonius, N.M 33).

Subjectively assessing the results of serial automated perimetry requires the analysis of complex three-dimensional numerical information (visual sensitivity presented spatially in time). It is also possible that clinicians might rely too much on the gray-scale schematic in the visual field printout, which is probably the most misleading part.

Subjective grading of changes in the visual field and optic disc has been used to categorize progressing and non-progressing eyes. In this type of assessment, there is minimal, if any, standardization of the criteria used for change. Advantages of this approach are the similarity to real clinical practice and ease of use. 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.

As neither are well-accepted standards for subjective evaluation and because criteria for perimetric deterioration are likely to differ among observers, the interpretation of visual field data can be inconsistent\(^{58}\). There are significant limitations, including the often reported poor inter-rater and intra-rater agreement\(^{61}\).

<table>
<thead>
<tr>
<th>Observer A</th>
<th>Definitely Stable</th>
<th>Probably Stable</th>
<th>Probably Progressing</th>
<th>Definitely Progressing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely Stable</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Probably Stable</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Probably Progressing</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Definitely Progressing</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5. An example of agreement /disagreement inter-observer on decision about glaucomatous visual field progression (Viswanathan AC, Crabb DP, McNaught AI, Wescott MC, Kamel D, Garway DF, Fitzke FW, Hitchings RA, 61).

B) Statistical Methods. Automated visual fields are essentially made up of a grid of numerical sensitivity values, making them amenable to an assortment of statistical analyses, both frequentist and Bayesian approaches, eased by available software packages.

In monitoring data it should keep in mind that although statistical findings belong to the strict domain of the mathematics, in many instances either the intuitive judgments\(^{58}\) or the clinical criteria belong to the scope of the empiric and arbitrary\(^{49}\).

a) Global indexes. In a time-sensitive environment, such as in many ophthalmic clinics, an accepted and quick assessment of glaucomatous defects for clinicians can be made using one set of methods for detecting visual field progression relies on estimates of change in summary measures of the field (so-called global indices) such as mean deviation (MD) and SLV. The MD is available on any test strategy and corresponds to the mean elevation or depression of the patients’ visual field
compared to a normal reference field. The calculation of the MD takes into account age-corrected normal values and the normal degree of variance at each of the 54 test locations used in a Humphrey 24-2 visual field. In an eye with reliable and stable visual fields, one should expect only very little fluctuation in the MD. If progression affects limited sectors of the visual field, its effect on MD is reduced by the influence of a large number of points that do not show progression.

The advantageous simplicity of these methods is outweighed by the fact that they largely or completely ignore the detailed spatial information contained within computerized field tests, and they are reported to be insensitive to glaucomatous change.

**b) Spatial patterns of sensitivity loss.**

The most important data obtained from perimetry concern the point wise depth of visual defects with respect to expected normal visual sensitivity at each of the tested visual field locations.

Methods considering change in parts of the fields or at individual locations are of course more sensitive. Three main methods-classes of checking changes in perimetry: (1) clinical trials (2) event analysis and (3) trend analysis.

(1) **CLINICAL TRIALS.** A number of multicentre trials have adopted objective criteria for evaluating visual field deterioration, which are helpful for standardizing interpretation of results within studies. Subjective clinical criteria used represent scoring systems that stratify field loss by score and define progression as score change over time.

The Collaborative Normal-Tension Glaucoma Study (Collaborative NTG Study), the Early Manifest Glaucoma Treatment Study (EMGT), the Advanced Glaucoma Intervention Study (AGIS) the Collaborative Initial Glaucoma Treatment Study (CIGTS), and the Ocular Hypertension Treatment Study (OHTS) have all used visual field criteria requiring specified minimum change, often within a cluster of test locations, before apparent progression is attributed to disease.

