

EXTENDED REPORT

Clinical evaluation of frequency doubling technology perimetry using the Humphrey Matrix 24-2 threshold strategy

P G D Spry, H M Hussin, J M Sparrow

Br J Ophthalmol 2005;89:1031–1035. doi: 10.1136/bjo.2004.057778

See end of article for authors' affiliations

Correspondence to:
Dr Paul G D Spry, Bristol Eye Hospital, Lower Maudlin Street, Bristol BS1 2LX, UK; paul.spry@ubht.swest.nhs.uk

Accepted for publication 1 December 2004

Aims: To evaluate performance of frequency doubling technology (FDT) perimetry using the Humphrey Matrix 24-2 thresholding program in a hospital eye service (HES) glaucoma clinic.

Methods: A random sample of individuals referred consecutively to the HES for suspected glaucoma were examined with 24-2 threshold FDT in addition to routine clinical tests. The discriminatory power of FDT and standard automated perimetry (SAP) were assessed using glaucomatous optic nerve head appearance as the reference gold standard.

Results: 48 of 62 eligible referred individuals were recruited. Glaucoma prevalence was 31%. Median test duration per eye was 5 minutes 16 seconds for FDT and 5 minutes 9 seconds for SAP. There was no significant difference ($p=0.184$) between proportions of individuals with reliable test results (FDT 75%, SAP 63%). Using a clinically appropriate binary criterion for abnormal visual field, sensitivity and specificity levels were 100% and 26% respectively for FDT and 80% and 52% for SAP. Both tests had higher negative than positive predictive values with marginal differences between tests. Criterion free receiver operator characteristic analysis revealed minimal discriminatory power differences.

Conclusions: In a HES glaucoma clinic in which new referrals are evaluated, threshold 24-2 FDT testing with the Humphrey Matrix has performance characteristics similar to SAP. These findings suggest threshold testing using the FDT Matrix and SAP is comparable when the 24-2 test pattern is used.

Frequency doubling technology (FDT) perimetry was introduced in 1997¹ and tests contrast sensitivity using spatial frequency doubled stimuli obtained with low spatial frequency stimulus undergoing counterphase flicker at a high temporal frequency. First generation instrumentation utilised 10° targets with suprathreshold screening strategies or thresholding algorithms. Stimulus arrays with up to 19 stimuli within the central 30° of the visual field were available. Reasonable performances have been obtained in evaluations of FDT performance in population screening^{2,3} and hospital glaucoma clinic^{4–6} environments. Given the relatively large size and low density of this FDT stimulus array, changes in the instrument design—specifically increases in the target spatial resolution—may improve visual field defect profile description. Prototype instrumentation using more test locations and a smaller stimulus size has been described, attaining a stimulus resolution increase. With this instrumentation FDT testing using a stimulus pattern equivalent to the Humphrey field analyser 24-2 test pattern resulted in a positive impact on discriminatory power for detection of initial glaucomatous visual field loss.⁷ A second generation instrument using similar small FDT stimuli, the Humphrey Matrix, became available for clinical use in 2003. To date, however, scant clinical data are available describing this instrument's performance.

The aim of this investigation was to evaluate the performance of threshold FDT perimetry using the Humphrey Matrix 24-2 thresholding program in a routine hospital eye service glaucoma clinic environment where referred glaucoma suspects are examined for the first time. Three specific aspects of performance were evaluated by comparison with standard automated perimetry (SAP), (1) the proportion of reliable test results; (2) test duration; and (3) clinical diagnostic capability.

METHODS

Data were collected during routine outpatient clinics at Bristol Eye Hospital. Relevant ethical and institutional approval was obtained. All patients gave informed written consent.

Study design and sample selection

A prospective case series design was employed. The reference population was individuals referred to the hospital eye service (HES) from any source because of suspected glaucoma. The sampling frame comprised all glaucoma suspect referrals to Bristol Eye Hospital. Such individuals are examined in specific "new glaucoma patient" clinics. Twenty five per cent of patients attending eight consecutive new patient clinics between October 2003 and January 2004 (simple random sample) were selected. This proportion was determined by physical ability to test participants within the clinic duration.

