Confidence intervals for change in automated visual fields

H DUNBAR HOSKINS, SCOTT D MAGEE, MICHAEL V DRAKE, AND MARTIN N KIDD

From the Foundation for Glaucoma Research and University of California, San Francisco, USA

SUMMARY Three successive fields of 136 eyes (86 patients) were extracted from our data base containing over 3000 visual fields performed on the Humphrey visual field analyser with program 30-2. Series of fields in which the second field was depressed relative to the first were selected for analysis to determine how much change between the first two fields was required to predict a downward trend as confirmed by the third field in the series. The data were stratified with respect to initial field damage. Seven regions of the visual field were analysed including the upper and lower temporal and nasal quadrants, the superior and inferior Bjerrum regions, and the whole field. Minimally damaged regions required between 4·7 and 5·6 dB change in mean sensitivity, whereas more damaged regions required between 5·5 and 7·2 dB change in mean sensitivity to have 95% confidence that the negative trend would be confirmed by the third field. The superior Bjerrum region was the most sensitive for the detection of change, and the lower temporal region was the least sensitive. We conclude from this series of data that large changes between two successive fields are required to be sure that the changes are due to disease rather than chance fluctuation. Where any doubt exists, the field should be repeated to confirm the reality of change.

Analysing visual field progression is an important part of the management of patients with glaucoma. Unfortunately detecting progression reliably may be difficult. Factors such as short-term (intratest) and long-term (intertest) fluctuation, patient fatigue, patient learning curve, and attention span can cause transient visual field changes which may mimic improvement or deterioration of the overall visual field.[1-4] Over a series of tests these factors should cancel themselves out, that is, errors in the positive direction and errors in the negative direction should be approximately equal, and overall mean sensitivity should stay roughly the same. In practice one must often repeat a visual field test several times before one can determine whether or not an increased depression is due to transient fluctuation or pathological deterioration. However, multiple visual field examinations require considerable expenditure of the patient's time and effort and may unnecessarily delay diagnosis.

In an effort to improve our ability to recognise

Correspondence to H Dunbar Hoskins, MD, Foundation for Glaucoma Research, 490 Post Street, Suite 1042, San Francisco, CA 94102, USA.

significant visual field deterioration we analysed a series of three visual fields to assess the predictive value of the second field in determining the likelihood of a decrease in sensitivity between the first and third field.

Patients and methods

We analysed the field of 136 eyes (86 patients), each of whom had at least three central 30° full threshold fields performed on the Humphrey visual field analyser with program 30-2. The patients were between 21 and 79 years of age (mean 54·7, SD 14·6) All patients had a corrected visual acuity of 6/12 (20/40) or better and pupillary diameter greater than or equal to 3 mm at the time of each test. All patients had glaucoma, with some areas of visual field depression present. The interval between the first and third field was between 4·6 and 30·5 months (mean 16·5, SD 5·4).

For each individual field test the data were analysed by quadrant, superior and inferior Bjerrum area, and for the total field. Fig. 1 displays the regions analysed. The mean sensitivity (MS) in each region
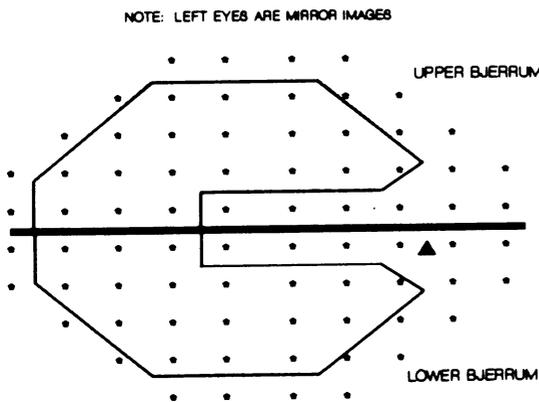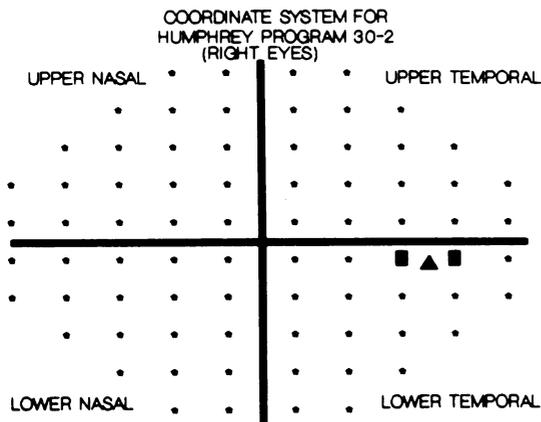
Fig. 1   *Co-ordinate system for the Humphrey visual field analyser program 30-2. Each spot represents a threshold sensitivity measured in dB. The two spots marked with a solid box in the lower temporal quadrant were removed from the statistical analysis. The triangle represents the approximate location of the blind spot. Points selected for the upper and lower Bjerrum regions are shown in the lower part of the figure.*

