British Journal of Ophthalmology 1996; 80: 304–308

Visual field test simulation and error in threshold estimation

S E Spenceley, D B Henson

Abstract

***Aim*—To establish, via computer simulation, the effects of patient response variability and staircase starting level upon the accuracy and repeatability of static full threshold visual field tests.**

***Method*—Patient response variability, defined by the standard deviation of the frequency of seeing versus stimulus intensity curve, is varied from 0·5 to 20 dB (in steps of 0·5 dB) with staircase starting levels ranging from 30 dB below to 30 dB above the patient's threshold (in steps of 10 dB). Fifty two threshold estimates are derived for each condition and the error of each estimate calculated (difference between the true threshold and the threshold estimate derived from the staircase procedure). The mean and standard deviation of the errors are then determined for each condition. The results from a simulated quadrantic defect (response variability set to typical values for a patient with glaucoma) are presented using two different algorithms. The first corresponds with that normally used when performing a full threshold examination while the second uses results from an earlier simulated full threshold examination for the staircase starting values.**

***Results*—The mean error in threshold estimates was found to be biased towards the staircase starting level. The extent of the bias was dependent upon patient response variability. The standard deviation of the error increased both with response variability and staircase starting level. With the routinely used full threshold strategy the quadrantic defect was found to have a large mean error in estimated threshold values and an increase in the standard deviation of the error along the edge of the defect. When results from an earlier full threshold test are used as staircase starting values this error and increased standard deviation largely disappeared.**

***Conclusion*—The staircase procedure widely used in threshold perimetry increased the error and the variability of threshold estimates along the edges of defects. Using earlier data, when available, overcomes this problem and reduces examination time.**

(*Br J Ophthalmol* 1996; **80:** 304–308)

School of Computer and Information Science, University of South Australia
S E Spenceley

Department of Ophthalmology, University of Manchester
D B Henson

Correspondence to:
Mr D B Henson, Department of Ophthalmology, University of Manchester, Royal Eye Hospital, Oxford Road, Manchester M13 9WH.

Accepted for publication 21 December 1995

Computer simulation of the visual field test makes it possible to investigate the interaction of patient characteristics and test strategies without the time and expense of an actual trial. The KRAKEN system[1 2] is one simulation procedure developed for use in the analysis of static full threshold perimetry. The system has been used to explore the effects of changing the variables of staircase algorithms such as the starting and final step sizes, the starting intensity, the number of reversals, and patient response variability.

The staircase strategy widely used to monitor glaucoma in programs such as the 30-2 and 24-2 on the Humphrey, Henson, and Octopus perimeters,[3] is a two reversal strategy with an initial step size of 4 dB changing to 2 dB after the first reversal. This strategy (4-2) is used in the current simulation with the last seen presentation being taken as the threshold estimate. The threshold estimate derived by the 4-2 strategy is subject to a certain degree of error which can be defined as the difference between the true threshold and the threshold estimated by the staircase.

A patient's sensitivity and response variability at a given retinal location are well described by the frequency of seeing (FOS) curve.[4–7] This S-shaped curve gives the percentage of stimuli seen at a range of intensities which straddle the threshold. The threshold sensitivity is normally defined as the intensity at which 50% of the stimuli will be seen while response variability is represented by the gradient or standard deviation of the FOS curve, steep gradients (small standard deviations) representing little variability and shallow gradients (large standard deviations) large amounts of variability.

Several researchers have investigated the gradient of the FOS curve and its relation to threshold deviation (the measured threshold minus that for an age matched normal eye)[8] in both normal and glaucomatous eyes.[4–7] Their results show a positive relation between the gradient and threshold deviation. They have also shown that variability is not constant within an eye and that the use of a global measure of fluctuation is inappropriate. In areas of the visual field where glaucomatous damage has already occurred the gradient of the FOS curve is often very shallow with standard deviations exceeding 10 dB. This value is much larger than that used in previous simulation studies.[1 2] The enhanced patient response variability found in the regions of a visual field defect[9–15] makes it difficult to differentiate between non-significant random variations in the visual field and true progression. It often requires several visual field results, taken over a period of time, before a decision on whether the visual field has progressed or not can be made.[16 17] Understanding the origins of the

