Diagnostic Performance of Visual Field Test Using Subsets of the 24-2 Test Pattern for Early Glaucomatous Field Loss

Yanfang Wang and David B. Henson

Purpose. To explore the diagnostic performance of threshold visual field tests using subsets of the standard 24-2 test pattern in detecting early/moderate glaucomatous field loss.

Methods. Normal (Brusini stage 0, $n = 2344$) and defective eyes (Brusini stage 2–3, $n = 2222$) from a database of visual field tests (6696 eyes/3586 patients, SITA standard 24-2 algorithm) were selected and resampled using a bootstrap method. The positive predictive values (PPVs) of each test location were calculated for the resampled datasets with a fail criteria of a single missed stimulus at a pattern deviation probability level of less than 0.01. Optimized test patterns started with the most frequent location of the maximum PPV in datasets. Eyes missing the location were removed and the PPV values of residual sample recalculated. The process was repeated until all defective eyes were detected. Receiver operating characteristic (ROC) curves were established for the PPV-optimized and five randomized patterns. Characteristics of visual field defects detected with subsets of optimized test pattern were established.

Results. With the PPV-optimized pattern, 95% of the field defects were detected with 30 locations and all with 43 locations. Areas under the ROC curve were greatest for the optimized pattern. With each increment in the number of test locations, the Mean Deviation of additionally detected eyes became more positive while Pattern Standard Deviation became less positive ($P < 0.001$).

Conclusions. Good diagnostic performance can be obtained with optimized subsets of the standard 24-2 test pattern that can provide substantial savings in test times. (*Invest Ophthalmol Vis Sci.* 2013;54:756–761) DOI:10.1167/iovs.12-10468

Glaucoma is one of the leading causes of blindness in the world, affecting an estimated 60.5 million people in 2010.[1] Measures of the IOP, the optic nerve head structure, and the visual field are widely used for the detection and management of glaucoma. Visual field measures quantify the functional loss and are important when evaluating a patient's quality of life and the effectiveness of treatment.[2–6]

The Swedish Interactive Threshold Algorithm (SITA) with the 24-2 distribution of test locations is the most widely used visual field test for the management of glaucoma. The 24-2 distribution is composed of 54 test locations distributed on a 6-degree square matrix displaced 3 degrees from the vertical and horizontal midlines. This pattern covers the central 24 degrees of the field with 2 additional stimuli at 27 degrees eccentricity above and below the horizontal midline in the nasal field. The 24-2 test has largely superseded the 30-2 test, which, in addition to the 54 test locations of the 24-2 test, has 22 extra test locations taking the total test area out to 30 degrees. The reduction in the number of test locations (76 to 54) reduced test times from approximately 6 to 5 minutes per eye (see Table 1). Additionally, the 24-2 was found to have a smaller variability than 30-2 test pattern and is less sensitive to artifacts due to the correcting lens or a droopy eye lid. These benefits have contributed to SITA 24-2 being the first choice for the detection and management of glaucoma in many institutions.

Whereas SITA 24-2 is widely used in the management of glaucoma, it is seldom used for routine screening in primary eye care. Suprathreshold tests are more commonly used for routine screening due largely to their shorter test times. The shorter test times of suprathreshold tests result from both a reduced number of presentations at each test location and, more often than not, a reduced number of test locations. The Humphrey, FDT, Octopus, and Henson 8000 have just 40, 17, 26, and 26 test locations in their basic glaucoma screening tests, respectively. In addition, the Octopus and Henson perimeters have a facility to extend a screening examination either by gaining more information from the existing test locations[11] or by increasing the number of test locations.[12] Being able to extend a test based on interim results allows rapid screening of suspect eyes with an option for more extensive testing in cases in which there is a screening failure. The criteria for a screening failure can be set to give high sensitivity, knowing that false positives, which tend to be high when the sensitivity of a screening test is high, will incur a simple extension of the test rather than reexamination with another test strategy.[12]

The informational value of each 24-2 test location at detecting glaucoma is likely to vary. Earlier work by Henson and colleagues[13,14] from an analysis of data obtained with the Friedmann Visual Field Analyzer found that stimuli in the extreme superior field (>20 degrees eccentricity) and around the blind spots give the least information, whereas stimuli in the superior and inferior arcuate areas give the most information. Removal of locations with low informational value is likely to have little effect on the diagnostic power of a visual field test but would shorten test times.

Interpoint topographic correlations of test locations in glaucoma[15] vary with test location. Several studies have explored the spatial pattern of interpoint correlation and found strong relationships between locations that follow the distribution of ganglion cell fibers.[16–18] Importantly, while the informational value of all locations within the arcuate regions may be high, the high correlation between locations might limit the value of additional test locations in these regions.

From the School of Biomedicine, University of Manchester, Manchester, United Kingdom.

Supported by the Manchester Academic Health Sciences Centre (MAHSC) and the NIHR Manchester Biomedical Research Centre.

Submitted for publication June 26, 2012; revised December 4, 2012; accepted December 13, 2012.

