



Testing for glaucoma with the spatial frequency doubling illusion

Ted Maddess^{a,*}, Ivan Goldberg^c, Jeffrey Dobinson^c, Stephen Wine^c, Alan H. Welsh^b,
Andrew C. James^a

^a Centre for Visual Sciences, Research School of Biological Science, Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

^b Centre for Mathematics and its Applications, Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

^c Prince of Wales Hospital, Sydney, Australia

Received 25 September 1998; received in revised form 17 May 1999

Abstract

We examined the performance of tests for glaucoma based on the spatial frequency doubling (FD) illusion. Contrast thresholds for seeing the FD illusion in four large visual field regions were measured from 340 subjects who were tested up to seven times over 2 years. Median sensitivities of 91% at specificities of 95% were obtained. Test–retest variability for the worst hemifield thresholds averaged $2.22 \text{ db} \pm 0.09 \text{ S.E.}$ for all tested groups, and significant progression was observed for glaucoma suspects over the seven visits, indicating that tests based on the FD illusion can detect diffuse early glaucomatous loss. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Glaucoma; Frequency-doubling; Illusion; Threshold; Hemifield

1. Introduction

Maddess and Henry (1990, 1992) presented data suggesting that the frequency doubling (FD) illusion may have special value for the diagnosis of glaucoma (see also Maddess, 1991, 1995). The FD effect, in which sinusoidal grating patterns presented at high temporal and low spatial frequency are seen to have twice their original spatial frequency, was originally described by Kelly (1966) (and see Tyler, 1974). Its monocular source (Tyler, 1974) suggested a retinal origin. The original concept behind the use of FD for glaucoma (Maddess & Henry, 1990) was that it might be produced by the Y-like retinal ganglion cells of the magnocellular visual pathway (M_y -cells). The illusion is produced by rectification (Tyler, 1974; Kelly, 1981) like that found in Y-cells (Victor & Shapley, 1979; Victor, 1988). Also, under the conditions where FD is seen, retinal gain control can make the nonlinear response component of Y-cells up to ten times larger

than their linear response component (Victor & Shapley, 1979). Whatever produces FD appears to have the same retinal density as M_y -cells (Maddess & Henry, 1990; Maddess, Hemmi & James, 1998). Correlative psychophysical and PERG studies support a Y-cell origin for FD (James, Maddess, Rouhan, Bedford & Snowball, 1995; Bedford, Maddess, Rose & James, 1997; Maddess, Bedford, James & Rose, 1997; Maddess & Kulikowski, 1999).

If M_y -cells were involved they would naturally have advantages as a litmus for glaucoma diagnosis because they are the largest cells in the M-pathway and they have a low retinal coverage factor (Maddess & Henry, 1990; Maddess et al., 1998). While there is good evidence that large ganglion cells are selectively lost in glaucoma (Quigley, Sanchez, Dunkelberger, L'Hernault & Baginski, 1987; Quigley, Dunkelberger & Green, 1988, 1989; Glovinsky, Quigley & Dunkelberger, 1991; Dawson, Brooks, Hope, Samuelson, Sherwood, Engel & Kessler, 1993; Glovinsky, Quigley & Pease, 1993; Smith, Chino, Harwerth, Ridder, Crawford & DeSantis, 1993) it is probably the low coverage factor of M_y -cells (Maddess et al., 1998) which would make detection of the loss of one of

* Corresponding author. Tel.: + 61-2-6249-4099; fax: + 61-2-6249-3808.

E-mail address: ted.maddess@anu.edu.au (T. Maddess)

their fellows relatively easy. The low spatial frequencies of FD stimuli also mean that test subjects need not be well refracted.

Anatomical (Quigley, Addicks & Green, 1982; Airaksinen, Lukowski, Drance & Price, 1986; Schulzer, Mikelberg & Drance, 1987; Quigley et al., 1989; Glovinsky et al., 1993) and ERG studies (Bach, Pfeiffer & Birkner-Binder, 1992; Falsini, Colotto, Porciatti & Porrelo, 1992; James & Maddess, 1996; Maddess, James, Goldberg, Wine & Dobinson, 1999) suggest that there is considerable diffuse ganglion cell loss even in the central retina. On the other hand diffuse loss is not a feature of early visual field changes assessed by conventional automated perimetry (Heijl, Lindgren & Lindgren, 1989; Lachenmayr, Drance, Chauhan, House & Lalani, 1991; Asman & Heijl, 1994). It would be interesting if an alternative diagnostic methodology could identify a form of diffuse loss associated with early glaucomatous damage. The present work therefore seeks to examine whether good diagnostic power can be obtained by testing relatively few, large, portions

of the visual field with FD stimuli and whether progression can be observed in glaucoma suspects. Some of this material has been reported previously at ARVO meetings (Maddess, Goldberg, Dobinson, Wine & James, 1995; Maddess, Goldberg, Dobinson, Wine, Welch & James, 1998).

2. Methods

2.1. Subjects

An initial cohort of 330 subjects were tested up to seven times on visits spaced at intervals of 3–4 months. Table 1 summarises: subject attendance, age, median Humphrey Field Analyser (HFA) MD and PSD for each subject group and their standard deviations, disc appearance, and refraction for the groups of primary interest. Refraction is expressed as spherical correction + half the cylindrical correction. Vertical cup to disc ratios, by contour and by pallor, were estimated by

Table 1
Subject data

Visit	1	2	3	4	5	6	7
<i>Age (years ± S.D.) and N subjects on each visit</i>							
Normal	55.1 ± 13.2 (44)	55.5 ± 13.3 (43)	54.9 ± 13.5 (37)	55.4 ± 13.4 (36)	56.5 ± 13.9 (28)	57.8 ± 15.0 (16)	51.5 ± 26.2 (2)
Weak	57.7 ± 11.0 (134)	58.8 ± 10.6 (126)	59.1 ± 10.5 (108)	60.1 ± 10.3 (94)	59.4 ± 10.4 (85)	60.2 ± 10.8 (56)	58.1 ± 10.8 (23)
Moderate	57.9 ± 13.7 (32)	57.9 ± 13.8 (31)	61.0 ± 12.8 (26)	63.8 ± 10.4 (21)	61.9 ± 11.4 (24)	62.1 ± 12.8 (18)	63.1 ± 11.6 (7)
Strong	60.4 ± 10.2 (48)	60.4 ± 10.2 (50)	61.5 ± 10.0 (47)	63.0 ± 8.4 (42)	62.6 ± 9.2 (42)	63.0 ± 8.8 (36)	64.6 ± 5.8 (13)
Glaucoma	62.6 ± 9.9 (70)	63.3 ± 9.5 (64)	63.1 ± 9.7 (63)	64.6 ± 9.3 (48)	63.9 ± 9.9 (45)	63.9 ± 9.3 (33)	64.8 ± 7.6 (14)
Total	58.9 ± 11.5 (330)	59.4 ± 11.3 (314)	60.0 ± 11.1 (281)	61.1 ± 10.7 (241)	60.8 ± 10.9 (224)	61.6 ± 10.9 (159)	61.5 ± 10.1 (59)
<i>HFA mean defect, PSD ± S.D. (decilogs)</i>							
	MD		PSD				
	Left	Right	Left	Right			
Normals	-2.17 ± 1.81	-1.87 ± 2.33	2.09 ± 1.41	1.97 ± 1.86			
Weak	-1.80 ± 2.05	-1.64 ± 1.93	1.92 ± 1.27	1.95 ± 1.08			
Moderate	-2.46 ± 1.87	-2.07 ± 2.82	2.29 ± 1.27	2.15 ± 1.58			
Strong	-2.06 ± 2.61	-2.40 ± 2.12	1.96 ± 1.47	2.02 ± 1.46			
Glaucoma	-6.66 ± 5.46	-6.79 ± 6.30	6.10 ± 4.11	7.13 ± 4.13			
<i>Vertical cup to disc ratios by contour and colour</i>							
	Mean	S.D.	Median	Min	Max		
<i>Contour</i>							
Normal	0.51	0.16	0.50	0.10	0.80		
Strong	0.64	0.15	0.70	0.30	0.90		
Glaucoma	0.71	0.18	0.80	0.20	0.95		
<i>Colour</i>							
Normal	0.43	0.16	0.40	0.00	0.60		
Strong	0.58	0.17	0.60	0.30	0.90		
Glaucoma	0.67	0.22	0.70	0.10	0.95		
<i>Refraction (dioptres)</i>							
Contour	Mean	S.D.	Median	Min	Max		
Normal	0.69	0.24	0.65	-4.25	6.50		
Strong	0.47	2.29	0.13	-7.50	5.63		
Glaucoma	0.15	2.82	0.25	-7.00	6.00		

Table 2
Eye diagnostic categories

Category	Criteria
Normal	IOP \leq 20 mmHg and normal HFA fields and normal discs o.u.
Weak	Fellow eye abnormal or IOP $>$ 20 mmHg or pseudoexfoliation syndrome (PXS) or pigment dispersion syndrome (PDS) or narrow angles or disc only abnormal
Moderate	IOP $>$ 20 mmHg and disc abnormal or IOP $>$ 20 mmHg and PXS or IOP $>$ 20 mmHg and PDS or IOP $>$ 20 mmHg and narrow angles
Strong	as for moderate and observed disc change but no scotoma on HFA 24-2
Glaucoma	as for moderate and scotoma

IG who is experienced with both methods. Disc condition was assessed from disc photography using Kodochrome film. We also recorded subject age, sex (191 females, 139 males), family history of glaucoma, corrected visual acuity and their pupil diameter during each test. HFA 24-2 automated perimetry was conducted on all subjects. Intraocular pressure (IOP) was measured on each visit by applanation tonometry. The highest IOP for a subject, typically occurring some years prior to our experiments, was also recorded. Normal subjects also had a complete eye examination including slit lamp investigation, HFA perimetry and fundus photography.