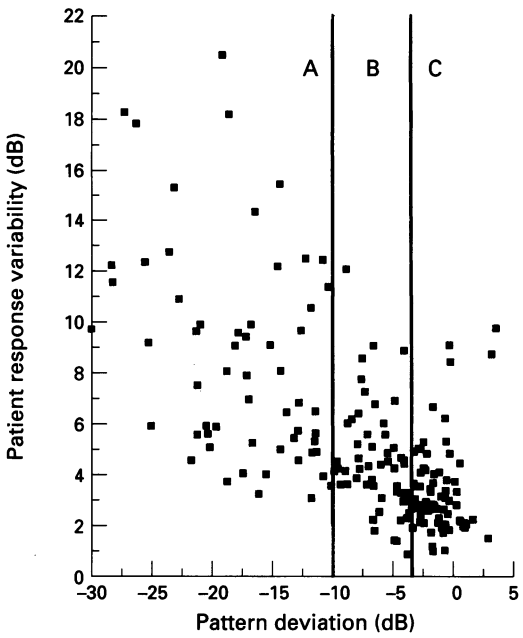

Figure 1 The relation between pattern deviation and patient response variability for 185 retinal locations in 60 glaucoma patients. Patient response variability is represented by the standard deviation of the frequency of seeing curve. The region marked A represents the 60 examples of most advanced loss, B the 60 examples of early loss, and C the 65 most normal results.

enhanced variability and, where possible, reducing its magnitude is an important objective in glaucoma research.

This paper seeks to determine how the error in the 4-2 strategy behaves when large degrees of patient response variability are present. It then presents an example which demonstrates that the error is larger and varies more at the edges of visual field defects. Finally, it demonstrates how the error is reduced, along with its variability, by using threshold data from an earlier examination as staircase starting values. This approach has the added advantage of reducing test time.

Method

Different combinations of patient response variability and staircase starting level are explored to determine the error in the threshold. Patient response variability, defined by the standard deviation of the FOS curve, is varied from 0·5 to 20 dB (in steps of 0·5 dB) with starting levels ranging from 30 dB above to 30 dB below the threshold (in steps of 10 dB). Each result is composed of 52 simulated threshold estimates from which the mean and standard deviations of the error are calculated.

The simulation is extended to a left superior quadrantic visual defect whose threshold deviation corresponds to −20 dB across the whole of the quadrant (see Fig 4). The response variability at each test location is set to match that found in patients with glaucomatous damage. This assignment used the FOS data collected by Chauhan *et al*[4 5] from 185 retinal locations in 60 glaucoma patients. The analysis of Chauhan's data began with the fitting of FOS curves to the data from each of the 185 retinal locations using the probit analysis of SPSS.[18]

Figure 1 plots response variability (standard deviation of the fitted FOS curves) against the pattern deviation for the 185 retinal locations tested by Chauhan *et al.*[4 5] Pattern deviation is similar to threshold deviation but includes a component which compensates for overall shifts in the eye's sensitivity.[8] The data given in Figure 1 show both an increase in magnitude and spread of response variability with severity of loss. The assignment of response variability within the stimulation was based on an empirical model of this distribution. This model isolated three ranges of pattern deviation values – advanced loss, intermediate loss, and no loss. The first range includes 60 retinal locations whose pattern deviations ranged from −29·92 to −9·57 dB, the second 60 retinal locations with pattern deviations ranging from −9·48 to −2·8 dB, the third 65 retinal locations with pattern deviations ranging from −2·69 to +3·58 dB. After first establishing which of the three ranges the pattern deviation fell within a value for response variability was assigned on the basis of the distribution within the response range.