Clinical examination

The Humphrey Matrix (Welch Allyn, Skaneateles Falls, NY, USA and Carl Zeiss Meditec, Dublin, CA, USA) test was performed in addition to standard ophthalmic examination within a routine outpatient clinic environment. Standard examination comprised assessment of corrected Snellen visual acuity and SAP using the Humphrey field analyser (HFA) Program 24-2 SITA-Fast (Carl Zeiss Meditec, Dublin, CA, USA) followed by ocular examination (anterior segment

Abbreviations: FDT, frequency doubling technology; GON, glaucomatous optic neuropathy; HES, hospital eye service; MD, mean deviation; NTG, normal tension glaucoma; OHT, ocular hypertension; ONH, optic nerve head; PDF, probability density function; POAG, primary open angle glaucoma; PSD, pattern standard deviation; PXF, pseudoexfoliative glaucoma; ROC, receiver operator characteristic; SAP, standard automated perimetry

examination, gonioscopy, Goldmann applanation tonometry, dilated posterior segment examination including optic nerve head examination with Volk binocular indirect ophthalmoscopy). The SITA-Fast thresholding strategy was employed in this study as it is the default in the clinic.

The FDT Matrix threshold program 24-2 was carried out either before or after SAP. FDT and SAP were performed in no particular order. A rest interval was provided between tests. Both visual field examinations tested the right eye first.

Stimulus characteristics of the Matrix FDT Program 24-2 test are similar to those previously described for the prototype 24-2 FDT device⁷ being 5° square with spatial frequency of 0.5 cycles/° and temporal frequency of 18 Hz. A ZEST thresholding strategy is used with a flat previous probability density function (PDF) and fixed termination criterion.⁸

SAP tests were performed by a pool of clinic staff trained in visual field testing. Subjects wore near spectacle correction if appropriate. For patients habitually wearing bifocal, varifocal, or tinted spectacles, full aperture trial lenses were used.

FDT tests were performed by one investigator (HMH). Individuals wore their distance refractive correction where appropriate. For patients habitually wearing bifocal, varifocal, or tinted spectacles, wide aperture correcting lenses designed for perimetry were used.⁹

The results of the first type of visual field tests (FDT or SAP) were not available to the individual operating the second test type.

Case definition

Case definition was selected to be independent of visual function to permit comparison of FDT with SAP. Individuals were classified as either glaucoma present or absent by identification of optic nerve head (ONH) signs consistent with glaucomatous optic neuropathy (GON) in either eye when examined by a single experienced consultant ophthalmologist (JMS) with a specialist interest in glaucoma. Diagnosis was by patient. The examining ophthalmologist was masked to FDT Matrix test results, but had access to SAP results.

Statistical analysis

Analysis of the data was split into three parts.

Reliability

Visual field test reliability was quantified by patient, not by eye. Standard outcome criteria for reliability were adopted:

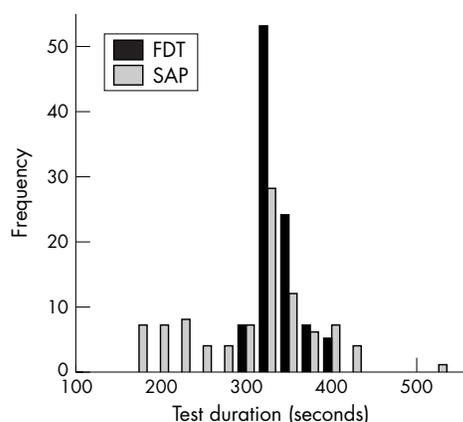


Figure 1 Bar chart showing distribution of visual field test durations for both FDT (solid bars) and SAP (shaded bars) pooled for both eyes of all recruited patients (n=48). Median durations were 316 seconds and 309 seconds for FDT and SAP respectively, with hypothesis testing (signed rank test) revealing that the difference was significant ($p < 0.001$).

fixation losses $< 25\%$ and both false positive and false negative responses $< 33\%$. Hypothesis testing for differences in number of unreliable individuals between the tests used analysis of paired proportions.

Test duration

Test times were quantified as time taken per eye (seconds) to perform the examination in each eye. Paired data were compared by eye.