Disclosure: **Y. Wang**, None; **D.B. Henson**, None

Corresponding author: David B. Henson, Research Department, 4th Floor, Manchester Royal Eye Hospital, Oxford Road, Manchester, UK, M13 9WH; david.henson@manchester.ac.uk.

Table 1. Test Duration of SITA Standard Visual Field Test

Test Strategy	Test Program	Subjects	Test Time, min	No. of Test Locations	Average No. of Presentations (Range)	References
SITA standard	24-2	28	4.59/4.54	54	NA	Wall et al.[7]
	30-2	90	6.6	76	NA	Budenz et al.[8]
	30-2	20	6.1	76	287(273–321)	Bengtsson et al.[9]
	30-2	50	NA	76	325	Bengtsson et al.[10]

NA, not applicable.

In the present work, we explore the relative performance of using subsets of the standard 24-2 test pattern at detecting early/moderate glaucoma defects. It is hypothesized that a subset of the 24-2 test pattern could serve as a potential fast screening method to detect early/moderate glaucomatous field defects with a potential to extend this test to a full 24-2 test pattern on screening failure.

Methods

A database of the visual field test results (SITA 24-2 algorithm, Humphrey Visual Field Analyzer; Carl Zeiss Meditec Inc., Dublin, CA) from 6696 eyes of 3586 patients with suspicious/diagnosed glaucoma (age 66.0 ± 13.0 years old, collected at Manchester Royal Eye Hospital) were classified into seven perimetric stages with the Brusini staging method.[19] Eyes with stage 0 were defined as normal group ($n = 2344$) and eyes with stage 2 and 3 were selected as a defective group ($n = 2222$) for further analysis. Eyes graded as borderline and stage 1 were excluded from the analysis, as there is considerable overlap in these grades between normal and early glaucomatous eyes. Eyes with advanced visual field loss (stages 4 and 5) were also excluded from the analysis on the basis that this level of loss does not offer a suitable diagnostic challenge. We included all eyes that met these criteria and did not exclude eyes on the basis of reliability indices.

Positive predictive value (PPV) is a statistical parameter used to estimate the performance of a diagnostic test. It is defined as the ratio between the true positives (TP) and total positive calls (false positives [FP] plus TP).

To reduce any bias that might result from a single sample of visual field data, we used a bootstrap method (Matlab, version 2008; Mathworks Inc., Natick, MA) to generate 200 datasets from the original sample. For each dataset, 4566 visual field tests were randomly selected, with replacement, from the original dataset. For each of the 200 datasets the PPV of the 54 test locations was calculated using a cutoff criterion of pattern deviation probability less than 0.01 to define a location as being normal or defective. For each of the 200 datasets, the location with the highest PPV was then identified and the most frequent location with the highest PPV was selected for the optimized test pattern. Visual field results of eyes in which this abnormal location was missing were then removed in each of the 200 datasets and the PPV of the residual sample calculated. This process was repeated until all defective eyes in the set had been detected (see Fig. 1).

The diagnostic performance of the optimized test patterns (sensitivity and specificity) were then calculated at each step (i.e., for 1, 2, 3... n test locations), from the original dataset ($n = 2344$ normal and 2222 defective eyes). The sensitivity was defined as the proportion of the abnormal eyes detected with defects at the n test locations and the specificity was the proportion of normal eyes detected with no defects at the n test locations.

In addition, five random test patterns were generated to establish the benefits of using optimized distributions. The five random series of test locations were generated with a Matlab program (version 2008; Mathworks). Receiver operating characteristics (ROC) curves were plotted for both the PPV optimized pattern and randomized test patterns based on the sensitivity obtained at a specificity level of approximately 95%, 90%, 85%, 80%, 75%, 70%, and 60%.

To analyze the characteristics of visual field defects detected with increasing number of test locations, the average and SD of mean deviation (MD), pattern standard deviation (PSD), and total number of defective locations, were calculated for 1–10, 11–20, 21–30, and 31–40 optimized test locations from the original dataset. One-way analysis of variance (SPSS 16.0 for Windows; SPSS Inc., Chicago, IL) was used to compare means of those parameters between the defined location groups.

Results

Visual Field Test Performances

With the optimized test patterns, all defective eyes were detected with 43 test points. The sensitivity increased logarithmically with the number of test locations (more than 95% with just 30) while the specificity decreased linearly (99% for 1, 62.1% for 43) (Fig. 2).

Figure 3 gives the optimized distribution of 10, 20, 30, and 43 test locations for the defective group. Slightly more locations were observed in the superior hemifield when fewer locations were involved (e.g., 10 and 20 test locations). The 11 locations (2 in the blind spot) that did not contribute anything to the performance were distributed throughout the central field and often (9/11) fell within the arcuate areas (5 superior, 4 inferior).

Compared with the randomized patterns, the PPV-optimized pattern showed a better performance with a larger area under the ROC curve (see Fig. 4).

Characteristics of Visual Field Defects Detected with PPV-Optimized Locations

Figure 5 displays the characteristics (MD, PSD, and defect numbers) of glaucomatous eyes detected with PPV-optimized test locations. With the increasing number of test locations, MD became less negative, whereas PSD and defect numbers became less positive ($P < 0.001$). This finding highlights how more advance defects can be detected with fewer test locations.