Patient diagnoses were taken to be the worst of the diagnoses of their two eyes, hence if one eye was glaucomatous the subject was deemed to have glaucoma and so on. *Normal subjects were normal in both eyes.* Eyes were considered to have glaucoma if they had a reproducible glaucomatous scotoma on repeated examination with HFA 24-2 perimetry. Glaucomatous eyes also had at least two other indications of glaucoma including, ocular hypertension, vertical disc cupping or asymmetry. In addition to eyes with distinct scotomas there were 75 eyes in the study having two or more other clinical signs of glaucoma and which, over a period of several years before the study, had shown a disc change, as demonstrated by repeated fundus photography, but which, on repeated HFA perimetry showed no scotoma. We will refer to these eyes as strong suspects. Note that by the rule mentioned above a *patient* with one strong suspect eye and a less affected eye would be classed as a strong suspect *patient*. We classified other eyes as either weak or moderate suspects. Moderate suspects had two or more indicators of glaucoma, IOP $>$ 20 mmHg and one other sign of

disease (see Table 2) but no observed disc change. Weak suspects had one indicator of disease such as elevated IOP (Table 2) or were eyes that were apparently normal but whose fellow eye was abnormal in any way. The diagnostic categories for eyes are summarised in Table 2.

The suspects and glaucomas were divided into *open angle* types, comprising primary open angle glaucomas and a few normal tension glaucomas, and *blockage mechanism* types, where there was potential or actual mechanical blockage of the outflow of aqueous humour as in: pseudoexfoliation syndrome (PXS), pigment dispersion syndrome (PDS), and anatomically narrow angles. IG made the initial diagnoses and these were checked by either SW or JD. The long-term nature of the study meant that we could not deny treatment or changes in treatment deemed necessary by the patients' physician. Of the initial 99 glaucomatous eyes, 90 were receiving topical beta-blockers, 53 were on miotics, and one received Propine only. Of the 53 eyes on miotics all but two were receiving beta-blockers as well. Laser trabeculoplasty had been administered in 35 eyes.

The research presented followed the tenets of the Declaration of Helsinki. Informed written consent was obtained from the subjects after explanation of the nature and possible consequences of the study were explained to them. The research was approved by the ANU's Human Experimentation Ethics Committee under protocol M 881.

2.2. Visual stimuli

Test stimuli were presented to four regions of the visual field (Fig. 1). The circular region occupying the central 10° of the visual field was filled with a 0.5 cycle per degree (cpd) grating, all other regions being filled with 0.25 cpd sinusoidal grating patterns, these being near optimal (Maddess et al., 1998; Maddess & Severt, 1999a,b). The achromatic stimuli (6500 K) were presented on a Barco Calibrator monitor (mean luminance 45 cd m²). The video frame rate was 101.5 Hz (non-interlaced). Grating patterns contrast was reversed every two video frames, providing square-wave modulation of the grating contrast at 25.38 Hz. This rate was selected as being well within the FD range and permitting thresholds where 8 bit brightness resolution would suffice (Maddess & Henry, 1992).

The experiments required subjects to find a contrast threshold for seeing the FD illusion by a method of adjustment (MOA). Stimuli were initially presented at 100% contrast on every trial so that subjects could see what the FD stimulus looked like (a very few subjects could not see this with one or the other eye). The subjects were then instructed to use a track ball whose function had been demonstrated to them to 'make the pattern fainter and fainter until you no longer see the pattern. When you can no longer see the pattern you

should roll the ball to make the pattern more vivid. Then you should make careful adjustments of the ball until you find the point where the striped pattern is only just visible'. The instructions to make the patterns appear fainter or more vivid referred to lowering or raising the image contrast. Audio feedback tones indicated if and when they had reached the limits of the trackball's contrast adjustment range. On their first visit subjects were asked to report what they saw at threshold, and specifically whether or not they could just see the a striped pattern with the same number of stripes as at the outset. This was used to provide some assurance that the thresholds obtained were for seeing the FD illusion.

2.3. Statistical analysis

The objective of this analysis was to see if the structure of the data permitted a method that was on average able consistently to discriminate normals from glaucomas. We explored several measures. One set of measures was the principal components of standardised threshold contrasts. The log transformation, implied by use of decibel contrasts ($20 \log_{10}(\text{contrast})$), did not resolve the inhomogeneity of the variance between the normals and glaucomas. Note that the variance of each

group, including the normals, was inhomogeneous. This suggested the use of quadratic discriminant analysis (QDA) to use all the data to estimate a function that best distinguished the normal and other diagnostic cases. Linear discriminant analysis (LDA) was also conducted for comparison (although the method was inappropriate given the different covariance structures for normals and glaucomas, see Fig. 3a). Logistic regression (LR) models based on either on the principal components or the worst of the superior and inferior hemifield thresholds were also examined. As a way of providing the best estimates of the covariance of each population, so as to build the best possible discriminant models, we first formed the discriminant models on data pooled from all visits. The models derived from the pooled data were then be applied to data from each visit and the average performance of the models could thus be assessed. We coded the quadratic discriminant model in Matlab based on Johnson and Wichern (1992). The logistic regressions were performed with Genstat. For the ROCs of Fig. 4 the risk factors were based on Fisher's linear discriminant function, and the quadratic classifier was the likelihood ratio assuming separate variance matrices (Johnson & Wichern, 1992). In the case of Fig. 5 we used the calculated probabilities based on the logit scores. The covariance model and the

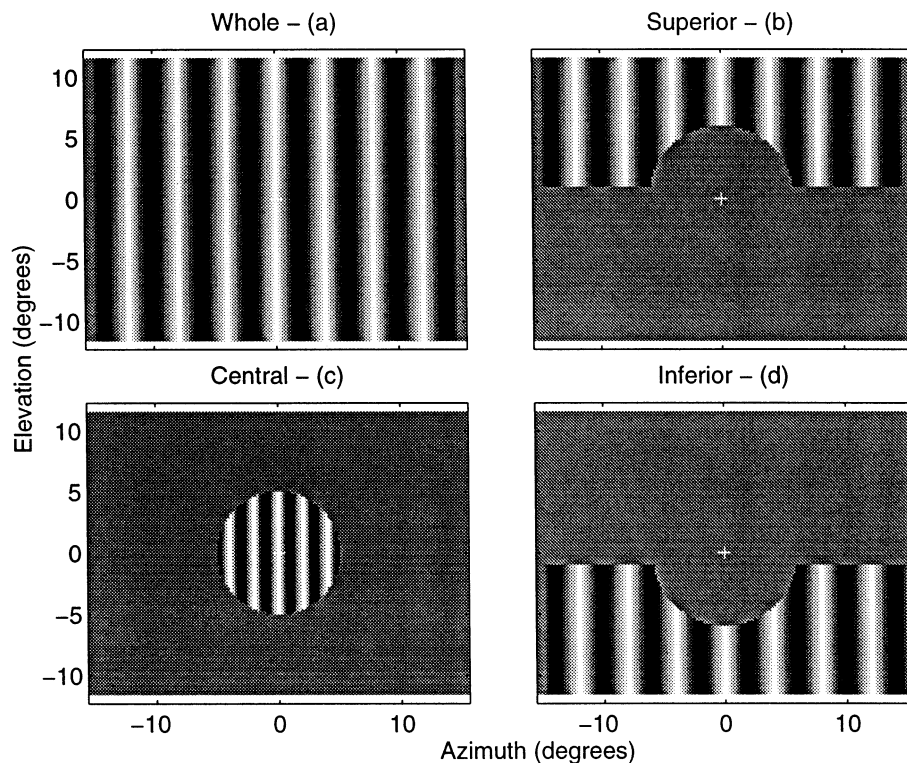


Fig. 1. Visual stimuli. Contrast thresholds were determined for these four regions of the visual field. (a) The whole field stimulus, and (b) the superior and, (d) inferior stimuli contained sinusoidal gratings at 0.25 cypd, while (c) the central 10° stimulus was set to 0.5 cpd. The contrast of all gratings was modulated at 25 Hz to produce the spatial frequency doubling illusion. Subjects were shown the grating patterns at 100% contrast at the outset of each threshold test sequence. Either three or six thresholds (MOA) were obtained for each field position and each subject on each of up to seven visits. Visits were at intervals of 3–4 months.

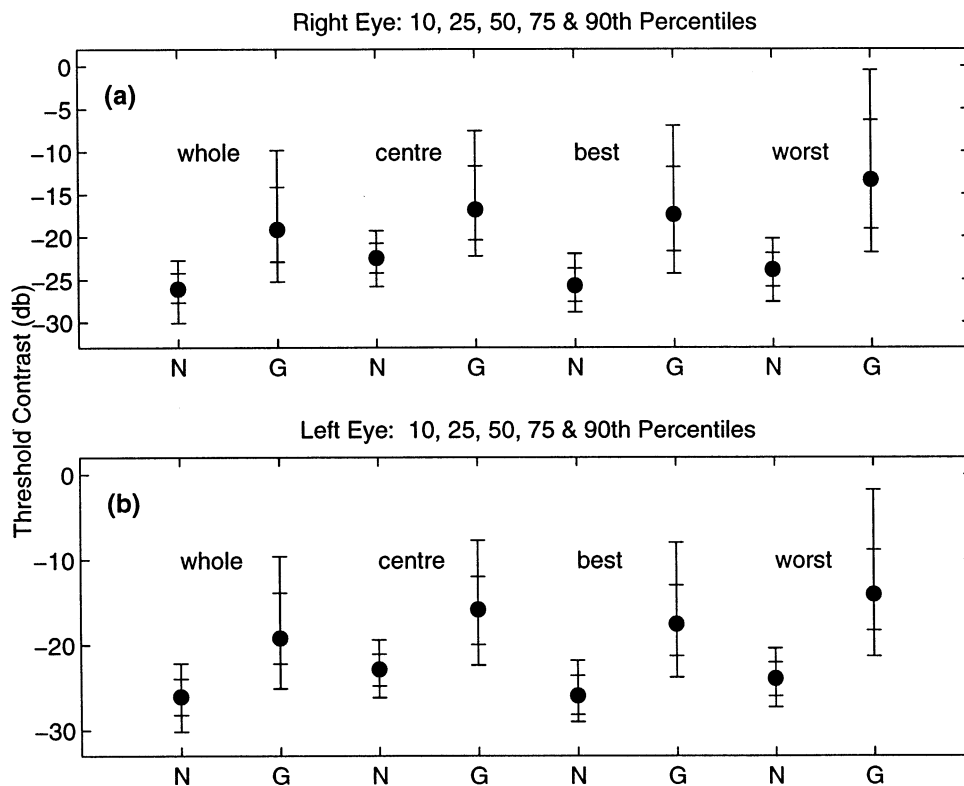


Fig. 2. Summary of raw contrast thresholds (decibels) for the two eyes and the four stimulus regions shown in Fig. 1. Data are for all seven visits from normal (N) and glaucoma (G) subjects. Dots are the 50th percentile of the distributions while the error bars show the 10, 25, 75 and 90th percentiles. The data for the Superior and Inferior visual fields (Fig. 1b and d) were sorted to form the worst and best hemifield thresholds for each eye.

regression models of Tables 7 and 8 were fitted using the restricted maximum likelihood (REML) method. The regression models presented in the section on the 'kink' test were done with SPSS.