was determined by adding the individual threshold values in dB and dividing by the number of values. The two points closest to the blindspot in the lower temporal region were eliminated from the analysis.

Each series of fields was coded for each region as to possible relationships between fields 1, 2, and 3. Fig. 2 shows the possible arrangements of field scores. For the analysis we considered only series where a respective region of the second field was depressed relative to the first field. After this selection 79 eyes were available for analysis.

Regression models were constructed to determine

what regional change between the first and second field was required to predict that the third field would show a depression relative to the first field. Depression of the third field was defined as the MS of field 3 being 1 dB below that of field 1. Each region must have had a drop in total sensitivity of 1 dB times the number of spots in that region. This represented a 19 dB total drop for the quadrant regions or a 74 dB drop for the entire field.

The data were stratified in accordance with degree of field damage at the first field. Each region with an MS of greater than 25 dB was placed in one group labelled 'minimal field damage', and regions with an MS of less than or equal to 25 dB were placed in another group labelled 'moderate field damage'.

Results

Table 1 displays the mean sensitivities and standard deviations for each of the regions where the second field was depressed relative to the first.

Tables 2 and 3 report the results for the regression models studied to determine the amount of change required between the first and second fields to predict

Table 1   *Summary of three fields for each region*

Quadrant	First		Second		Third	
	Mean	SD	Mean	SD	Mean	SD
UN	21·1	4·1	18·7	4·5	19·2	4·8
UT	22·3	5·2	19·8	5·9	20·7	5·7
LN	24·2	4·0	21·8	4·9	22·2	4·9
LT*	26·0	3·2	23·9	4·1	24·6	4·0
Total	22·6	4·9	20·6	5·4	20·9	5·4
SB	21·4	3·9	19·1	3·8	19·6	4·2
IB	24·5	3·6	22·5	4·0	23·1	4·2

*With blind spot points removed. UN=upper nasal. UT=upper temporal. LN=lower nasal. LT=lower temporal. SB=superior Bjerrum. IB=inferior Bjerrum.

Table 2   *Model: Drop MS field 1 to 2 versus change MS field 1 to 3. Minimally damaged fields*

Region	R^2	Slope	Intercept	p
UN	0·05	0·17	1·88	0·1772
UT	0·22	0·27	1·60	0·0056
LN	0·14	0·22	1·66	0·0145
LT	0·05	0·12	1·75	0·1196
Total	0·05	0·08	1·71	0·1941
SB	0·07	0·19	1·98	0·1172
IB	0·16	0·29	1·37	0·0035

For abbreviations see Table 1.
   Regression coefficients as predicted from each model: R^2=R^2 value. Slope=slope of best fit line through points. Intercept= intercept of best fit line on the axis corresponding to the dependent variable. P=Probability that the slope is not equal to 0.

Table 3   *Model: Drop MS field 1 to 2 versus change MS field 1 to 3. Moderately damaged fields*

Region	R^2	Slope	Intercept	p
UN	0·30	0·34	1·85	0·0001
UT	0·22	0·33	2·40	0·0028
LN	0·23	0·30	2·43	0·0106
LT	0·05	0·10	2·32	0·3539
Total	0·12	0·22	1·86	0·0276
SB	0·30	0·35	1·67	0·0006
IB	0·16	0·21	2·09	0·0335

For abbreviations see Table 1. See footnote to Table 2.