While plausible and clinically reasonable, the empirical bases of these analytical methods have not been well described. Such systems were specifically designed
to detect meaningful change considered clinically significant by the originators of the scale. Also there is no evidence that the scales of scoring systems used necessarily reflect uniform increments of disease severity; consequently, each study probably interprets progression differently. For such staging of visual field damage, a great number of gradations are employed in their score, namely 20. The greater the number of gradations, the greater the sensitivity is. However, at a certain point the gradations become meaningless, because the examination method lacks the ability to make a meaningful distinction between the adjacent stages. Subjective clinical scoring systems may therefore be unable to detect subtle visual fields 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 ideal scoring system must balance between stages that are sufficiently close together that they will detect meaningful clinical change, and yet sufficiently far apart that they can be used to describe changes that actually can be reproducibly noted.

2) THE EVENT ANALYSIS. (Glaucoma change probability program) is an empirical method to detect change that compares a single visual field to the results of one or two earlier tests. In this form of analysis, a follow-up examination is compared to a baseline of a single or mean of two or more examinations. Progression is then defined on the basis of a somewhat arbitrarily chosen cut-off, for example a given number of test locations with a given amount of change on a given number of consecutive follow-up examinations (fig. 4). The typically binary nature of this approach (stable/ progressing) is subject to limitations similar to those discussed for subjective grading. Event analysis defines progression to have occurred if threshold sensitivity at a particular field test location changes more than the expected variability of a baseline pair of visual fields. The method of identifying visual field change is to derive statistical limits of test-retest variability in a group of patients with stable glaucoma (tested frequently in close succession e.g., over a few weeks, so that any differences between the test
Figure 6. **Test-retest analysis.** The changes (i.e. the differences between baseline and follow-up examinations) are compared to the test–retest variability from a separate sample of patients. If the observed changes exceed the test–retest limits (typically estimated by empirical 5th and 95th percentiles), actual change is likely to have occurred. The analysis is based upon detailed empirical knowledge of the variability found at all stages of glaucomatous visual field loss through knowledge gained in extensive multi-centre clinical trials in North America, Europe, and Asia. The plain language analysis (GPA Alert™) is based upon the criteria used for ten years in the Early Manifest Glaucoma Trial. (HFA; Carl Zeiss Meditec, Dublin, CA).

results can more likely be attributed to variability rather than to true change). These limits are then applied to visual fields of patients who are observed over time. If, at any given test location, the difference between the current test and a previously established baseline is beyond the lower limit of test-retest variability, the location has likely worsened (fig 6). The advantage of this approach lies in its theoretical ability to detect an event of change with as few as three tests. If the two baseline fields are for any reason not representative of the true baseline, the comparison made by this method is more unstable than with regression.

In event-based analysis, some subset of t1 through tn-1 is used to determine a baseline threshold, and then tn is examined in the context of that baseline. The *glaucoma change probability* (GCP) analysis program is available in the market with the Humphrey field analyzer (HFA; Carl Zeiss Meditec, Dublin, CA).

3) **TREND ANALYSIS.**
With trend analysis, the pattern of change in sequential visual fields is modeled over time. Linear regression is a simple method to estimate trends in longitudinal data such as those of serial automated perimetry. Univariate linear regression is performed with time as the independent variable. Trend analyses relate the magnitude of change to the variability observed within the individual data series and therefore do not rely on population derived variability limits that can be inappropriate in individual patients (fig. 8).

It has the potential to identify those persons whose trend is sufficiently strong to exceed the variability in the data and thereby to reach statistical significance. To make reasonable estimates of the rate of change with this technique, several visual fields examinations are required.

![Event analysis diagram](image)

**Figure 7.** Illustration of event analysis criteria at one test location. *Solid horizontal line*: the mean of the two baseline test results. If the deviation at a subsequent follow-up examination is more negative than the lower 5% limit of test–retest variability (*dotted line*), deterioration is flagged with a *filled triangle*. In this example, the progression criterion has been met after the last observation (3 years). Artes PH, Nicolela MT, LeBlanc RP, Chauhan BC. (3)

So the different parameters derived from visual field tests can be evaluated by regression analysis: using regression of mean sensitivity of certain zones of the visual field, point by point linear regression (point wise), clustering of points with
significant progression, separating MD of normal and pathological zones or observing result reproducibility in consecutive examinations etc. The Humphrey perimeter software includes a sole PLR analysis function that calculates a global slope for MD beyond five visual fields.