After assigning a value for the response variability to each of the test locations the simulation conducted a 24-2 full threshold visual field test. The only difference between this program and that used to test patients is the replacement of the patient responses with those from the simulation algorithm. The simulation used random stimulus presentation of the test targets and started with four seed locations (one in each quadrant at a horizontal and vertical eccentricity of 9 degrees) with starting values set at 30 dB. In this simulation the normal setting for age was set to be 30 dB across the whole of the central field. The results from the seed locations were used to establish starting values for the neighbouring locations. The results at these secondary locations are then used as starting values for more neighbouring locations, and so on. Repeat estimates were made at 10 preselected locations (as per standard 24-2 program) which included the seed locations. When repeat estimates were available the results from the second estimate were used in the following analysis.

The whole process, assignment of response variability and simulated examination, was repeated 100 times. The mean and standard deviation of the threshold estimates were then calculated for all test locations.

To establish the effect of using earlier threshold data as the starting values for the 4-2 strategy the results from the 100 simulations were used as starting values for a further 100 simulated examinations. Each simulation took the estimated threshold at each test location of the corresponding previous simulation as a starting value rather than use 'normal age values' or neighbouring locations. The simulation assigned new levels of response variability to each location of each re-run. The mean and standard deviation of the threshold estimates were again calculated for all test locations.

Results

The average error in threshold determination for each of the seven starting levels versus patient response variability (standard deviation of the FOS curve) is given in Figure 2. The direction of the error is biased towards the starting level with its magnitude dependent

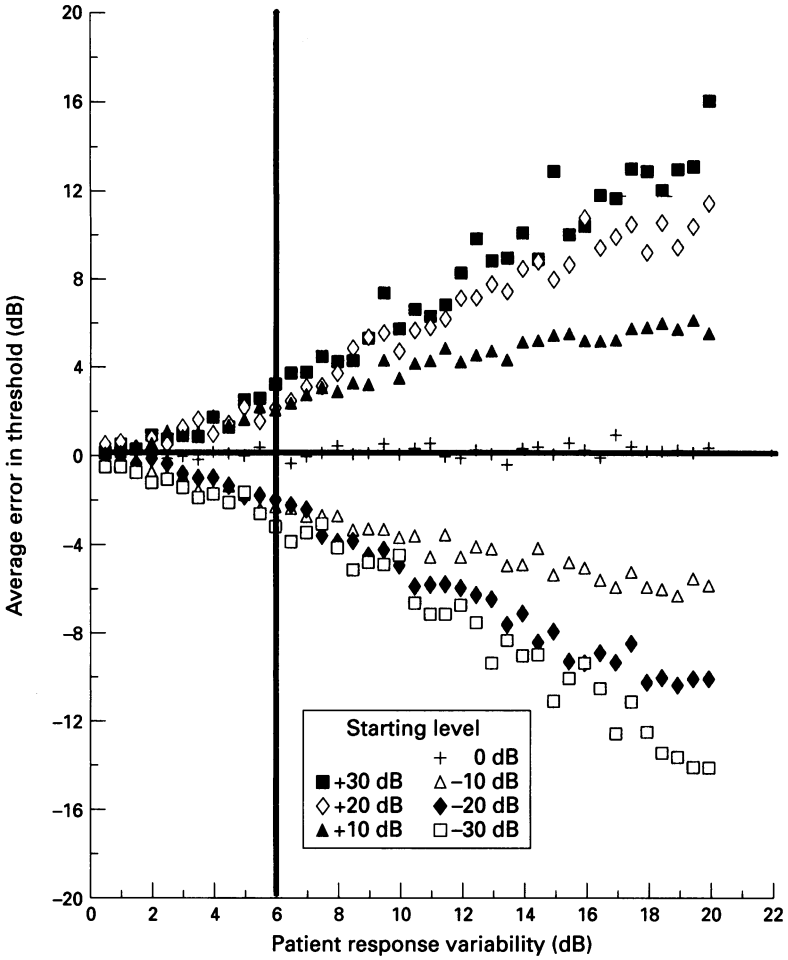

Figure 2 The relation between the average error (of 52 measures) in final threshold determination and patient response variability for seven different staircase starting levels. Patient response variability is represented by the standard deviation of the frequency of seeing curve. The vertical line separates small variabilities (<6 dB) from the larger ones.

upon patient response variability and the starting level. The average error is close to zero when the starting level of the staircase corresponds with the threshold.