Table 4   *Required change field 1 to 2 in dB*

Region	Degree of damage	
	Minimal	*Moderate*
Upper nasal	*	6·50
Upper temporal	4·74	6·91
Lower nasal	5·22	7·20
Lower temporal	*	*
Total field	*	5·49
Superior Bjerrum	*	5·76
Inferior Bjerrum	7·85	5·55

*No significant regression model found for these regions.
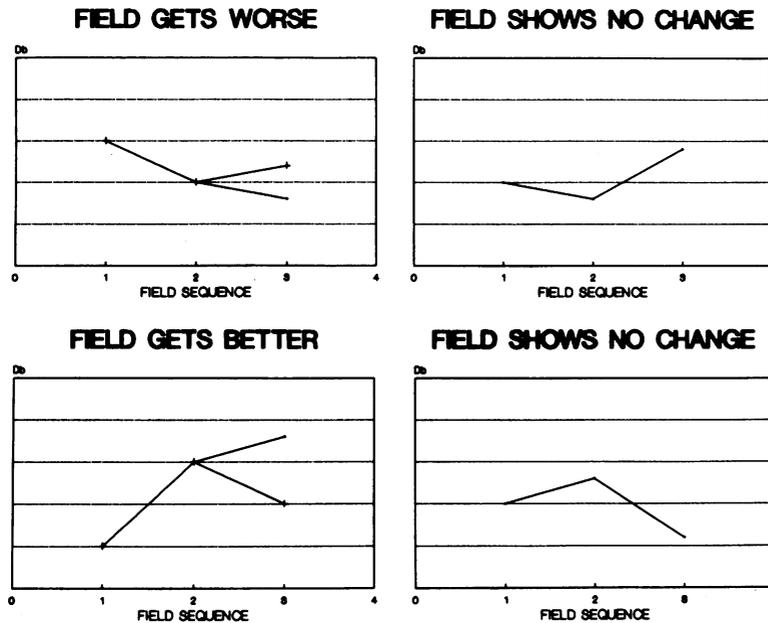

Fig. 2   *Possible patterns of response for each patient's series of three consecutive fields. A series was considered worse if the third field was at least 1 dB below the first field, or the slope of the best fit line through the three fields was negative.*
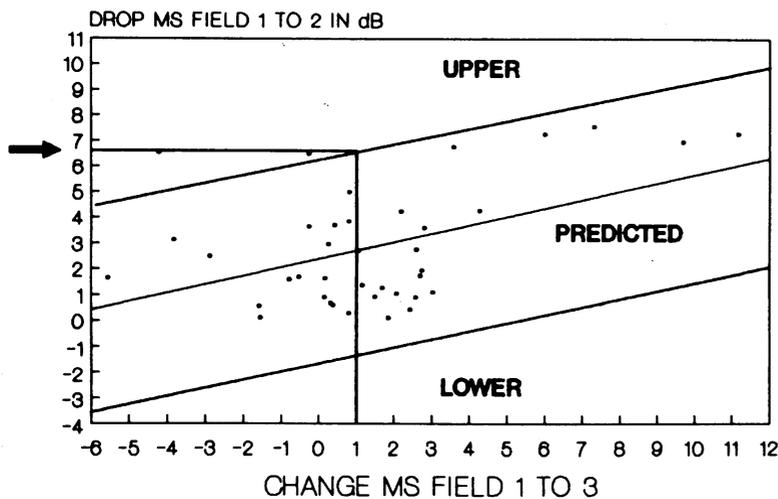

Fig. 3   *Graphical representation of the method used to determine the 95% confidence level for the amount of change required between the first two fields which would predict that the third field would confirm a downward trend in the series.*

a downward trend in the series as confirmed by the third field. The regression equations were solved for the dependent variable (drop required between fields 1 and 2) by means of the change in MS between field 1 and field 3 as the independent variable. This value plus the product of the standard error of the model and the t statistic, adjusted for degrees of freedom and desired confidence interval, yields the upper confidence level for the point where the patient's field is confirmed as getting worse.