Given the relative lack of specificity of MD for glaucoma change, point wise linear regression (PLR) was developed as a technique for dealing with the inter-test variability in threshold seen with automated visual fields. It examines the trend of threshold sensitivity at each test location over time and provides an estimate for the rate of change at each location in the visual field.

Point wise linear regression (PLR) determines the slope of the best fitting linear regression across $t_1, \ldots, t_n$, and flags a given location as “progressing” if the slope is greater than some predetermined confidence limit and is statistically significant.

Again, the confidence limit is based on a database collected from a large sample of subjects. In spite of the obvious claims of PLR’s being a clinically useful tool for examining longitudinal visual field data, there is no consensus on what value of regression slope and $P$-value constitutes progression and whether it should be maintained in subsequent fields. Similar to event-based analysis, the definition of progression using this approach is arbitrary and can be based on the number of visual field locations showing a given magnitude of slope and/or statistical significance.

Let two parameters to keep in mind: slope and probability. A standard criterion example: a point was considered to be progressing if the regression slope was $\leq-1.0\ dB/year$ with $P \leq 0.01$. Improvement at a test location was described as a regression slope $\geq1.0\ dB/year$ in the presence of $P \leq 0.01$. This is the simplest commonly used PLR criteria in published studies, and it is used clinically as an indicator for change in some centers.

In Spain Gonzalez de la Rosa, as a part of his, up to now unpublished new program TNT, has used linear regression operating on the cumulative defect curve, native to the Octopus Field Analyzer system (Haag-Streit, Bern), also known as Bebie curve, in an explicit method of sectoring progression analysis. On ordering
from best to worst a series of previously filtered thresholds with random fluctuation, the deviation of those situated at the same level in the order is supposed to acquire much higher stability than habitually measured fluctuation at each position of the visual field. He analyses 8 sectors of each curve. Trend-based approaches are available in visual field analysis programs such as Statpac 2, Progressor; (OBF Laboratories UK Ltd., Wiltshire, UK), Peridata (Peridata Software GmbH, Huerth, Germany), TNT Threshold Noiseless Trend, (La Laguna University, Spain).

Figure 8. Illustration of trend analysis with point-wise linear regression using the criteria of the two-omitting method. A test location was classified as having progressed if the slope of the linear regression line was more negative than -1 dB/yr and significantly different from zero at P< 5%. These criteria must be met when the last examination is excluded (a) and also when the penultimate examination is excluded (b). Solid lines: regression obtained by omitting the last or penultimate observation (open circles). Dotted line: regression including all observations. In this example, the location was not classified as having progressed because the slope was not significant when the last observation was excluded (a). Artes PH, Nicolela MT, LeBlanc RP, Chauhan BC.(3).

**Choice of the best algorithm**

To be useful, a new perimetric method should demonstrate superior psychometric properties, including low variability, strong validity, high sensitivity and specificity, and high predictive values (positive and negative).
As with any other diagnostic test, the basic performance of a progression detection algorithm can be described in terms of sensitivity and specificity. The ideal clinical test for detection of progressive visual field loss should be sensitive to subtle changes in visual field sensitivity, and it should discriminate between stability and progression on a test-by-test basis. However, at present this ideal has not been achieved.