The standard deviation of the error in threshold determination for each of the seven starting levels versus response variability is given in Figure 3. There is a steady rise in the standard deviation as response variability increases, for all categories. The effect of staircase starting level upon the standard deviation of the error becomes more important as the patient response variability increases. For situations where the response variability is fairly small (<6 dB) all staircase starting levels give similar results while beyond this level the standard deviation increases with the starting level (starting level being defined as the difference between the staircase starting intensity and the threshold).

The mean threshold estimates for the quadrantic defect using 'normal age values' and neighbourhood calculations are given in Figure 4B. In the normal area of the visual field the mean values are within a few tenths of a decibel of the true value. In the defective quadrant the mean error varies with location reaching values of between 2·2 and 3·8 dB towards the edge of the defect. In comparison, when the staircase starting value is set to equal the threshold estimate of a previous examination this increase towards the border of the defect largely disappears (see Fig 4C).

The standard deviation of the threshold estimates when using 'normal age values' and neighbourhood calculations are given in Figure 5A and the results from using earlier data in Figure 5B. In both these figures it can be seen that the standard deviation of the threshold estimates is larger within the defective quadrant than within the normal area of the visual field. This reflects the established relation between response variability and pattern deviation. In Figure 5A the standard deviation is greatest on the defective side of the defect edge having an average value of 5·2 dB. In comparison, when results from a previous threshold estimate are used as staircase starting values the increase in standard deviation along the edge of the defect largely disappears (average value 4·1 dB). The differences in the standard deviations between these two strategies is highlighted in Figure 5C where the results from Figure 5B have been subtracted from those in Figure 5A. Positive values indicate the standard deviation is smaller when using earlier threshold data as staircase starting values.

The use of earlier threshold data as staircase starting values for the 4-2 strategy also reduces the number of presentations needed to complete the examination. On average, the number of presentations reduced from 300 to 281, a small but significant ($p<0·001$) difference.

Discussion

One of the characteristics of glaucomatous visual field loss is an increase in the patient's response variability at and around the region of a visual field defect.[9–15] The data of Chauhan *et al*[1 2] have shown that standard deviations of the FOS curve can reach values as high as 20 dB and frequently reach levels of 10 dB (see Fig 1). In previously published simulations response variability has been confined to the levels found in patients with threshold deviations which fall within normal range (response variability <6 dB). Within this range the average bias in the threshold estimate is less than 2 dB and the variability of the threshold estimates is relatively independent of the staircase starting level (see Figs 2 and 3). When the range of patient response variability is extended to include those values likely to be seen in areas of glaucomatous visual field damaged there is not only a further increase in both the average bias and the standard deviation of the threshold estimates but also an increasingly important effect of staircase starting level.

The algorithm used in the full threshold strategy found within most computerised perimeters utilise, whenever possible, the threshold levels of neighbouring locations to decide upon suitable staircase starting values. These algorithms speed up the examination by reducing the number of presentations within any given examination.

At the edge of visual field defects algorithms that utilise neighbourhood averaging will generate staircase starting values which differ from the threshold by a larger amount than in other regions of the visual field. This will bias