Fig. 3 graphically displays the 95% confidence level for change from field 1 to field 2 required to predict that field 3 will be depressed by at least 1 dB relative to field 1 in the upper temporal region of moderately damaged fields. Table 4 reports the amount of change in MS required between fields 1 and 2 to predict with 95% confidence that field 3 will be below field 1 by at least 1 dB for each' of the regions studied for both minimally and moderately damaged fields.

Figs. 4 and 5 demonstrate plots of required change from fields 1 to 2 versus varying confidence intervals for confirmatory change in field 3 in minimally and moderately damaged fields.

Fig. 4

Fig. 5

Figs. 4 and 5   Plots of drop in MS required to predict a downward trend in the series as confirmed by the third field versus confidence intervals for this confirmation. Fig. 4 shows required change in minimally damaged fields. Fig. 5 shows required change for moderate to severely damaged fields.
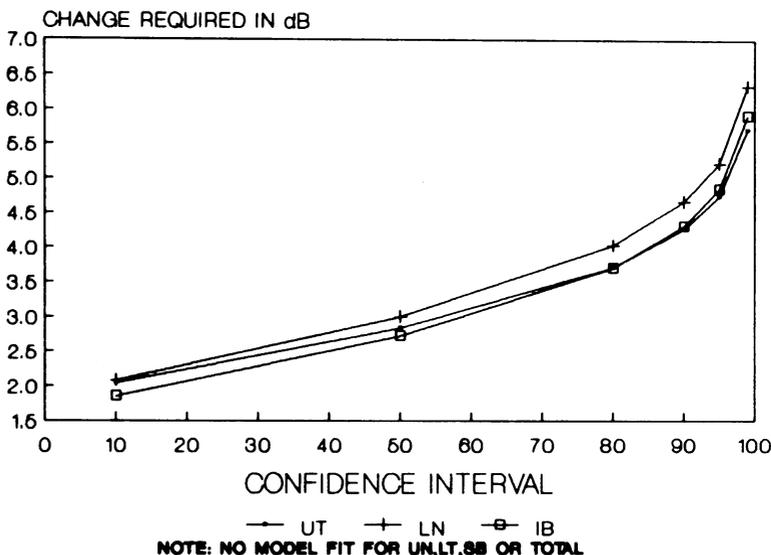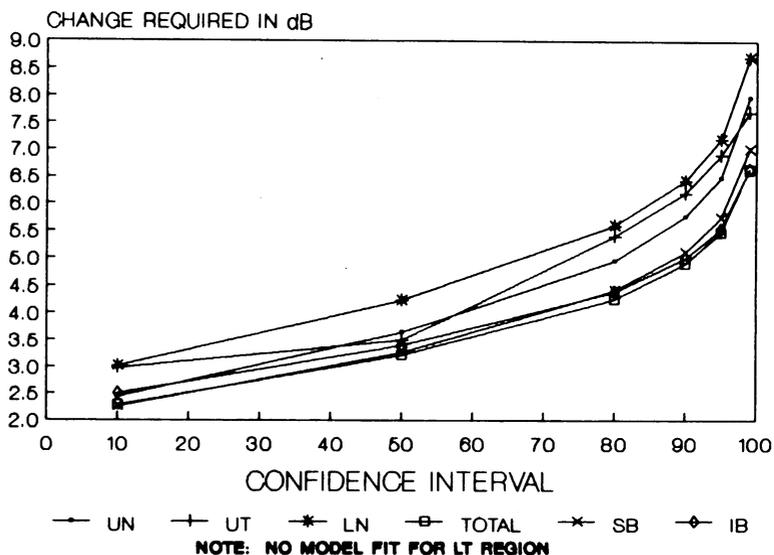
Discussion

The visual field examination is but one of a series of tests and evaluations that ophthalmologists use to determine whether or not a patient's glaucoma is progressing. Our data indicate that, for eyes with visual acuity better than or equal to 20/40 and pupil size greater than or equal to 3 mm, a decrease in mean sensitivity of between 4 and 7 dB (depending on the region analysed) is required to predict with great reliability a continuing downward trend in sensitivity. The converse is also true: as the change from field 1 to 2 becomes smaller, the probability of a continuing downward trend is less. An average loss of 3·2 dB in moderately to severely damaged fields corresponds to the 50% confidence interval. In other words, when the average loss per point in a given region between the first two fields in our series is 3·2 dB, 50% of the time the third field will confirm the downward trend.