3. Results

There were no significant effects of age, sex, family history or pupil size. There was also no effect of blockage mechanism *versus* open angle glaucomas, therefore in subsequently presented analyses all glaucomas are lumped together. On their first visit all subjects confirmed that at their threshold they could just see a pattern with the same number of stripes as at 100% contrast, i.e. the FD pattern. Further analysis suggested that there was some benefit from sorting data obtained from the inferior and superior visual fields into the best and worst thresholds respectively for each eye before further analysis. The raw thresholds for each visual field region (Fig. 1) and eye are summarised in Fig. 2. Notice that the worst hemifield data are the least overlapping: the 10th percentiles of the thresholds from glaucoma subjects being about level with the 75th percentile of the same thresholds from normal eyes. The distributions of Fig. 2 are based on measurements from

206 pairs of normal eyes, 266 left eyes with glaucoma and 202 right eyes with glaucoma. We examined the significance of the difference between the thresholds whole, central, worst, best for normals and glaucomas for the first visit to provide an example. The t -values for the right eye were 8.72_{Wh}, 7.34_{Ce}, 11.71_{Wo}, 9.69_{Be}, and for the left eye 8.80_{Wh}, 7.77_{Ce}, 10.24_{Wo}, 8.35_{Be}, the subscripts denoting the visual field region. There were 83 degrees of freedom and this meant that the worst performer, Central threshold from the right eyes ($t = 7.34$) was significant at $P < 0.4 \times 10^{-9}$, the other thresholds being several orders of magnitude more significant. Such comparisons, however, are meaningless if we are interested in the ability of a test to diagnose individuals drawn from particular populations, therefore we must look at sensitivities and specificities.

Fig. 2 also demonstrates that even the distributions of the tacitly log transformed (decibel) thresholds are skewed, more so for the glaucomas. Any discriminant analysis that seeks to use all the data to obtain the best possible discriminant function will have to account for this lack of normality and equality of distributions between the normal and glaucoma subjects. Sorting the thresholds into worst and best hemifield thresholds aggravated the problem of non-normal distributions. Fig. 3 illustrates an alternative approach and also a

further feature of the data structure, namely asymmetry of thresholds between eyes of glaucoma subjects. Our data set could in practice span an eight dimensional space, four measures having been obtained from each eye. We examined the structure of the data by taking the principal components (PCs) of the (standardised) thresholds for each eye. In general over the seven visits the first PC for each eye accounted for 90% of the variance, the second PC accounting for about 5% (Table 3). For the second PC to be considered it would have to contribute about one quarter of the variance, that is have an eigenvalue of one or greater (i.e. one variable's proportion of the variance). The first PC was essentially the mean of the four thresholds (whole, centre, inferior and superior) with slightly more weight

Table 3

First two principal components of the threshold data^a

	Right eye		Left eye	
	PC1	PC2	PC1	PC2
r^2	0.895	0.053	0.892	0.056
who	0.460	-0.361	0.465	-0.411
cen	0.409	-0.654	0.426	-0.636
inf	0.519	0.081	0.505	0.181
sup	0.593	0.660	0.590	0.628

^a For summary purposes all the data for the seven visits has been treated as one data set. The first two columns summarise the PCs for the right eye and the last two columns the left eye. The row labelled r^2 indicates the proportion of variance accounted for by each PC. The bottom four rows represent the weights (eigenvectors) applied to each of the whole (who), central (cen), inferior (inf) and superior (sup) visual field data to construct the PCs. Note that the proportion of variance (r^2) for the second PC is only about 5%.

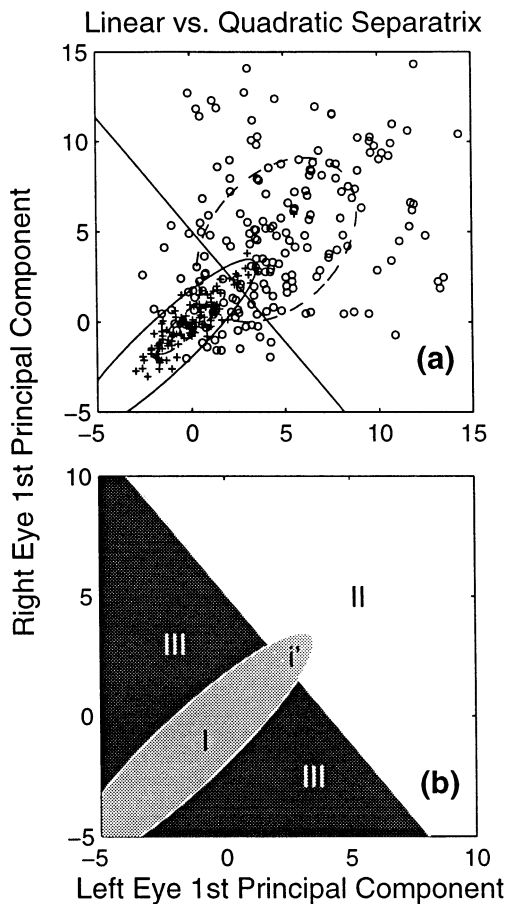


Fig. 3. Threshold data structure (a) shows a plot of the left versus right eye first principal components (PCs) of the threshold data for normals and glaucomas. The original data were in decibel contrast and the PCs are essentially the mean plus a constant. Thresholds for the glaucomatous eyes tend to be more different and so lie off the diagonal more. Dashed contours show one SD for the two populations. The solid straight line is the separatrix from a linear discriminant analysis (LDA). The solid ellipse around the normal data is the separatrix from a quadratic discriminant analysis (QDA). ROCs for these two cases are shown in Fig. 4b. (b) A diagrammatic representation of (a) showing that the QDA exploits the differences in many glaucoma subject's eyes by correctly classifying many glaucomas in region III and normals in the small region i' .

given to the superior and inferior fields (Table 3). For comparison PCs were also formed on the data set whole, central, best and worst thresholds and the resulting eigenvalues and vectors were very similar: the proportions of variance accounted for by the first component being 90.8 and 90.4% for the right and left eyes, respectively. The weights of the eigenvectors were very similar to those in Table 3 with best substituted for inferior, and worst for superior.

The fact that the data can be represented largely by the first PCs from the two eyes means that we can represent the eight-dimensional data set quite well in a two-dimensional plot as in Fig. 3. There are no units as the input threshold data for the PC analysis were standardised, but for simplicity we will refer to the PC data as thresholds, they being essentially scaled means of the four thresholds obtained per eye. The 316 pairs of data from the two eyes of normal and glaucoma subjects for whom six averages were obtained for each visual field location on each visit are shown. The most obvious feature of the data is that the distributions of the data for normal (+) and glaucoma subjects (O) are very differently shaped. The two dashed ovals of Fig. 3a indicate one standard deviation. A standard linear discriminant analysis (LDA) as formulated by Fischer, assuming that the distributions are the same shape, explicitly averages the covariance matrices from the data sets to be discriminated. Quadratic discriminant analysis (QDA) still uses both distributions to estimate the best separatrix but without the averaging and thus copes with the differing distributions. The solid diagonal line in Fig. 3a is the best fitting separatrix for LDA.