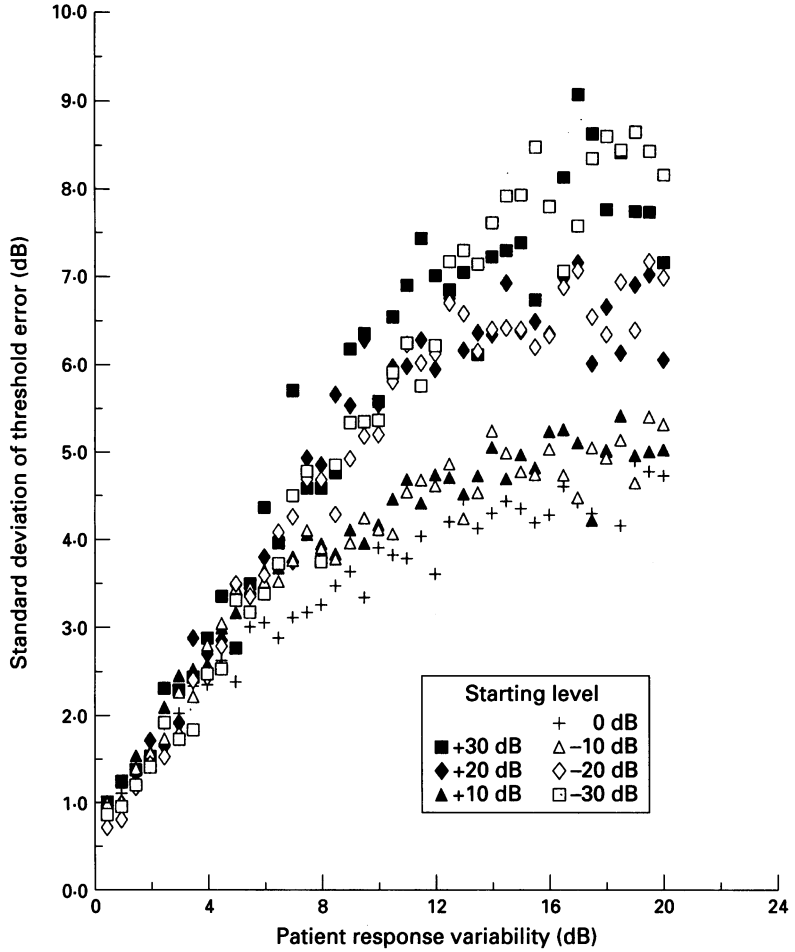

Figure 3 The relation between standard deviation of error in final threshold determination and patient response variability for seven different staircase starting levels. Patient response variability is represented by the standard deviation of the frequency of seeing curve.

the threshold estimates towards the average threshold of the neighbouring locations, it will raise the average threshold on the defective side of the defect border and lower the average threshold on the normal side of the border. The magnitude of the bias is dependent upon the response variability (see Fig 2). In eyes with glaucomatous damage, where the response variability is high in the region of a defect, the magnitude of this bias will be exaggerated. This predicted effect can be seen in the data of Figure 4B where the average threshold estimate on the defective side of the defect border is raised by an average of 2·8 dB.

On the 'normal' side of the defect border the predicted effect, a lowering of the average threshold, is not apparent (the average error for the nine locations bordering the defect is only −0·2 dB). There are two explanations for this; firstly, the response variability on the normal side of the border is lower than that on the defective side which, as can be seen from Figure 2, will reduce the bias for any given staircase starting level. Secondly, the simulation starts testing the four seed locations at the expected threshold (30 dB). While this coincides with the threshold in the normal quadrants it is 20 dB away from the threshold in the defective quadrant. To establish threshold estimates the simulation will, therefore, require more presentations at the seed location in the defective quadrant than it will in the other three quadrants. This in turn will result in the threshold estimates being established at the normal seed locations before the defective location and the test proceeding from the normal to the defective quadrant more often than from the defective to the normal. At the border of the defect the staircase starting level will, therefore, be closer to the normal value of 30 dB than it will to the defective value of 10 dB which can account for the absence of the predicted bias on the normal side of the defect border.

The relation between response variability and threshold deviation means that within the area of the defect there is an increase in the variability of threshold estimates (see Fig 5A). As discussed above, along the edges of the defect the influence of neighbourhood averaging will generate staircase starting levels which will differ from the thresholds by a larger extent than in other regions of the visual field. This in turn will increase the standard deviation of the threshold estimates on the defective side of the defect border. Such an effect can clearly be seen in Figure 5A where the average standard deviation of threshold estimates on the defective side of the defect border is 5·2 dB compared with 3·9 dB for all other defective locations.

Haefliger and Flammer[12 20] have already reported that patients with glaucoma give more variable results both at the edge of glaucomatous defects and in the region surrounding the blind spot. The results from this simulation indicate that some of this increase can be attributed to the threshold algorithm.