It should be noted that no statistically significant regression model could be constructed for the lower temporal region, even when the two points closest to the blind spot were removed from the analysis. This is due to the fact that there was very little deterioration between the successive fields in this region.

What is perhaps most striking about our findings is that the degree of depression between the first and second fields must be quite extensive before the confidence level is high enough for the third field to confirm depression. If on the Humphrey field analyser a normal field might have an MS of 30 dB in a given quadrant, then approximately 20% (6 dB) loss of this sensitivity must occur to have a high predictive value.

Large regions, namely, the entire field, require less change per spot, but more total loss, than smaller regions or individual points. In the analysis using moderately damaged fields the upper temporal quadrant required a 6·9 dB change per point, whereas the total field required only a 5·49 dB change per point to be 95% sure that a third field would confirm this downward trend.

Less total change is required by region than by total field. From the above example the upper temporal quadrant requires a total change of 6·9×19 points, or 131 dB, whereas the total field would require 5·49×74 points (excluding the two points nearest the blind spot), or 406 dB. Therefore analysing individual regions for change is more sensitive than looking at the entire field.

More importantly, regional analysis may detect change that is not evident from full field analysis, as demonstrated by the fields in Fig. 6. In this series the drop in total mean sensitivity between fields 1 and 2 is 3·5 dB. Field 3 is depressed relative to field 1 by 3·1 dB indicating the 40–50% confidence level. However, the lower Bjerrum region is depressed by 7·8 dB, which represents the 99% confidence interval from our models. Field 3 confirms the change for this region.

Change in mean sensitivity, even though large, may be due to fluctuation rather than pathological change. Figs. 7 and 8 demonstrate grey scales of two series of visual fields where the second field is depressed relative to the first by 3·95 dB (70% confidence) and 5·25 dB (92% confidence) respectively. The third field in both these series failed to confirm this downward trend.

Regional analysis can confirm field deterioration suspected from looking at the entire field. Fig. 9 shows a series of fields where the MS of the second field is depressed relative to the first by 5·22 dB (92% confidence). In this series the lower temporal quadrant is depressed by 6·74 dB in the second field, corresponding to the 99% confidence interval. This region confirms the loss suspected in the total field analysis at a very high confidence level.
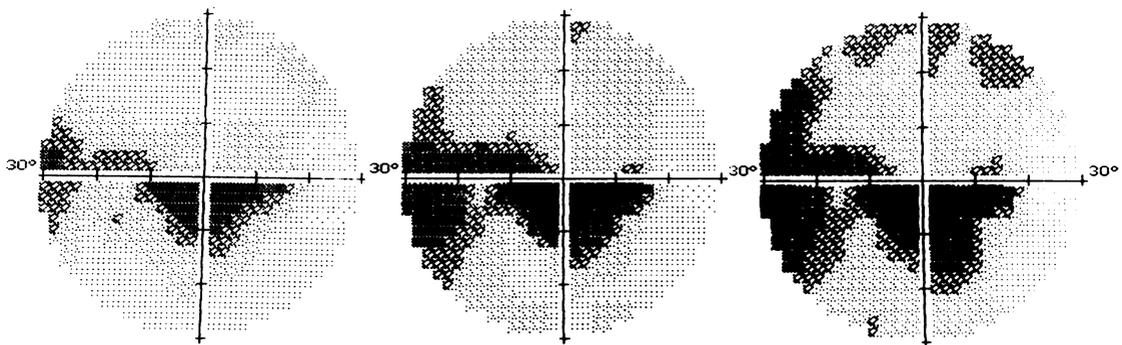