Clinical diagnostic ability

Two analyses were used to evaluate this aspect of test performance. The first quantified performance of both FDT and SAP visual field tests using a binary outcome criterion to define abnormal visual field test result. This approach enables estimation of discriminatory power (sensitivity and specificity) and predictive values. This analysis approach is entirely dependent upon the criterion selected to denote abnormality. The criterion was therefore pragmatically selected in order to be generalisable to that used in a typical clinical environment, consisting of glaucoma hemifield test "outside normal limits" and/or $p < 0.05$ with the pattern standard deviation (PSD) global index in one or both eyes. The same criterion was used to dichotomously categorise both SAP and FDT results, and therefore it is important to note that the methodological derivation of these indices is identical for FDT and SAP. Note that normative data collection methodology for the matrix is available in the literature.¹⁰

The second analysis approach was criterion-free analysis of continuous variables using receiver operator characteristic (ROC) curves. This approach was used to evaluate mean deviation (MD), and PSD and was performed by eye.

Statistical tools

Analysis of proportions was performed by hand using statistical tables.¹¹ Parametric and non-parametric comparison of paired proportions used SigmaStat 2.0 (Jandel Corporation, San Rafael, CA, USA). ROC analysis of continuous variables versus binary case definition was performed using Intercooled Stata 7.0 (Stata Corporation, College Station, TX, USA).

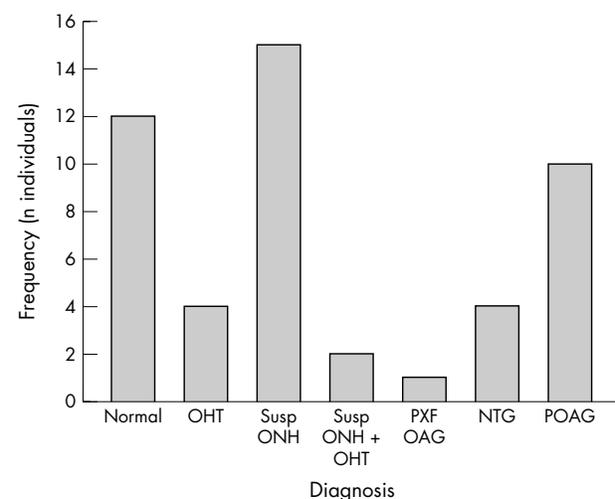


Figure 2 Frequency distribution of diagnosis among the sample (n=48). Individuals with glaucomatous optic neuropathy (n=15) came from three diagnostic categories, normal tension glaucoma (NTG), primary open angle glaucoma (POAG) and pseudoexfoliative glaucoma (PXF OAG). Suspects (n=21) comprised individuals with ocular hypertension (OHT) and those with suspicious optic nerve head (ONH) appearance.

RESULTS

Sixty two individuals were randomly selected for participation; 48 (77.4%) were recruited, with the major reason for non-participation being failure to attend the outpatient appointment (n = 9). Other reasons comprised dementia (n = 2), postural problems (n = 2), and deafness (n = 1).

The mean (SD) age of individuals recruited was 67.3 (13.5) years. The male:female ratio was 1:1. The age and sex characteristics of non-participants was similar to participants.

Information about test duration is provided in figure 1. Among the entire sample (both eyes data pooled), the distribution of test durations with FDT was approximately symmetrical, although SAP was negative skewed. The median test duration was 316 seconds for FDT and 309 seconds for SAP. Quantification of test duration spread by standard deviation (SD) revealed that this was greater for SAP than FDT, with SDs of 70 seconds and 24 seconds respectively. The difference in median durations was statistically significant (signed rank test, $p < 0.001$.)

A higher proportion of individuals exhibited reliable tests in both eyes with the Humphrey Matrix (36/48, 75%) than SAP (30/48, 63%), although this difference was not significant at $p < 0.05$ (test of paired proportions, $p = 0.184$). Of the 12 individuals who were unreliable with FDT, six of these were also unreliable with SAP.

Of the sample of 48 individuals, 15 (31.2%) were found to satisfy the glaucoma case definition (GON in one or both eyes), 21 (43.8%) were identified as suspects that required subsequent monitoring within the HES and 12 (25%) were normal. Further diagnostic information is provided in figure 2.