One of the reasons that there is little consensus about which method of visual field detection offers the best specificity and sensitivity in diagnosing true deterioration is the lack of an independent gold standard: the absence of a reference standard method. Of the several different methods and criteria of determining visual field progression that have been proposed, there is no agreement about which is the best. Indeed, the method that is most suitable for use depends on the purpose to which the results are to be put.

a) When the outcome of an eye’s being labeled as progressing would be the patient’s undergoing a risky or costly change in clinical management, a method should be used that is known to falsely label very few stable eyes as progressing (i.e., a high degree of specificity).

b) Conversely, when failing to treat a progressing eye would be much more damaging than treating a stable eye, it is better to use a test that will correctly flag progressing eyes quicker (i.e., a more sensitive test).

As the frequency of testing increases, the sensitivity of PLR increases. Three tests per year provide a good compromise between sensitivity and specificity.

The predictive values

Although statistical methodology is well developed for comparing diagnostic tests in terms of their sensitivities and specificities, comparative inference about predictive values is not. The positive and negative predictive value measure the probability of disease conditional on the test result. The PPV is the probability of progression, given that the test is positive, and the NPV is the probability of stability, given that the test is negative. The bayesian predictive value has greater...
clinical relevance than the sensitivity and specificity and is more directly applicable in patient care.

**Stability or Progression**

The key-principle is the failure of a gold standard. In the final analysis we have no independent qualifier of glaucoma or its progression since the classes of tests that are used to measure its severity or progression are the very ones that are used to define it.

Hence when glaucomatous progression, defined on the event of trend-based analysis of SAP, is used to evaluate a new psychophysical test, the latter is by definition at a disadvantage. An independent qualifier, such as a blood test, that can definitively classify progression and non-progression does not exist.

When comparing qualitatively different methods of progression measure (functional and structural), rather than comparing absolute changes (for example, the loss of visual field sensitivity over time in dB/year or the loss of neuro-retinal rim area in mm²/year) the *evidence of change index* (EOC) analyses compare the strength of the evidence that the measurements have changed over time. This is accomplished by relating the magnitude of the observed trend (signal) to the observed variability of the measurements (noise). The EOC analyses therefore based on an ordinal grading of the statistical significance (p-value) of the trend (fig. 9). This analysis does not attempt to quantify the amount of progression, but it quantifies the *strength of the evidence* that change has taken place.

In general, progression of visual field loss over 5 years occurs in the majority of glaucoma patients. Because there is no gold standard for evaluating progressive visual field loss and results from the various methods differ; it is advisable to apply more than one method.

The fresh specialized literature displays relatively a lot more of research on PLR’ methods. Although PLR software features, in the theory, a better performance in handling glaucoma progression, surprisingly its availability is tangled.
Compared with AGIS criteria PLR detects progression in a similar proportion of eyes. Point wise linear regression may be superior to AGIS methods since it identifies fewer visual field series as improving\(^47\).

In data modelling PLR techniques have been shown to have high specificity and to detect progression in more patients but they may require more tests (or time) than event-analysis techniques\(^59\).

Trend analyses software have a useful advantage over event analyses in clinical practice in that, instead of a binary outcome, a rate of change may be estimated. This is particularly useful at the earliest stages of the disease process and in ocular hypertension. A measured rate of progression may assist in the assessment of a patient’s risk of development of functionally significant visual loss and in the decision to commence or alter treatment.

Including clinical and perimetric data, Nouri-Mahdawi and cols.\(^46\) addressed the important issue of predicting the course of the entire visual field series at 8 years using information available only during the first 4 years of follow-up and PLR methods (using multivariate logistic regression). A summary index such as sum of slopes, would allow the clinician to calculate a probability for future worsening of glaucoma.

Event-based analyses (GPA), at present prevalent in the Perimetry market, is well-founded rather on statistical empirical rules, primarily based on the variability of test retest measurements. If a test provides low variability compared to the measurement range, then it is likely that more progressive events can be detected.
by this technique. Conversely, if the test has higher variability there will be fewer detectable events. Because the events are defined statistically with respect to measurement variability, they may not be related to the same biological event in the eye.