Fig. 6 *Grey scales of three sequential visual fields performed on a 58-year-old male with glaucoma. There is a 3·5 dB change in total mean sensitivity between fields 1 and 2. The lower Bjerrum region shows a change in mean sensitivity of 7·8 dB which corresponds to the 97% confidence interval and is confirmed by the third field.*

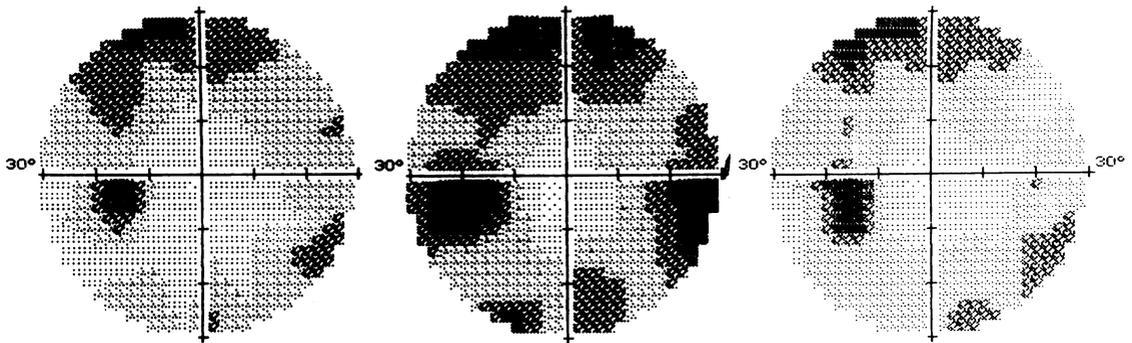*H Dunbar Hoskins, Scott D Magee, Michael V Drake, and Martin N Kidd*

Fig. 7   *Grey scales of three sequential visual fields performed on a patient with primary open angle glaucoma. The mean sensitivity of the second field is depressed 3·95 dB relative to the first. The third field, performed four months later, has returned to nearly the same mean sensitivity as the first.*
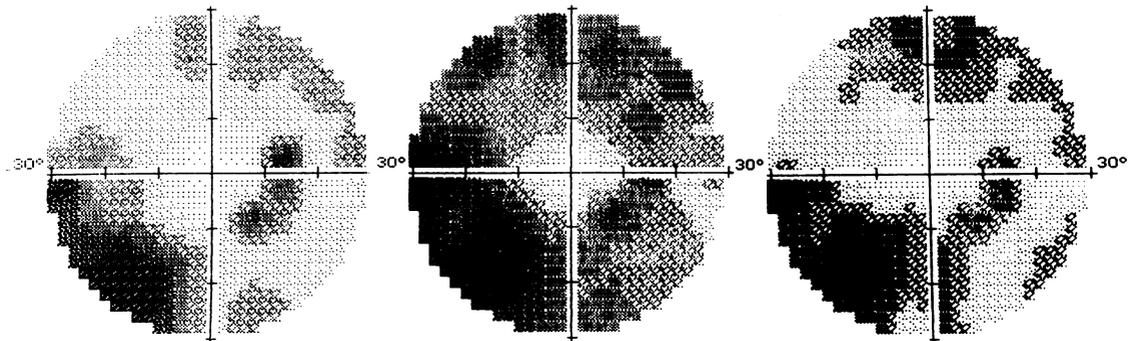
Fig. 8   *Grey scales of three sequential visual fields performed on a 61-year-old patient with primary open angle glaucoma. The second field is depressed 5·25 dB relative to the first, but the third field has returned to a mean sensitivity 0·82 dB greater than the first.*
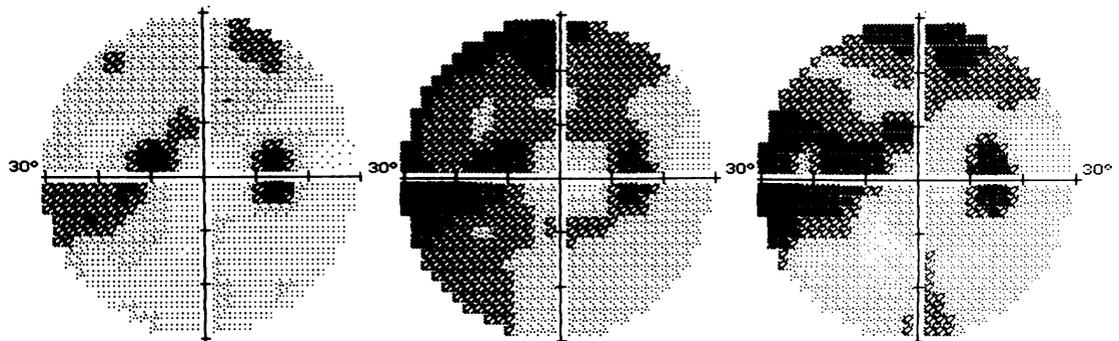
Fig. 9   *Grey scales of three sequential visual fields performed on a 76-year-old male with glaucoma. The 5·22 dB drop in mean sensitivity was confirmed with a third field. The upper and lower temporal quadrants changed in mean sensitivity by 6·26 and 6·74 dB respectively, corresponding to 95% confidence intervals as predicted by our models for moderately damaged fields.*