The results of analysis based on collapsing visual field data into a binary outcome of abnormal visual fields are given in table 1. Overall, it can be seen that both visual field tests exhibited higher sensitivity than specificity, with FDT exhibiting higher sensitivity and SAP higher specificity. Predictive values were consistent with these sensitivity and specificity estimates, as evidenced by negative predictive values exceeding positive predictive values for both tests. Overall, FDT had marginally higher predictive values than SAP.

In order to gain some insight in the performance of FDT at differing levels of visual loss on standard perimetry, the proportion of individuals with abnormal FDT at varying degrees of SAP abnormality was investigated (see fig 3). The proportion of individuals having abnormal FDT results increased with increasing SAP defect severity, quantified by MD.

It can be seen from criterion free analysis based on continuous variables (see fig 4) that performance of both test types for each variable substantially exceeded chance. The SAP area under ROC demonstrated minimally higher discriminatory performance than FDT.

DISCUSSION

This study provides evidence that performance of threshold 24-2 FDT perimetry is comparable to SAP 24-2 in the key

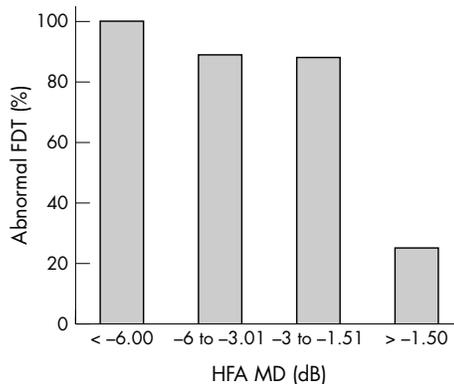


Figure 3 Bar chart showing proportion of sample with abnormal FDT test results stratified by standard automated perimetry (SAP) mean deviation (MD).

areas of proportion of reliable test results obtained, test duration, and diagnostic accuracy. The pragmatic experimental design allows findings to be generalised to hospital eye service (HES) clinics used for initial assessment of patients with suspected glaucoma.

In terms of reliability, the proportion of reliable test results by FDT perimetry (75%) was similar to that described for SAP in other reports of perimetrically naive patients.³ More patients were found to be reliable with FDT than SAP (62.5%), although this difference was not significant.

The sample comprised a variety of visual field defect magnitudes. Average test time for FDT was just over 5 minutes. Although this was statistically significantly greater than SAP, it is reasonable to suggest that the

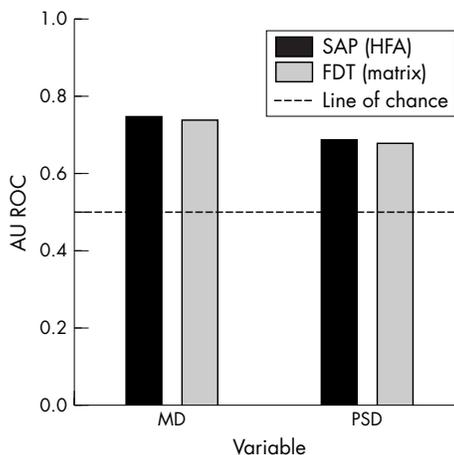


Figure 4 Results of criterion free (ROC) analysis for global indices mean deviation (MD) and pattern standard deviation (PSD) for both FDT and SAP.

Table 1 Sample estimates of discriminatory power and predictive values for SAP and FDT to detect individuals classified as having glaucomatous optic neuropathy in either eye

	SAP (HFA 24-2 Sita-Fast)	FDT (Matrix 24-2)
Sensitivity, % (95% CI)	80 (44 to 100)	100 (72 to 100)
Specificity, % (95% CI)	52 (38 to 66)	26 (14 to 38)
Positive predictive value (95% CI)	0.26 (0.14 to 0.38)	0.31 (0.18 to 0.44)
Negative predictive value (95% CI)	0.93 (0.86 to 1.00)	1.00 (0.72 to 1.00)