With either trend or event analyses, there are a large variation in the progression rates obtained with different fail-safe criteria (fig 7 and 8). *Progression suspicion* and progression *confirmation* criteria require more examinations of patients but afford better specificity. Clearly, the key is which way the method flags eyes in the middle ground that could be judged to be either progressing or stable. Normally, tests that are more sensitive are less specific, and vice versa. Comparing the performance of event and trend analysis methods about detecting early deterioration is difficult because of the criterion choice. The judgment of whether one technique is more sensitive to change than another requires that the specificities of the competing tests be equalized.

The choice between different criteria involves an essentially arbitrary balance between sensitivity and specificity, neither of which are known or are easily determined. Sensitivity, specificity, and time to detect progression are influenced by the progression criteria selected and, in the absence of an independent reference standard for progression (gold standard) performance may be judged by data modeling. Agreement on progression between expert clinicians is poor or moderate at best, and this precludes it from being a good arbiter of different methods.

**Guidelines for progression tests**

Some stringent norms are established for correct assessment of technological clinical devices. Lately, the following guidelines and considerations on the characteristics that would make in special a determination of glaucoma progression acceptable have been proposed, literally:

1. The measurement and criteria used in clinical studies should have good performance in recognizing progression and in separating the groups being
compared (stable versus unstable, rapid deterioration versus less rapid, and so forth). This requires good reproducibility of the tests and that the parameter used have small and known biological fluctuation in an individual from one time to another.

2. The measured values and criteria that provide good sensitivity and specificity in recognizing progression of the disease may very well differ from those that are sensitive and specific in recognizing the presence or absence of the disease.

3. The measurement should be accepted as relevant to the course of the disease and its clinical outcome. This may include direct measurements of structure (axon count) or function (vision) exemplified, for example, by retinal nerve fiber measurements, optic disk parameters, electrophysiological measurements of ganglion cell activity, visual field as well as other visual psychophysical testing.

4. In principle, an acceptable measurement may also be a short-term surrogate of long-term outcome, that is, some predictive parameter for which there has been documented linkage to future progression and outcome. Currently, intraocular pressure is accepted as a surrogate for expected future course. The choice of another suitable surrogate depends not only on whether there is a demonstrated association between the surrogate and the outcome, but also on whether the treatment alters the surrogate in a manner that would alter the outcome. The former requirement does not always imply the latter.

5. The outcome parameter should be capable of reliably measuring a small increment of change compared with the range of change in that parameter over the course of the disease from its onset to the point of visual impact on the patient’s lifestyle.

6. Validation of any candidate measurements and criteria is necessary. Comparison of a newly proposed criterion to existing progression criteria is helpful, but if a new test or criterion does not match an existing “gold” standard, it may be better rather than worse than the criteria used previously. Cross-sectional analysis of the proposed parameter at various stages of disease compared with its variability is an indication of potential sensitivity to recognize change and may also define the range of the disease over which the parameter may be useful.
Simulation of progression

To detect progression and evaluating new methods necessitates large, longitudinal sets of visual fields, which are usually difficult and time consuming to obtain. An alternative method of assessing longitudinal data sets is by computer simulation. The logic of the simulation blends in itself real data with arranged data. The advantages of these machine-made approaches are numerous, including:

a) Computer simulation of the tests with introduction of confounding variables that affect test outcome, may easily estimate relative performance of several proposed tests by modeling surrogate standard. So we can survey the specificity of any progression test on a glaucoma fake stable series. Elsewhere by simulating flimsy glaucoma progressing series we can probe sensibility. By introducing confounding variables it could also be validated the discriminative abilities between subjective expert opinion and objective proposed tests (man vs. machine).

b) A control of variability parameters (adding or removing noise). The true status (noise-free) of the eyes of a series is unknown, and so comparing test results from different methodologies with the actual status is impossible. Therefore, it is obvious to look for ways to specify the noise-free behavior of an eye over time. The key to any accurate visual field simulation is how well this noise or variability is estimated. Factors affecting variability can be mathematically composed, so allowing comparison of outcomes of statistical analysis from simulated visual field series with no variability with those exhibiting typical glaucomatous variability.

c) Getting visual field series (stable or progressive) with chosen levels of fluctuation and progression. When noise is added in, the series can be tested for progression, and the test results compared with this specified behavior.

d) To handle the frequency of tests interpolated, and the length of follow-up.