The increased variability seen at the edges of a defect results from having an algorithm which uses 'normal age values' and the thresholds of neighbouring test locations to calculate suitable staircase starting values. When the staircase starting value corresponds

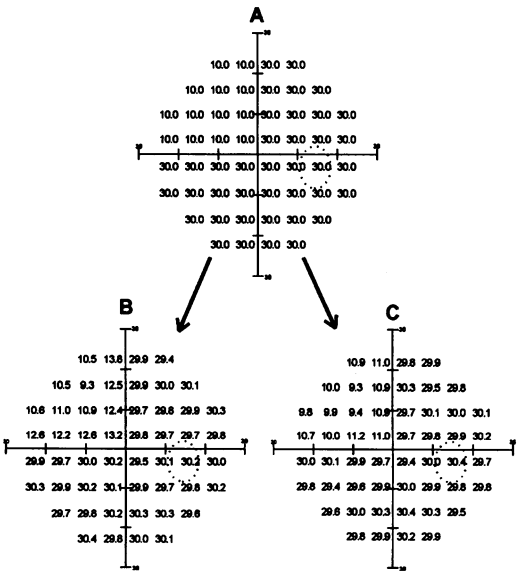

Figure 4 Visual field charts giving (A) the threshold values used in the simulations; (B) the mean threshold values from the 100 simulations using the testing algorithm which sets staircase starting values on the basis of 'normal age values' and the thresholds of neighbouring locations; (C) the mean threshold values from 100 simulations using threshold data from an earlier examination as staircase starting values.

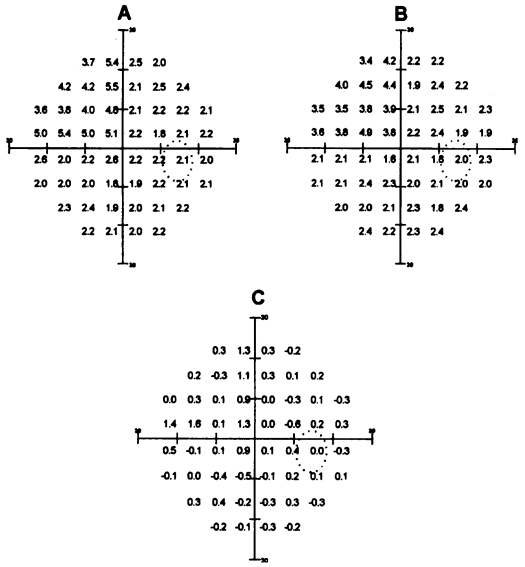

Figure 5 Visual field charts giving (A) the standard deviation of the threshold estimates from the 100 simulations using the testing algorithm which sets staircase starting values on the basis of 'normal age values' and the thresholds of neighbouring locations; (B) the standard deviation of the threshold estimates from the 100 simulations using threshold data from an earlier examination as staircase starting values; (C) the difference in the standard deviations of the threshold estimates (A) – (B).

to the threshold the average bias is zero irrespective of the response variability (see Fig 2). Given that many patients being seen in an ophthalmic department will have previous visual field records and that glaucomatous defects progress at a relatively slow rate[21] then starting the staircase off at values corresponding to the previous threshold estimates is likely to reduce the bias. Figure 4C shows that this is in fact the case with the bias largely disappearing when such a strategy is adopted. Figure 5B shows that in addition to the bias being reduced the standard deviation of the threshold estimates also reduces – that is, there will be less variability in the threshold estimates from one session to the next.

The absence of an increase in the variability of threshold estimates on the normal side of the defect border is again attributed to both the reduced response variability at these locations and the tendency for the simulation to proceed from the normal to the defective regions more often than from the defective to the normal.

Using earlier threshold data has the added advantage of reducing the number of presentations and thereby reducing examination time.