difference of 7 seconds is not clinically important. This test duration should be contrasted with reports from both threshold strategies of the first generation FDT device, such as program C20 (seventeen 10° test targets within central 20°) and also the prototype 24-2 FDT device ("Quadravision"). Reports from first generation C20 FDT testing in glaucoma patient groups of varying size and composition described average test times between 4 minutes and 6 minutes,^{5 12-14} therefore, similar to our data with the threshold program 24-2 on the Matrix device, despite lower spatial stimulus resolution. Data collected with the Quadravision prototype 24-2 FDT device yielded an average test time in 23 patients with early glaucomatous visual field loss of approximately 12 minutes.⁷ The differences between these devices and the Matrix FDT may be attributed to use of a thresholding strategy whereby both first generation FDT perimeter and the prototype 24-2 device employ a modified binary search (MOBS) thresholding strategy¹⁵ unlike the matrix, which uses a Bayesian strategy, specifically ZEST,¹⁶ with a fixed number of presentations at each test location and a flat previous probability density function (PDF). Both simulated and empirical laboratory FDT data show that use of ZEST, albeit with a dynamic termination criterion, can produce results of similar accuracy and reliability in approximately 50% of test time.^{17 18} Furthermore, fixed termination criterion use does not impact on detection of threshold inaccuracies,¹⁹ although the effect of a flat PDF remains unknown.

In terms of diagnostic accuracy FDT and SAP produced broadly similar results, yielding considerably higher sensitivity than specificity. Obviously, alternative outcome criteria could have been applied: this particular criterion being selected because it was considered typical of that applied by clinic staff. The impact of this criterion demonstrated that it was relatively liberal (that is, higher sensitivity). Criterion free analysis obtained minimal differences in discriminatory power between the two test types. It is possible, however, that this may have differed if the SITA-Standard thresholding algorithm had been used for SAP, although this would have increased duration. It is also important to consider that a potential source of differential information bias existed in favour of SAP because the examining ophthalmologist had access to the results of this test.

Although considerable published data exist on performance of the first generation FDT perimeter, to our knowledge there are no reports on clinical performance of the Humphrey Matrix in the literature. Many research groups have reported on the clinical performance of the first generation FDT screening (suprathreshold) tests with the broad consensus of high levels of discriminatory power for identification of glaucomatous visual field loss, which appears to be associated with the degree of achromatic field loss.¹⁴ Specifically, some investigators have described reasonable discriminatory power in population screening studies for the presence of any abnormality likely to cause a visual field defect.³ A glaucoma population screening study obtained optimal levels of FDT discrimination of 92% sensitivity and 93% specificity.² Similarly, other studies report good discriminatory performance with the same screening tests in hospital glaucoma clinic environments,⁴⁻⁶ although these studies may have limited generalisability owing to differing selection criteria. Fewer data are available on threshold strategies of this instrument, with a single available comparative report of suprathreshold and full threshold strategies suggesting that threshold FDT perimeter improved discriminatory power compared with suprathreshold screening tests.⁵ Specifically, for threshold 24-2 FDT, a cross sectional study performed with the prototype device found a greater proportion of abnormal test locations with threshold 24-2 than threshold C20, both in individuals with early

glaucoma and high risk glaucoma suspects, implying modest sensitivity improvements. The shape of visual field defects was also found to be better characterised.⁷

In the context of population screening, the aim of an initial or primary screening round is to identify as many individuals with the disease as possible while taking into account the costs and benefits of disease detection, capability of the continuing care system, and disease prevalence. In the context of a low prevalence disease such as chronic open angle glaucoma where the course of untreated disease is generally slow and repeated testing is feasible, it can be argued that selection of primary population screening tests and abnormality criteria should compromise sensitivity to achieve optimal specificity and positive predictive values. The purpose of this is to avoid excessive demands (high false positive rates) upon referral centres. However, for secondary screening of enriched prevalence populations, such as those failing a primary screen, the priority becomes identification of all cases of disease within the re-screened group, therefore demanding optimal sensitivity and negative predictive value—that is, compromised specificity. With this approach in mind, the performance of FDT in this study's enriched glaucoma prevalence population suggests suitability for use in a secondary screening environment, such as initial assessment of glaucoma suspects within the HES, or referral refinement schemes.²⁰ The estimations of negative predictive values and positive predictive values suggest that FDT is at least as valuable in this capacity as SAP.