The simulation progression program published by Spry and cols. permits generation of simulated visual field series with chosen levels of fluctuation and progression (fig 10). The initial and final fields of the series are real measured clinical fields chosen and new intermediate fields are simulated by interpolation.
The central sense of algorithms is straightforward: for *simulation of damaged but stable* visual field test results, the initial and final fields are the same, and therefore no progression is present. 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 (fig 10). Otherwise, for *simulation of progression* the initial and final fields are different.

In the Gardiner’ “Virtual Eye” simulation program \(^{11-13}\) (fig 11) the response variability is calculated adding only one measure of noise at each test location directly related to its point wise sensibility. The noise is determined by independent random samples from a normal distribution formula. The pointwise response variability was represented by the function: 

\[
\log_e (SD) = A \times \text{sensitivity}(dB) + B,
\]

where the constants \(A\) and \(B\) are -0.081 and 3.27, respectively.
So, for example, if the true sensitivity at a point is 28dB, then the simulated sensitivity is drawn randomly from a normal distribution with mean 28dB and standard deviation 2.72dB. So only the short-term fluctuation is intended. Naturally, the Virtual Eye can be used to simulate an entire visual field or some clustered points rather than just one point. In the above figure we see a gray-scale simulation of three series over a six tears period of “follow-up”: noise-free progressing, noisy progressing and noisy stable.

Figure 11. Virtual Eye simulations. The figure shows three series of greyscales, representing annual field testing over a six year period. The first column is a noise-free series with six points progressing at 2dB/year, and all other points remaining stable. The second and third columns are produced by using the Virtual Eye simulation to add noise to each point in each field; it becomes very difficult to distinguish the series based on the eye with six progressing points (the middle column) from the series based on a completely stable eye (the right-hand column). (Gardiner, S.K 13)
Conclusions

Within the past 20-30 years, a large amount of information has been obtained concerning the glaucoma visual fields behavior. Nowadays, we are proud of new devices applying more or less complicated techniques resources on selected patients’ databases. But in reading over the content of this review one realizes that four quite immovable problems overshadow glaucoma evaluation as it relates to monitoring the rate of any change that has occurred in a visual fields series.

1) The tenacity of an unstable psychophysical behavior in signal and noise (variability).

2) Although statistical methods belong to the strict domain of the mathematics, in many instances the operators’ clinical criteria belong rather to the scope of the heuristic.

3) It may be that one day a test will be developed that can be seen as an independent gold standard. However, this future gold standard is not available for the glaucoma patients at hand.

4) Although predictive values could be more important in clinical practice, the parameters of accuracy (sensitivity and specificity) are much more popular between perimetry-operators.

abenju@us.es
REFERENCES


6) Chen PP: Correlation of visual field progression between eyes in patients with open-angle glaucoma. *Ophthalmology* **109**: 2093-9, 2002


17) Glynn RJ, Rosner B.: Multiple imputation to estimate the association between eyes in disease progression with interval-censored data. Statist. Med. 23;3307–3318, 2004


33) Jansonius, N.M. Bayes theorem applied to perimetric progression detection in glaucoma: from specificity to positive predictive value. *Graefe’s Arch Clin Exp Ophthalmol.* 243;433–437, 2005


49) Schulzer M. Errors in the diagnosis of visual field progression in normal-tension glaucoma., *Ophthalmology* **101**: 1589–1594, 1994


56) Tarek M, Spaeth GL, Bitterman A: Rate and amount of visual loss in 102 patients with open-angle glaucoma followed up for at least 15 years. *Ophthalmology.* **110**: 900–907, 2003