Conclusions

The results of these simulations show that the utilisation of previous threshold measures as starting values, such as the full threshold from an earlier data program on the Humphrey 600 range of perimeters, reduces the error and variability of the threshold estimates. It also has the added benefit of speeding up the examination. As the monitoring of glaucomatous loss is currently compromised by the large degree of variability seen in the results from full threshold perimetry the adoption of such a strategy would seem a simple and worthwhile option.

This work has been financially supported by the EPSRC (grant number GR/H54393) and made possible by the kind contribution of data from Balwantray C Chauhan at Dalhousie University.

1 Shapiro LR, Johnson CA, Kennedy RL. KRAKEN – a computer simulation procedure for static, kinetic, suprathreshold static and heuristic perimetry. In: Heijl A, ed. *Perimetry update 1988/89*. Amsterdam: Kugler & Ghedini, 1989: 431–8.
2 Johnson CA, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci* 1992; **33:** 2966–74.
3 Henson DB. *Visual fields*. Oxford: Oxford University Press, 1993.
4 Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency of seeing curves in glaucoma. In: Mills RP, ed. *Perimetry update 1992/93*. Amsterdam: Kugler & Ghedini, 1993: 543–9.
5 Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma and patients with glaucoma. *Invest Ophthalmol Vis Sci* 1993; **34:** 3534–40.
6 Olsson J, Heijl A, Bengtsson B, Rootzen H. Frequency-of-seeing in computerized perimetry. In: Mills RP, ed. *Perimetry update 1992/93*. Amsterdam: Kugler & Ghedini, 1993: 551–6.
7 Weber J, Rau S. The properties of perimetric thresholds in normal and glaucomatous eyes. *Germ J Ophthalmol* 1992; **1:** 79–85.
8 Heijl A, Lindgren G, Olsson J, Asman P. Visual field interpretation with empiric probability maps. *Arch Ophthalmol* 1989; **107:** 204–8.
9 Werner EB, Drance SM. Early visual field disturbances in glaucoma. *Arch Ophthalmol* 1977; **95:** 1173–5.
10 Donovan HC, Weale RA, Wheeler C. The perimeter as a monitor of glaucomatous damage. *Br J Ophthalmol* 1978; **62:** 705–8.
11 Holmin C, Krakau CET. Variability of glaucomatous visual field defects in computerised perimetry. *Graefes Arch Ophthalmol* 1979; **210:** 235–50.
12 Werner EB, Saheb N, Thomas D. Variability of static threshold responses in patients with elevated IOPs. *Arch Ophthalmol* 1982; **100:** 1627–31.
13 Flammer J, Drance SM, Zulauf M. Differential light threshold short and long term fluctuation in patients with glaucoma, normal controls and patients with suspected glaucoma. *Arch Ophthalmol* 1984; **102:** 704–6.
14 Langhorst CT, Van den Berg TJTP, Van Spronsen R, Greve EL. Results of a fluctuation analysis and defect volume program for automated static threshold perimetry with the scoperimeter. *Doc Ophthalmol Proc Ser* 1984; **42:** 1–6.
15 Werner EB, Bishop KI, Davis P, Krupin T, Petrig B, Sherman C. Visual field variability in stable glaucoma patients. *Doc Ophthalmol Proc Ser* 1986; **49:** 77–84.
16 Heijl AH, Lindren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol* 1989; **180:** 130–5.
17 Chauhan BC, Drance SM, Douglas GR. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci* 1990; **31:** 512–20.
18 *SPSS for Windows Advanced Statistics 6.1*. SPSS Inc, Michigan Avenue, Chicago, Illinois, 1993: 249–62.
19 Haefliger IO, Flammer J. Increase of the short-term fluctuation of the differential light threshold around a physiologic scotoma. *Am J Ophthalmol* 1989; **107:** 417–20.
20 Haefliger IO, Flammer J. Fluctuation of the differential light threshold at the border of absolute scotomas: comparison between glaucomatous visual field defects and blind spots. *Ophthalmology* 1991; **98:** 1529–32.
21 Hart WM, Becker B. The onset and evolution of glaucomatous visual field defects. *Ophthalmology* 1982; **89:** 268–79.