The other obvious feature of the data is that while normal subjects vary in their threshold criteria they nevertheless use same criteria for both eyes, hence their thresholds are spread out along the diagonal. By com-

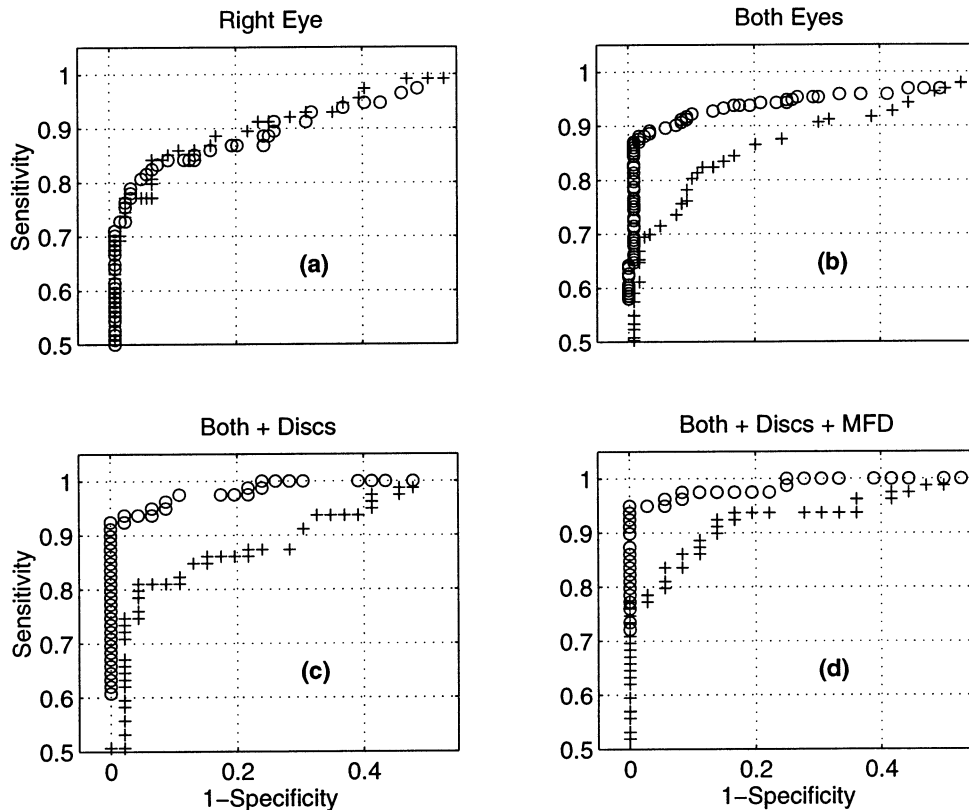


Fig. 4. ROC curves comparing various LDA and QDA cases. Only the upper left quadrant of the ROC curves are shown. ROCs for LDA are shown as +, QDA as \circ . (a) Right eye data only. (b) The case shown in Fig. 2 for left and right eye first PCs. QDA improves sensitivity even at high specificities. (c) Optic disc cupping added. (d) Humphrey 24-2 mean defect (MD) added.

parison individual glaucoma subjects' thresholds (\circ) tend to be quite different for their two eyes so their data points lay away from the diagonal, and hence away from where the data of normal subjects lie. The solid oval of Fig. 3a is the best fitting separatrix for QDA. Fig. 3b shows the obvious advantage of QDA in this case, the glaucoma subjects whose data lie in region III are now diagnosed correctly, as are those normal subjects whose data lie in region i'.

Fig. 4 demonstrates the effect upon sensitivity and specificity of making use of the fundamental structure of the data through QDA. In the ROC curves of Fig. 4a–d the results for LDA are shown as (+) and those for QDA as (\circ). Table 4 summarises sensitivity and specificities computed as the simultaneously highest obtained for a given ROC curve. As would be predicted both methods yield the same result when only data from one eye is used (Fig. 4a). When data from both eyes are incorporated, as suggested by Fig. 3a, sensitivity and specificity increase (Fig. 4b). Fig. 4c shows the effect of adding optic disc cupping to the diagnosis. Adding the HFA MD made little improvement.

Similar discriminant models can be constructed using logistic regression, one advantage of which is that the assumptions of the distributions being multivariate normal are not required. In an example of a linear logistic

model the logit function could be expressed as a function of subjects' left and right eye first PCs, C_{left} and C_{right} : $\log(p/(1-p)) = k_0 + k_1 C_{\text{left}} + k_2 C_{\text{right}}$. The closest equivalent to the QDA whose results are summarised in

Table 4

Percent sensitivities (sens) and specificities (spec) obtained for linear and quadratic discriminant models^a

	Linear		Quadratic	
	Sens	Spec	Sens	Spec
<i>(a) 3 Repts/visit</i>				
Right eye	85.6	81.7	85.0	86.4
Both eyes	84.0	84.9	82.5	95.6
Both + discs	87.5	91.2	87.0	97.8
Both + discs + MD	88.8	95.0	88.9	97.2
<i>(b) 6 Repts/visit</i>				
Right eye	86.0	86.6	86.0	84.0
Both eyes	84.5	83.2	89.6	94.1
Both + discs	84.8	84.8	92.4	97.8
Both + discs + MD	88.6	88.9	94.9	97.2

^a The figures given are the highest simultaneous sensitivities and specificities, that is the condition for which the two values are most similar. (a) For the seven visits where three repetitions/visit were available. (b) For the first four visits in which six repetitions/visit were conducted. The discriminant functions were formed on all the data to provide an average model.

Quadratic Logistic Regression

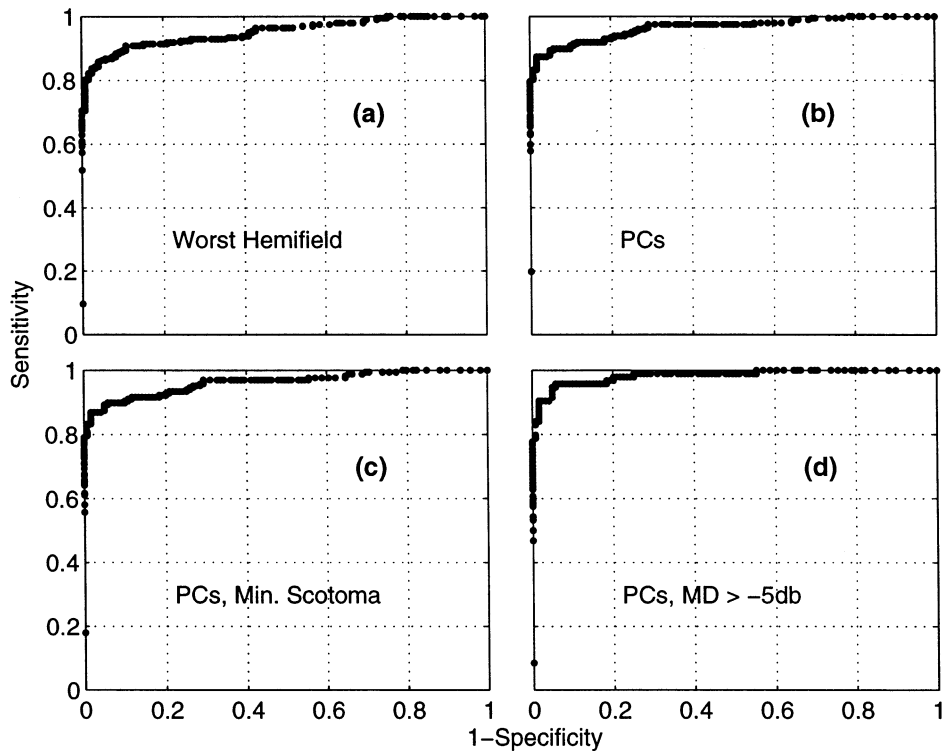


Fig. 5. ROC curves from logistic regression. QLR is more robust than QDA in permitting non-normal variance structures. Results were in agreement with QDA. (a) ROC curve for QLR case where the inputs are the worst hemifield threshold (average of six reps on each visit). (b) ROC like the case of Fig. 3b with left and right first PC data as the input. (c) As for (b) but only minor scotomas included. (d) As for (b) but only subjects with minor Humphrey Mean defects (> -5 db).

Fig. 4b is the quadratic logistic regression (QLR) model:-

$$\log(p/(1-p)) = k_0 + k_1 C_{\text{left}} + k_2 C_{\text{right}} + k_3 (C_{\text{left}})^2 + k_4 (C_{\text{right}})^2 + k_5 C_{\text{left}} C_{\text{right}} \quad (1)$$

The ROC for the model where the C_{left} and C_{right} values are the principal components from the left and right eyes, is shown in Fig. 5b. Notice that Figs. 4b and 5b are very similar. The ROC of Fig. 5a was produced using a QLR model where the left and right measures were the Worst hemifield thresholds from the two eyes (sens 88.3%, spec 91.6%). The sensitivities and specificities are almost as good as in Fig. 5b (sens 89.9%, spec 93.3%). Fig. 5b in turn is very similar to the equivalent case for QDA (both eyes Table 4b and Fig. 4b: sens 89.6%, spec 94.1%). Some of our glaucoma subjects had quite severe visual field losses and so we removed these subjects and reconstructed the ROC curve (Fig. 5c). There was no change (sens 89.9%, spec 93.3%). Since the HFA MD data had little positive effect (Fig. 4d) we decided to remove those subjects with mean defects greater than 5 db (Fig. 5d). This improved sensitivity and specificity (sens 95.7%,

spec 93.3%). The pupil diameter of the two groups was: > 5 db, $2.905 \text{ mm} \pm 0.061 \text{ SE}$; and ≤ 5 db, $2.754 \text{ mm} \pm 0.060 \text{ SE}$. Although ≤ 5 db group's pupils were significantly smaller ($t = 1.731$, $df = 412$, $P = 0.044$), the average difference of 0.151 mm is unlikely to be sufficient to explain the result.

These models are quite complex and so their performance needs to be confirmed for multiple data sets. So far the models presented have been formed on all the data regardless of the visit on which it was obtained. This in effect gives us an average model. Of course the models were formed on repeated measures from the same subjects and this would produce over estimated measures of goodness of fit of these discriminant functions, however we are not attempting to make such estimates. We did confirm that the average model performed as expected on if applied to each data set.

Fig. 5a,b summarises the performance of average models for the cases where we collected three averages per visit (5a) and six averages per visit (5b). The model used is the QLR model described above employing left and right eye first PCs. The typical performance, as illustrated by the median, is 86.7% (sens) and 93.8% (spec): this compares well with the performance obtained when the data from all visits was pooled ($N =$

550) 88.0% (sens) and 91.0% (spec). For the six average per visit cases the median performance, 90.8% (sens), 94.9% (spec), compared favourably with that obtained for the average model applied to all the data ($N=316$) 89.9% (sens) and 93.3% (spec) (see text above on Fig. 5). Table 6 summarises the results of applying the QLR model to subjects having a variety of conditions in their two eyes, from both eyes only weakly suspect, to both eyes glaucoma.