Data obtained in this study suggest that threshold 24-2 FDT has desirable characteristics for initial glaucoma detection in HES environments. However, at the present time care should be taken not to interpret this as suggesting that this technology may have a role in monitoring individuals for progressive glaucoma and therefore longitudinal data upon sensitivity to change are awaited with interest.

Authors' affiliations

P G D Spry, H M Hussin, J M Sparrow, Bristol Eye Hospital, Lower Maudlin Street, Bristol BS1 2LX, UK

Commercial relationship: PGDS has received research funds from Welch-Allyn, Skaneateles, NY, USA. HMH and JMS have no commercial relationships.

REFERENCES

- 1 Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci* 1997;**38**:413-25.
- 2 Yamada N, Chen PP, Mills RP, et al. Screening for glaucoma with frequency-doubling technology and Damato campimetry. *Arch Ophthalmol* 1999;**117**:1479-84.
- 3 Cioffi GA, Mansberger S, Spry PGD, et al. Frequency doubling perimetry and the detection of eye disease in the community. *Trans Am Ophthalmol Soc* 2000;**98**:195-9.
- 4 Quigley HA. Identification of glaucoma-related visual field abnormality with the screening protocol of frequency doubling technology. *Am J Ophthalmol* 1998;**125**:819-29.
- 5 Burnstein Y, Elish NJ, Magbalon M, et al. Comparison of frequency doubling perimetry with Humphrey visual field analysis in a glaucoma practice. *Am J Ophthalmol* 2000;**129**:328-33.
- 6 Casson R, James B, Rubinstein A, et al. Clinical comparison of frequency doubling technology perimetry and Humphrey perimetry. *Br J Ophthalmol* 2001;**85**:360-2.
- 7 Johnson CA, Cioffi GA, Van Buskirk EM. Frequency doubling technology perimetry using a 24-2 stimulus presentation pattern. *Optom Vis Sci* 1999;**76**:571-81.
- 8 Johnson CA, Gunvant P, Fingeret M, et al. *A primer for frequency doubling technology (FDT) perimetry using the Humphrey Matrix*, Welch-Allyn, Zeiss, 2004.
- 9 Henson DB, Earlam RA. Correcting lens system for perimetry. *Ophthalmic Physiol Opt* 1995;**15**:59-62.
- 10 Anderson AJ, Johnson CA, Fingeret M, et al. Characteristics of the normal database for the Humphrey Matrix Perimeter. 2004.
- 11 Altman DG. *Practical statistics for medical research*. Boca Raton: Chapman Hall/CRC, 1991.
- 12 Kondo Y, Yamamoto T, Sato Y, et al. A frequency-doubling perimetric study in normal-tension glaucoma with hemifield defect. *J Glaucoma* 1998;**7**:261-5.

- 13 **Chauhan BC**, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci* 1999;**40**:648–56.
- 14 **Cello KE**, Nelson-Quigg JM, Johnson CA. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. *Am J Ophthalmol* 2000;**129**:314–22.
- 15 **Tyrell R**, Owens D. A rapid technique to assess the resting states of the eyes and other threshold phenomena: the modified binary search (MOBS). *Behav Res Meth Instr Comp* 1988;**20**:137–141.
- 16 **King-Smith PE**, Grigsby SS, Vingrys AJ, *et al*. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vis Res* 1994;**34**:885–912.
- 17 **Turpin A**, McKendrick AM, Johnson CA, *et al*. Performance of efficient test procedures for frequency-doubling technology perimetry in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2002;**43**:709–15.
- 18 **Turpin A**, McKendrick AM, Johnson CA, *et al*. Development of efficient threshold strategies for frequency doubling technology perimetry using computer simulation. *Invest Ophthalmol Vis Sci* 2002;**43**:322–31.
- 19 **Anderson AJ**. Utility of a dynamic termination criterion in Bayesian adaptive threshold procedures. In: Henson DB, Wall M, eds. *Perimetry update 2002/3*. The Hague, Netherlands: Kugler Publications, 2004:191–7.
- 20 **Henson DB**, Spencer AF, Harper R, *et al*. Community refinement of glaucoma referrals. *Eye* 2003;**17**:21–6.