Werner *et al.*[5] presented data suggesting that long-term fluctuation at any single point could exceed 35%. Thus a moderately damaged point corresponding to 23 dB could fluctuate 8 or more dB between subsequent tests. It was also shown that fluctuation increases with eccentricity and with amount of damage.[6][7] Our data confirm these findings. We also recognised that greater change is required to reach

the 95% confidence level in more severely damaged visual fields (Table 4).

Greater numbers of field examinations will lend themselves to more precise statistical analysis. However, time, cost, and the patient's endurance limit the numbers of fields that can reasonably be performed. We have examined the methodology followed by many practitioners, that is, the sequential comparison of routinely collected visual fields. We are not advocating this as the best use of computerised perimetry, but rather are indicating the magnitude of change required if the practitioner wishes to use this technique.

Arguments can be made that two or three baseline fields performed within a short period of time would provide a better starting point for recognition of change. This may be so but remains to be proved. Until then practitioners need to recognise the potential fluctuation in visual fields over time.

In practice one analyses multiple factors when making therapeutic decisions about glaucoma patients. The most prominent of these factors include the patient's intraocular pressure, the cup to disc ratio, the presence or absence of optic disc haemorrhage, and visual field deterioration. Unfortunately, data on the predictive validity of these factors are incomplete.

If doubt exists about the results of a particular field test, it is wise to repeat the test reasonably soon to obtain confirmation.

The visual field examination measures a biological function, and fluctuation is to be expected in the results. It is our goal to create an awareness of the magnitude of this fluctuation as it occurs in an actual clinical setting.

References

1 Flammer J, Drance SM, Augustiny L, Funkhouser A. Quantification of glaucomatous visual field defects with automated perimetry. *Invest Ophthalmol Vis Sci* 1985; **26:** 176–81.
2 Flammer J, Drance SM, Zulauf M. Differential light threshold: short- and long-term fluctuations in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol* 1984; **102:** 704–6.
3 Flammer J, Zulauf M. The frequency distributions of the deviations in static perimetry. In: Heijl A, Greve EL, eds. *Proceedings of the 6th international visual field symposium.* Dordnect: Dr W Junk, 1985.
4 Rabineau PA, Gloon BD, Tobler HJ. Fluctuations in threshold and effects of fatigue in automated static perimetry. In: Heijl A, Greve EL, eds. *Proceedings of the 6th international visual field symposium.* Dordnect: Dr W Junk, 1985.
5 Werner EB, Bishop KI, Sherman C. Visual field variability in stable glaucoma patients. Presented at the 7th International Visual Field Symposium, Amsterdam, September 1986.
6 Heijl A, Lindgren G, Olsson J. Variability of computerized threshold measurements across the central field in a normal population. Presented at the 7th International Visual Field Symposium, Amsterdam, September 1986.
7 Langerhorst CT, van den Berg TJLP, Greve EL. Shortterm and longterm fluctuation of thresholds in automated perimetry in normals, ocular hypertensives, and glaucoma patients. Presented at the 7th International Visual Field Symposium, Amsterdam, September 1986.