We decided to explore the correlation structure between the worst hemifield threshold measurements. We did this because these measurements had good diagnostic power and they were easy to comprehend. We considered only the first six visits thus providing 12 worst hemifield measurements (from two eyes \times six visits) for each subject. The input thresholds were the mean of the decibel thresholds obtained on the first three repetitions of each visit.

Our covariance model contained the between subject variance $\sigma_{\text{subject}}^2$, and the within subject variance which was decomposed into the between eye variance σ_{eye}^2 , the between visit variance σ_{visit}^2 , and the error variance σ_{error}^2 . Diagnostic plots showed that the response distribution is somewhat longer-tailed than the Gaussian distribution (not shown) but the model fit is otherwise acceptable. The longer-tailed response distribution tends to increase the standard errors and thereby make

Table 5
Single visit % sensitivities and specificities^a

Visit	Sensitivity	Specificity
<i>(a) 3 Repeats/visit</i>		
1	91.5	77.3
2	87.7	90.7
3	86.0	83.8
4	87.8	94.4
5	80.4	96.4
6	85.3	93.8
7	86.7	100
Mean	86.5	90.9
Median	86.7	93.8
S.D.	3.34	7.84
<i>(b) 6 Repeats/visit</i>		
1	91.6	88.6
2	86.9	97.6
3	90.7	92.3
4	91.0	100
Mean	90.0	94.6
Median	90.8	94.9
S.D.	2.13	5.15

^a Quadratic logistic regression function (Eq. (1)) sensitivities and specificities for several visits. (a) For all seven visits where three repetitions/visit were conducted. (b) For the first four visits in which six repetitions/visit were available. The discriminant functions were formed on all the data and then applied to each visit to examine the reliability of the resultant functions. S.D. indicates standard deviations of the means.

Table 6
Group % sensitivities and specificities^a

$P>0.5$	Sens	Spec	Eye diagnoses
<i>3 Repeats</i>			
32.5	59.6	59.0	rws & lws
42.7	63.2	60.6	rms & lms
51.9	68.4	66.5	rss & lss
93.2	93.2	92.6	rgl & lgl
73.5	79.6	77.6	rws & lgl or lw s& rgl
100	100	99.5	rms & lgl or lms & rgl
87.5	87.5	86.7	rss & lgl or lss & rgl
<i>6 Repeats</i>			
28.0	60.7	60.8	rws & lws
34.0	63.8	62.2	rms & lms
46.3	68.5	65.6	rss & lss
93.4	93.4	92.4	rgl & lgl
75.9	79.3	78.2	rws & lgl or lw s& rgl
100	100	99.2	rms & lgl or lms & rgl
80.0	100	79.8	rss & lgl or lss & rgl

^a The columns are: $P>0.5$ the proportion of subjects assigned a probability of being glaucomatous of 0.5 or greater; sens and spec the highest simultaneous sensitivity and specificity; eye diagnoses the logical selection criteria for inclusion in the group. The acronyms are related to the diagnostic classes of Table 1. So rws means right eye weak suspect, lss means left eye strong suspect, and rgl means right eye glaucoma, etc. Logical AND and OR are designated by & and or. Thus rgl & lgl refers to subjects where both eyes are glaucomatous; rms & lgl or lms & rgl indicates one eye is moderately suspect while the other is glaucomatous, but which eye has which condition doesn't matter. The titles 3 and 6 Repeats refer to the number of repeats averaged on each visit. The proportion of normals assigned a at least a 50% probability of being normal was 0.94 for the 6 Rep case and 0.88 for the 3 Rep case.

inferences more conservative. The correlation coefficients may be computed from the variance components, of particular interest are the correlation between eyes on the same visit $\rho_{\text{eyes/visit}}$, and the correlation between visits on the same eye $\rho_{\text{visit/eye}}$. The results are given in Table 7.

The subjects were grouped according to eye diagnoses: N/N being normal in both eyes, G/G both eyes glaucoma, S/S both eyes suspect (weak, moderate or strong), G/S one eye suspect the other glaucoma, and a small group whose diagnoses were initially listed as changing over the 6 visits. In fact this change group was caused by errors in classification that have been rectified for all the other analyses shown, but we left them unaltered because the original error was with the classification of some suspect eyes and was due to the marginal nature of that diagnosis. Thus, the change group represents borderline cases that we have left in for interest's sake. As would be expected the change group performs close to the G\G group. Note that the units of the first four rows are db². For all groups the between visit variation, σ_{visit}^2 , is about 5 db², providing a mean standard deviation of 2.22 db \pm 0.09 S.E. in the thresholds obtained from visit to visit for all groups.

The between subject variance, $\sigma_{\text{subject}}^2$, was especially large for the G/G group ($45.74 \text{ db}^2 \pm 14.58 \text{ se}$) yielding a standard deviation of about 7 db. The variation in thresholds between eyes was greatest for the G/S group ($23.34 \text{ db}^2 \pm 5.46 \text{ s}$). The correlation between eyes across the visits, $\rho_{\text{visit/eye}}$, was on average reasonable, being $0.906 \pm 0.023 \text{ S.E.}$ for the G/G group.

We also wished to examine the relationship between worst hemifield threshold and Humphrey MD, disc cupping, and IOP measured on the test day. Disc cupping was measured in two different ways, by contour and by colour. In our first analysis we fit a mixed linear regression model for each of these other four variables. In view of differences in the variance structure for different subgroups in the study, this analysis was first carried out separately on the five different subgroups. Table 8 details the regression coefficients for these diagnostic variables and their standard errors. Note that coefficients near zero indicate weak to no relationship, while the sign of a coefficient indicates the nature of the correlation.

There is no significant relationship between the worst hemifield threshold measurement and the IOP for any of the groups but there is a significant relationship between worst and the other diagnostic variables for every group of subjects. The lack of correlation with IOP is expected as the non-normal subjects were mainly medicated. The relationship between the worst and the two cupping variables is strong and very similar. The coefficients are positive showing that high cupping corresponds to a high worst hemifield threshold. The relationship between the worst hemifield threshold measurement and MD is significant but the regression coefficients are relatively small (note that worst and MD had the same sign).

We also examined the relationship between worst hemifield threshold for each eye and, MD, cupping and IOP simultaneously. Disc cupping by colour was excluded because is essentially the same quantity as cupping by contour (see Table 8). As before, when the measures were fit separately, the MD and cupping data are significant but IOP is not. We do not show these figures because the regression coefficients for MD and IOP were very similar to those in Table 8 although the

regression coefficients for cupping were about 40% smaller.

We also examined progression over the seven visits using the logit (log odds) values from the logistic regression model based on the left and right eye principal components (Fig. 5b). The logit values are fitted values of $\log(p/1-p)$ from Eq. (1) for each subject and visit. Figure 6 shows the logit values for the 316 cases considered for Fig. 5b. The insert on the right of Fig. 6 illustrates that the 10th and 90th percentiles of the distributions of the logit values for normals and glaucomas are just abutting. If the logit values increased for a given set of subjects over the course of our study this would indicate disease progression.

Given the 2 year duration of our study we offered patients new or supplemented treatment when their ophthalmologist (IG) decided such treatment was needed. That some subjects had supplementary changes in their treatment provided the possibility of examining disease progression relative to the treatment time. Treatment supplements were typically an increase in the dose or frequency of topical beta-blockers or occasionally laser trabeculoplasty. There are two possible expectations for such treatment changes. Firstly, one might expect to observe a decline in performance before treatment, followed by a partially or completely arrested decline in performance. Alternately, since the decision to treat would typically be preceded by some pathological change, one might expect average post treatment performance to become worse, given that the increased treatment would only act to preserve the vision remaining after the increased pathology. Under the latter hypothesis synchronising subject data with respect to augmented treatments may simply synchronise our measurements to clinical presentation of progression.

We included those subjects who had a treatment change after their first and before the seventh visit, who attended at least five visits, and who had a simultaneous augmentation of treatment in both eyes. We found that there were 20 such subjects across the various non-normal diagnostic classes that attended a total of 117 visits. Using the symbols W,M,S,G to denote weak, moderate, strong, glaucoma the eye-wise diagnoses, then the 20 subjects can be diagnostically

Table 7
Visit 1 to 6 variance (σ^2) and correlation (ρ) structure of the Worst Hemifield thresholds^a

	N/N	G/G	S/S	G/S	Change G/S
$\sigma_{\text{subject}}^2$	6.52 ± 2.78	45.74 ± 14.58	14.20 ± 1.80	13.08 ± 6.68	37.27 ± 13.47
σ_{eye}^2	7.81 ± 1.80	15.45 ± 4.19	4.27 ± 0.47	23.34 ± 5.46	3.67 ± 1.32
σ_{visit}^2	4.92 ± 0.69	3.69 ± 0.72	5.62 ± 0.33	5.14 ± 0.80	5.45 ± 1.02
σ_{error}^2	1.80 ± 0.21	2.67 ± 0.37	1.84 ± 0.09	2.85 ± 0.34	2.01 ± 0.32
$\rho_{\text{eyes/visit}}$	0.543 ± 0.088	0.732 ± 0.079	0.764 ± 0.032	0.410 ± 0.157	0.882 ± 0.059
$\rho_{\text{visit/eye}}$	0.681 ± 0.049	0.906 ± 0.023	0.712 ± 0.032	0.820 ± 0.050	0.846 ± 0.066

^a The units for the first four rows describing the variance components are decibels contrasts squared. The errors are standard errors.

Table 8
Relationship between worst, MD, cupping and IOP^a

	N/N	G/G	S/S	G/S	Change G/S
MD	0.049 ± 0.007	0.079 ± 0.012	0.049 ± 0.005	0.081 ± 0.009	0.035 ± 0.011
Contour	0.928 ± 0.316	1.765 ± 0.481	0.771 ± 0.139	1.601 ± 0.431	0.852 ± 0.401
Colour	0.767 ± 0.276	1.581 ± 0.471	0.770 ± 0.126	1.456 ± 0.370	0.958 ± 0.425
IOP	0.008 ± 0.008	0.001 ± 0.009	0.002 ± 0.003	−0.005 ± 0.007	0.011 ± 0.011

^a The figures are the regression coefficients describing the relationship between the worst hemifield thresholds over the first six visits and both eyes for four diagnostic measures. For the MD coefficients the units are db (worst hemifield threshold)/dl (HFA threshold). For contour and colour refer to vertical cup to disc ratios (Section 2) the units can be thought of as db (worst)/cup. The units for the IOP coefficients are ar db/mmHg.

described as follows: W/W 5, W/M 2, W/S 1, W/G 1, M/M 1, S/S 1, S/G 1, G/G 4. Thus, 15 of the 20 were glaucoma suspects. We compared these 20 with the 25 normal subjects who made five or more visits (total 140 visits). For example across the seven visits the slope (not shown) of the logit values for normals was -1.90 ± 0.068 S.E. ($P = 0.06$). This marginally significant slope suggested a slight learning effect (more negative logits indicating better performance cf. Fig. 6). The slope for the treatment group was 0.07 ± 0.41 S.E., while not different from 0 ($P = 0.78$) this was significantly different from the slope of the normals ($P = 0.004$) suggesting that the glaucomas progressed relative to the normals.

We then fitted linear regression models to the treatment group selecting various subgroups to examine what if anything was happening with their progression relative to the date of treatment. Subgroups were defined using the criteria of the seriousness of their diagnosed condition (weak, strong, glaucoma) and the severity of disease indicated by their logit values. Fig. 7a illustrates performance of five persons who were classed as strong suspect or glaucoma in at least one eye and whose logit value was less than or equal to five. Thus, their logit values based on their combined thresholds indicated mild disease even if their condition in at least one eye presented clinically quite badly. In this case a so called *kink* model (two slopes) model (methods) was a better fit ($P = 0.031$) than a single line model, the total model being significant at $P = 0.006$ ($F = 5.98_{2,37}$). The kink model fits a slope for the whole epoch plus a slope from day 0 onward (treatment took place between day -1 and day 0). The slope over the whole epoch was -0.264 ± 0.414 S.E. and from treatment day 0 onward the slope increased by 1.562 ± 0.699 S.E. ($P = 0.032$), a post treatment progression of 1.30. The decline in the overall slope is suggestive of the performance of the normals. Providing their logit values were lower than five other selections of subjects tended to show simple progression, that is their performance declined linearly but showed no significant kink. An example

is shown (Fig. 7b) for the 15 subjects who were weak to strong suspects and whose logit values were less than five. For seven subjects with logit values greater than five there was no measurable progression (Fig. 7c) only a high mean logit value indicative of glaucoma.

Since the data in question are time series data it is possible that serial correlations might affect the veracity of our error estimates, however, correlations from 'lag-one' plots and Durban–Watson statistics indicated no such significant correlations. More complex regression models containing additional subject-wise effects (means) were sometimes more significant than the regressions presented, but these models lead to the same conclusions with respect to the kinks and progression so we have elected to present the less complex regression results.

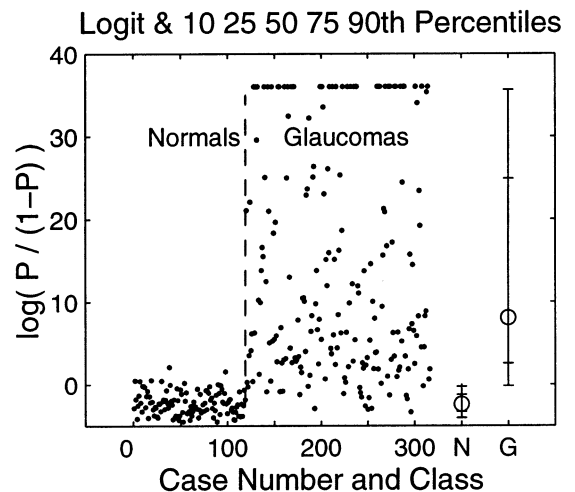


Fig. 6. Illustration of the observed logit (log odds, Eq. (1)) values for 316 subject visits for which averages of six thresholds for each region were obtained from normal and glaucoma subjects. The vertical dashed line separates the logit values for the normals and glaucomas. The two plots on the left of the figure summarise the distributions of logits for normals (N) and glaucomas (G) where dots are the 50th percentile of the distributions while the error bars show the 10, 25, 75 and 90th percentiles.

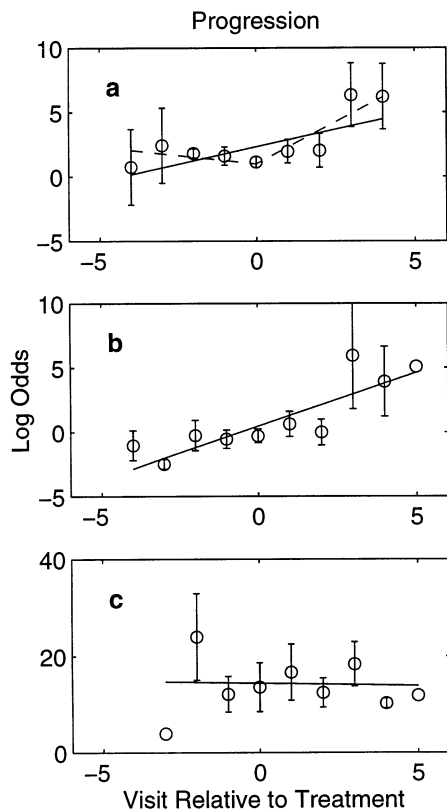


Fig. 7. Progression of the logit values relative to the visit after the first treatment change (Day 0). Two regression models were fitted for each case. The first was a simple line plus a mean effect, the second model included the sum of an additional line beginning at day 0. The second model thus permitted quantification of any significant kink in the data characterised by either an increased or a decreased slope related to treatment. (a) The case for subjects whose diagnosis was strong suspect or worse in at least one eye and whose logit value was ≤ 5 (cf. Fig. 6). Here the three parameter kink model (dashed line) was significantly better ($P = 0.031$) than a single line plus mean effect (solid). The slope over the whole period was -0.264 ± 0.414 S.E. and from Day 0 onward the slope increased by 1.562 ± 0.699 S.E. ($P = 0.032$), a post treatment progression of 1.30. (b) A significant slope ($P = 0.001$) of 0.832 ± 0.222 indicates progression in subjects classed as weak to strong suspects and Logits ≤ 5 . (c) For the same diagnostic groups as (a) but having logit values greater than five there was no measurable progression (slope of -0.098 ± 0.851 S.E., $P = 0.655$) only a mean effect of 14.42 ± 1.92 S.E. logits, $P = 0.0000$. For (a) to (c) error bars are S.E., no error bar indicates 1 data point only.

4. Discussion

The distributions of the thresholds for normal and glaucoma subjects overlapped little (Fig. 2). Even the poorest performing threshold, central, still being different from normals at the $P < 0.4 \times 10^{-9}$ level indicating that FD methods detect central ganglion cell loss (Glovinsky et al., 1993). The structure of the data (Fig. 3) required discriminant models that could cope with unequal, non-normal, covariance structures. The lack of concordance in the thresholds from the two eyes of the Glaucoma subjects (Figs. 3 and 7) meant that two eye models would perform better than single eye models

(Fig. 4b). Two types of discriminant models were tried for comparison: quadratic discriminant analysis (QDA), and a quadratic form of logistic regression (QLR). QLR had the advantage of being more tolerant of non-normal distributions and multicollinearity (the latter evidenced by the large first PCs). QDA permitted easy comparison with linear discriminant models. There was good agreement: 89.9% sensitivity for 93.3% specificity with QLR; 89.6% sensitivity for 94.1% specificity for QDA (cf. also Fig. 4b and Fig. 5b).

Extracting the first principal components succinctly represented our eight dimensional data set in two dimensions that accounted for about 90% of the variance (Table 3). The eigenvectors placed more weight on the inferior and superior thresholds, or their relatives the best and worst hemifield thresholds (Table 3). A corollary of this statement is that models based purely on the worst threshold provided similar performance to the PC models providing 88.3% specificity for 91.6% sensitivity. Also encouraging was the relatively small proportion of weak and moderate suspects that were diagnosed as glaucomas: 28% and 34% respectively for six reps (Table 5). Thus, measurements from relatively small numbers of quite large portions of the visual field can yield good performance.

Given that the principal components are essentially an average of the separate thresholds (Table 3) suggests that there is diffuse as well as local ganglion cell loss in glaucoma, at least when FD stimuli are used. This conclusion is supported by several studies (Quigley et al., 1982; Airaksinen et al., 1986; Schulzer et al., 1987; Glovinsky et al., 1993). Quigley et al. (1982) have described a subject who, although missing 40% of his retinal ganglion cells, showed no field losses, had no visual acuity loss and had only moderately cupped discs. More recently Quigley et al. (1989) have determined that a uniform decrease of -5 db across the retina obtained with automated perimetry would correspond to a loss of approximately 20% of all retinal ganglion cells. Anatomical (Glovinsky et al., 1993) and ERG studies (Bach et al., 1992; Falsini et al., 1992) also indicate macular as well as peripheral retinal damage in early glaucoma.

Nevertheless diffuse loss as assessed by standard automated perimetry is not a good indicator of early glaucomatous loss (Heijl et al., 1989; Lachenmayr et al., 1991; Asman & Heijl, 1994). Loss of reliability visual field measures at fairly large scales (Heijl, 1993) might be however, which could be indicative of changes to the retinal gain control system with which FD is associated (Maddess et al., 1998). Interestingly, a related stimulus to FD, high frequency flicker thresholds, provides similar diagnostic power whether measured in the central or peripheral visual (Tyler, 1981). These related stimuli also show good correlation with elevated IOP (Tyler, Ryu & Stamper, 1984). It is worth noting

that trials where seven spatial frequencies in the range 0.063–0.813 cpd, 27 Hz, presented in a 5° aperture at 15° (nasal) eccentricity showed that intermediate spatial frequencies around 0.5 cpd are the most reliable for glaucoma diagnosis (Maddess & Severt 1999a,b). Thus FD stimuli have an advantage over pure flicker. The results of Tyler and the present results may together indicate that diffuse changes associated with glaucoma are a feature of perimetry conducted with FD-like stimuli.

An advantage of FD perimetry methods therefore might be that a mean defect obtained with these methods could be more closely allied with glaucomatous damage. Indeed, two examples of a lack of concordance between HFA mean defects and the currently described thresholds were presented. In the first case adding the mean defect to the discriminant model did little to improve sensitivity and specificity (Fig. 4d). At the same time removing persons with HFA defects worse than -5 db actually improved performance (Fig. 5d, sens 95.7%, spec 93.3%). Finally, although there was significant correlation between MD and the Worst hemifield threshold, the regression coefficients (Table 8) indicated a weak relationship. By contrast, disc-cupping improved discriminant models (Fig. 4c) and produced larger regression coefficients (Table 8) in agreement with previous results (Maddess & Henry, 1990).

The discriminant models discussed so far were formed on all the data pooled over the seven visits to indicate average performance. We then applied these average models to the data from each of the seven visits in turn. Median performance was a good match to the performance of the models applied to all the data (Table 5). We found that averaging six repetitions of the threshold data provided somewhat better performance than averaging three repetitions, sensitivity being improved (Tables 4 and 5). Thus three repetitions of measurements from the two hemifields might provide a rapid screening procedure with good specificity. Scotomas are an obvious feature of glaucoma however (e.g. Capriolli, 1991), and so for long term monitoring of patient condition a perimeter testing a few more smaller regions with FD stimuli might be best (James & Maddess, 1996), the Humphrey model 710 (Maddess, 1991, 1995) being a good example. Recent work seems to confirm its value as a screening device (Kondo, Yamamoto, Sato, Matsubara & Kitazawa, 1998; Quigley 1998; Sponsel, Argano, Trigo & Mensah, 1998).

Johnson and Samuels (1997) have performed similar tests to those done here using more stimulus regions. Those authors restricted themselves to single eye discriminant models and their reported performance matches our single eye models. Another difference between the two studies was the use by Johnson and Samuels (1997) of a Modified Binary Search (MOBS) (Tyrrell & Owens, (1987) threshold procedure. The test

retest variability here was assessed over 6 visits and two eyes for the worst hemifield thresholds to be an average standard deviation of $2.22 \text{ db} \pm 0.08 \text{ S.E.}$ By comparison Johnson and Samuels (1997) using a MOBS method reported $\pm 2\text{--}3\%$ test–retest difference at 95% confidence for normal eyes and $\pm 6\text{--}10\%$ for Glaucomatous eyes. For comparison we translate our results into absolute contrast providing a 95% confidence range of about $\pm 3.41\%$ for normals, and $-8.25\text{--}13.76\%$ for glaucomas. The MOBS method (four reversals) was therefore about 1.3 times more accurate than the MOA Worst threshold method (Fig. 5a) (three repetitions). We have not calculated the test-retest variability for six repetitions or the PC based measures but would expect them to provide improved performance in line with their improved sensitivity and specificity.

The performance of many threshold schemes (Shelton, Picardi & Green, 1982; Stillman 1989) including MOA (Hesse, 1986) do not differ greatly. Methods that estimate both the threshold and the slope are probably superior to MOA (e.g. Yager & Beard, 1994). The test–retest variability reported here is similar to that reported by Johnson and Samuels (1997), and is also slightly smaller than for clinical trials estimating high frequency flicker thresholds reported for 804 subjects using a sophisticated staircase procedure (Tyler, 1989).

A factor that may have assisted us is the Eureka effect of Ahissar and Hochstein (1997). Those authors found that perceptual learning of difficult threshold tasks similar those found in perimetry are greatly enhanced by showing the subjects the test targets at suprathreshold levels just once for about 30 s. Subjects shown these suprathreshold stimuli had initial performance levels normally obtained only after thousands of trials (hence the name Eureka), and subsequent learning was more stereotypical. The Eureka effect requires that the stimuli be shown in the exact location in which the threshold test will be determined. In our trials subjects were always shown the test stimulus at the beginning of each trial, hence they saw the stimuli under the conditions for the Eureka effect on every trial.

The progression data showed evidence of learning effects, the normals improving performance was as evidenced by a negative slope $-1.90 \pm 0.068 \text{ S.E.}$ in their logit values as a function of subject visit. This slope was significantly different from the flatter performance of the treatment group ($P = 0.004$). Synchronising subjects' data with respect to the date of treatment increase lead to an increase in the slope of their logit values post treatment suggesting that whatever triggered the ophthalmologist's decision to supplement treatment had caused damage. Significant, progression was demonstrated by an increase in logit values of $0.832 \pm 0.222 \text{ S.E.}$ (Fig. 7b, $P = 0.001$) in the 15 treatment change subjects who whose eye diagnoses ranged

from weak to strong suspect. No progression was measurable in the subjects with logit values > 5 (Fig. 7c, slope of -0.098 ± 0.851 S.E., $P = 0.655$). It is possible that any small progression in this group was swamped by their large logit values (mean = 14.42 logits ± 1.92 S.E.).

There is now considerable evidence that the FD illusion is produced by M_y -cells (James et al., 1995; Bedford et al., 1997; Maddess et al., 1997, 1998; Maddess & Kulikowski, 1999). If M_y -cells produce the FD illusion that would go a long way toward explaining the good performance of FD stimuli for glaucoma testing (Maddess, 1991; Maddess & Henry, 1992; Johnson & Samuels, 1997; Kondo et al., 1998; Quigley, 1998; Sponsel et al., 1998; Maddess & Severt, 1999a,b). In that case the good performance results from two factors: cell size and low retinal coverage factor (Maddess & Henry, 1992; Maddess et al., 1998). Several reports indicate that the M_y -cells are larger than M_x -cells (for summary see Maddess et al., 1998). At the same time there is now considerable evidence (Quigley et al., 1987; Quigley et al., 1988; Quigley et al., 1989; Glovinsky et al., 1991; Glovinsky et al., 1993; Smith et al., 1993) that large retinal ganglion cells are damaged more by glaucoma. The coverage factor (the number of cells seeing each point in visual space) of M_y -cells is perhaps < 1 (Maddess et al., 1998). By contrast the coverage factor for P-cells is close to 20 (e.g. Crook, Lange-Maleki, Lee & Valberg, 1988). So, if a person lost 1 P cell and 1 M_y -cell on a given part of their retina one might expect a change in sensitivity of about 2.5% for the P-system ($1 - (19/20)^{0.5}$, assuming Poisson noise and summation of all the P-cells for that point) compared to a 100% change for the M_y -cell system. Random M_y -cell losses might thus produce detectable diffuse loss, while quite large focal lesions might be required to detect P-cell loss.

Another issue is whether or not the FD illusion is observed at threshold (Kelly, 1966; Richards & Felton, 1973; Kulikowski, 1975). In a spatial frequency matching paradigm a pure FD effect has been reported for 2% contrast 1 cpd gratings at high flicker rates (Kulikowski, 1975). This is half the contrast of the lower 95% confidence limit for our normal subjects of 3.82% (see also Maddess & Kulikowski, 1999). FD at low contrast is not surprising because FD appears to be produced by a rectifying nonlinearity (Tyler, 1974; Kelly, 1981), rectification being hard down to small signal sizes. In comparisons of contrast thresholds derived for detecting the orientation of FD targets and the current MOA method we have shown that the orientation threshold is seven times lower than for the MOA threshold (Maddess & Severt, 1999a). Nevertheless there was good diagnostic concordance between the two methods suggesting the same mechanism was being tested. It is clear that our MOA thresholds here are well within the range where FD is observed in normals.

We cannot say that our glaucoma subjects definitely saw FD at threshold. On their first visit, however, all subjects were asked if they saw the same striped pattern at what they claimed to be their threshold as they had seen at the beginning of the test. All subjects replied in the affirmative except for a few serious glaucoma cases who reported seeing a blank quiescent screen and thus recorded thresholds of 100% contrast. This might suggest that at these spatiotemporal conditions there is no other mechanism with which to see these patterns other than that which produces FD. At very low contrasts one might see FD patterns with low reliability making it difficult to distinguish FD from a pattern of a differing spatial frequency. Any attempt to identify the frequency is complicated if the pattern is a sum of the original and an FD pattern as this leads to an illusory apparent fineness, even in static gratings (Maddess & Kulikowski, 1999).

Overall it appears that tests employing FD stimuli applied in even relatively few, large, regions of the visual can yield good diagnostic performance and could potentially serve at least as a screening device, even comparison of the two hemifields yielding good diagnostic power. Even the Central thresholds of glaucomatous eyes were well distinguished from those of normal eyes. It appears assessing relatively large portions of the visual field with FD stimuli may detect a form of diffuse loss diagnostic of glaucoma. The losses detected with FD appear to be sensitive enough to track progression, even in subjects who do not yet show conventional field defects.

References

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.
- Airaksinen, P. J., Lukowski, R., Drance, S. M., & Price, M. (1986). Colour vision and retinal nerve fiber layer in early glaucoma. *American Journal of Ophthalmology*, *101*, 208–213.
- Asman, P., & Heijl, A. (1994). Diffuse visual field loss and glaucoma. *Acta Ophthalmologica (Copenhagen)*, *72*, 303–308.
- Bach, M., Pfeiffer, N., & Birkner-Binder, D. (1992). Pattern electroretinogram reflects diffuse retinal damage in early glaucoma. *Clinical and Visual Science*, *7*, 335–340.
- Bedford, S., Maddess, T., Rose, K. A., & James, A. C. (1997). Correlations between observability of the spatial frequency doubled illusion and a multi-region PERG. *Australian and New Zealand Journal of Ophthalmology*, *25*, 91–93.
- Caprioli, J. (1991). Automated perimetry in glaucoma. *American Journal of Ophthalmology*, *111*, 235–239.
- Crook, J. M., Lange-Maleki, B., Lee, B. B., & Valberg, A. (1988). Visual resolution of macaque retinal ganglion cells. *Journal of Physiology*, *396*, 205–224.
- Dawson, W. W., Brooks, D. E., Hope, G. M., Samuelson, D. A., Sherwood, M. B., Engel, H. M., & Kessler, M. J. (1993). Primary open angle glaucoma in the rhesus monkey. *British Journal of Ophthalmology*, *77*, 302–310.
- Falsini, B., Colotto, A., Porciatti, V., & Porrelo, G. (1992). Follow-up study with pattern ERG in ocular hypertension and glaucoma

- patients under timolol maleate treatment. *Clinical and Vision Science*, 7, 341–347.
- Glovinsky, Y., Quigley, H. A., & Dunkelburger, G. R. (1991). Retinal ganglion cell loss is size dependent in experimental glaucoma. *Investigative Ophthalmology and Visual Science*, 32, 484–491.
- Glovinsky, Y., Quigley, H. A., & Pease, M. E. (1993). Foveal ganglion cell loss is size dependent in experimental glaucoma. *Investigative Ophthalmology and Visual Science*, 34, 395–400.
- Heijl, A. (1993). Perimetric point density and detection of glaucomatous visual field loss. *Acta Ophthalmologica (Copenhagen)*, 71, 445–450.
- Heijl, A., Lindgren, A., & Lindgren, G. (1989). Test-retest variability in glaucomatous visual fields. *American Journal of Ophthalmology*, 108, 130–135.
- Hesse, A. (1986). Comparison of several psychophysical procedures with respect to threshold estimates, reproducibility and efficiency. *Acustica*, 59, 265–273.
- James, A. C., & Maddess, T. (1996). Glaucoma testing using non-linear systems identification techniques. *Australia, Patent No.*, 667, 702.
- James, A. C., Maddess, T., Rouhan, K., Bedford, S., & Snowball, M. (1995). Evidence for M_γ-cell involvement in the spatial frequency doubled illusion as revealed by a multiple region PERG for glaucoma. *Journal of the Optical Society of America VSIA Technical Digest*, 1, 314–317.
- Johnson, C. A., & Samuels, S. (1997). Screening for glaucomatous visual field loss with frequency doubling perimetry. *Investigative Ophthalmology and Visual Science*, 38, 413–425.
- Johnson, R. A., & Wichern, D.W. (1992). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall, pp. 475, 491.
- Kelly, D. H. (1966). Frequency doubling in visual responses. *Journal of the Optical Society of America*, 56, 1628–1633.
- Kelly, D. H. (1981). Nonlinear visual responses to flickering sinusoidal gratings. *Journal of the Optical Society of America*, 71, 1051–1055.
- Kondo, Y., Yamamoto, T., Sato, Y., Matsubara, M., & Kitazawa, Y. (1998). A frequency-doubling perimetric study in normal-tension glaucoma with hemifield defect. *Journal of Glaucoma*, 7, 261–265.
- Kulikowski, J. J. (1975). Apparent fineness of briefly presented gratings: balance between movement and pattern channels. *Vision Research*, 15, 673–680.
- Lachenmayr, B. J., Drance, S. M., Chauhan, B. C., House, P. H., & Lalani, S. (1991). Diffuse and localized glaucomatous field loss in light-sense, flicker and resolution perimetry. *Graefes Archives of Clinical and Experimental Ophthalmology*, 29, 267–273.
- Maddess, T. (1991). Method and apparatus for use in diagnosis of glaucoma. USA. *Patent No.* 065,767, 5.
- Maddess, T. (1995). Early detection of glaucoma. USA, *Patent No.* 5912723.
- Maddess, T., Bedford, S., James, A. C., & Rose, K. A. (1997). A multiple frequency, multiple region pattern electroretinogram investigation of nonlinear retinal signals. *Australian and New Zealand Journal of Ophthalmology*, 25, 94–97.
- Maddess, T., Goldberg, I., Dobinson, J., Wine, S., & James, A. C. (1995). Clinical trials of the frequency doubled illusion as an indicator of glaucoma. *Investigative Ophthalmology and Visual Science*, 36, 335.
- Maddess, T., Goldberg, I., Dobinson, J., Wine, S., Welch, A. H., & James, A. (1998). Testing for glaucoma with the frequency doubled illusion. *Investigative Ophthalmology and Visual Science*, 39, 26.
- Maddess, T., Hemmi, J., & James, A. C. (1998). Evidence for spatial aliasing effects in the Y-like cells of the magnocellular visual pathway. *Vision Research*, 38, 1843–1859.
- Maddess, T., & Henry, G. H. (1990). Density of Nonlinear visual units and glaucoma. *Investigative Ophthalmology and Visual Science*, 31, 230.
- Maddess, T., & Henry, G. H. (1992). Nonlinear visual responses and visual deficits in ocular hypertensive and glaucoma subjects. *Clinical and Visual Science*, 7, 371–383.
- Maddess, T., James, A. C., Goldberg, I., Wine, S., & Dobinson, J. (1999). Parallel PERG perimetry for glaucoma using the frequency doubling illusion. *Investigative Ophthalmology Visual Science* (submitted).
- Maddess, T., & Kulikowski, J. (1999). Apparent fineness of compound gratings. *Vision Research*, 39, 3404–3416.
- Maddess, T., & Severt, W. (1999a). Head to head trials of eight tests using the frequency doubling illusion to diagnose glaucoma. *Investigative Ophthalmology and Visual Science*, (Ft. Lauderdale, USA), 40, 5700.
- Maddess, T., & Severt, W. L. (1999b). Testing for glaucoma with the frequency doubling illusion in the whole, macular and eccentric visual fields. *Australian and New Zealand Journal of Ophthalmology*, 27, 198–200.
- Quigley, H. (1998). Identification of glaucoma-related visual field abnormality with the screening protocol of frequency doubling technology. *American Journal of Ophthalmology*, 125, 819–829.
- Quigley, H., Sanchez, R., Dunkelberger, G. R., L'Hernault, N. L., & Baginski, T. A. (1987). Chronic glaucoma selectively damages large optic nerve fibres. *Investigative Ophthalmology and Visual Science*, 28, 913–920.
- Quigley, H. A., Addicks, E. M., & Green, R. W. (1982). Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fibre loss and visual field defect in glaucoma, ischaemic neuropathy, papilloedema, and toxic neuropathy. *Archives of Ophthalmology*, 100, 135–146.
- Quigley, H. A., Dunkelberger, G. R., & Green, W. R. (1989). Studies of retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *American Journal of Ophthalmology*, 107, 453–464.
- Quigley, H. A., Dunkelburger, G. R., & Green, W. R. (1988). Chronic human glaucoma causing selectively greater loss of large optic nerve fibres. *Ophthalmology*, 95, 357–363.
- Richards, W., & Felton, T. N. (1973). Spatial frequency doubling: retinal or central. *Vision Research*, 13, 2129–2138.
- Schulzer, M., Mikelberg, F. S., & Drance, S. M. (1987). A study of the value of the central and peripheral isoptres in assessing visual field progression in the presence of paracentral scotoma measurements. *British Journal of Ophthalmology*, 71, 422–427.
- Shelton, B. R., Picardi, M. C., & Green, D. M. (1982). Comparison of three adaptive psychophysical procedures. *Journal of the Acoustic Society of America*, 71, 1527–1533.
- Smith, E. L. III, Chino, Y. M., Harwerth, R. S., Ridder, W. H. III, Crawford, M. L. J., & DeSantis, L. (1993). Retinal inputs to the monkey's lateral geniculate nucleus in experimental glaucoma. *Clinical and Vision Science*, 8, 113–139.
- Sponsel, W., Argano, S., Trigo, Y., & Mensah, J. (1998). Clinical classification of glaucomatous visual field loss by frequency doubling perimetry. *American Journal of Ophthalmology*, 125, 830–836.
- Stillman, J. A. (1989). A comparison of three adaptive psychophysical procedures using inexperienced listeners. *Perception and Psychophysics*, 46, 345–350.
- Tyler, C. W. (1974). Observations on spatial-frequency doubling. *Perception*, 3, 81–86.
- Tyler, C. W. (1981). Specific deficits of flicker sensitivity in glaucoma and ocular hypertension. *Investigative Ophthalmology and Visual Science*, 20, 204–212.
- Tyler, C. W. (1989). Two processes control variations in flicker sensitivity over the life span. *Journal of the Optical Society of America A*, 6, 481–490.
- Tyler, C. W., Ryu, S., & Stamper, R. (1984). The relation between visual sensitivity and intraocular pressure in normal eyes. *Investigative Ophthalmology and Visual Science*, 25, 103–105.

- Tyrrell, R. A., & Owens, D. (1987). A new technique to rapidly assess the resting states of eyes and other threshold phenomena: The Modified Binary Search (MOBS). *Behavior Research Methods, Instruments and Computers*, 20, 137–141.
- Victor, J. D. (1988). The dynamics of the cat retinal Y cell subunit. *Journal of Physiology*, 405, 289–320.
- Victor, J. D., & Shapley, R. M. (1979). The nonlinear pathway of Y ganglion cells in the cat retina. *Journal of General Physiology*, 74, 671–689.
- Yager, D., & Beard, B. L. (1994). Age differences in spatial contrast sensitivity are not the result of changes in subjects criteria or psychophysical performance. *Optometry and Visual Science*, 71, 778–782.