Estimated Test Time for Subsets of 24-2 Pattern

Based on results from previous studies (see Table 1), the average SITA Standard test time for one test location is approximately 0.08 to 0.09 minutes in healthy subjects.[7–10] Using these figures, the test times, sensitivity, and specificity for a range of different test patterns has been calculated (see Table 2). For test patterns composed of 30 stimuli, test times would reduce to less than 3 minutes from an average of 4.6 minutes for all 54 test locations and for patterns of 20 stimuli to less than 2 minutes.

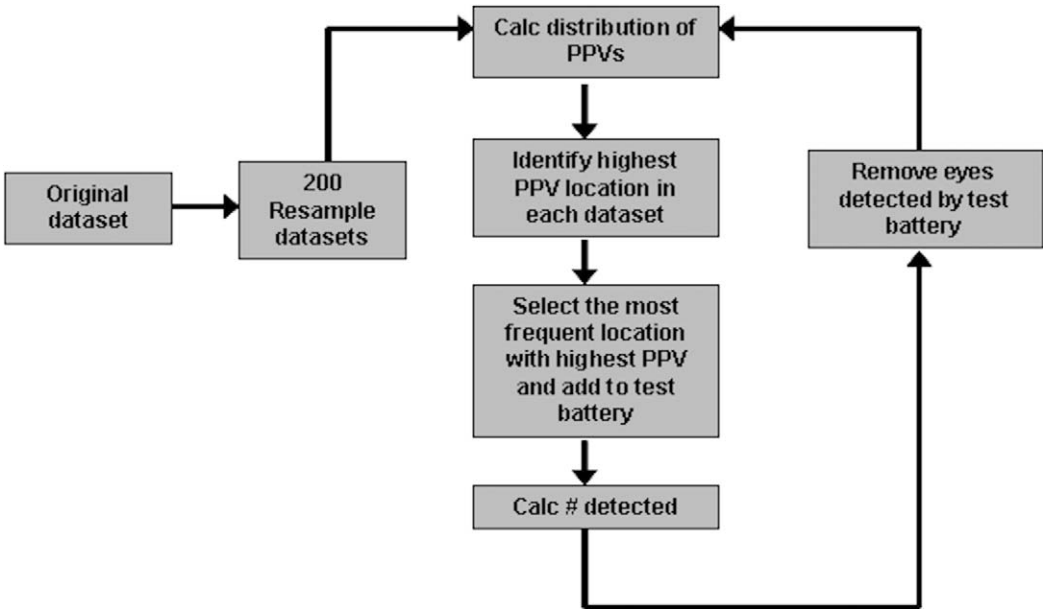

FIGURE 1. Flowchart of PPV calculation.

DISCUSSION

We have established the diagnostic performance of visual field tests using subsets of the 24-2 test pattern on eyes with mild/moderate glaucomatous loss (Brusini stages 2 and 3). This was done with optimized distributions (greatest PPV) and random distributions of the standard 24-2 test pattern. Using a cutoff criterion of a single missed location (pattern deviation probability value < 0.01), we were able to obtain high sensitivity with just 30 test locations (95.5%). The specificity for this number of test locations was 76.3%. In a protocol in which a patient is initially tested with a truncated test set with an extension to a full 24-2 test pattern on screening failure (1 missed location at a pattern deviation probability level of <0.01), this would reduce test times to approximately 2.5 minutes in approximately 75% of eyes with no visual field loss.

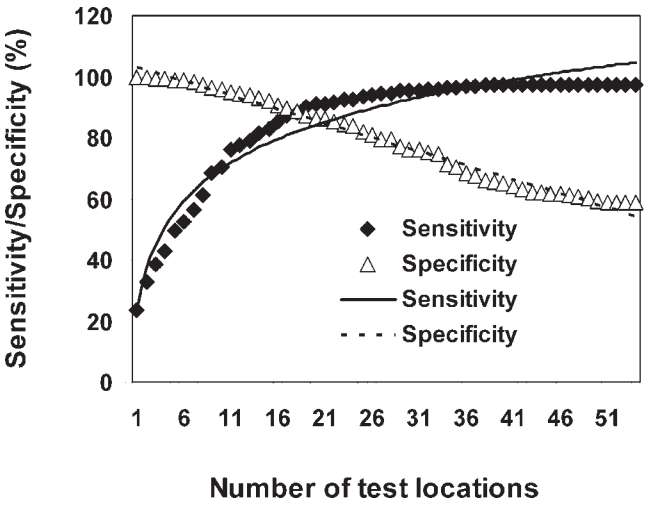

FIGURE 2. The sensitivity and specificity of the optimized test patterns with increasing numbers of test locations. The *solid line* represents a logarithmic relationship between the sensitivity and increased test locations and the *dashed line* represents a linear relationship between the specificity and increased test locations.

Such strategies would be particularly beneficial when examining those with ocular hypertension and those at risk of glaucoma (no visual field loss), since the usage of full 24-2 test pattern is not cost efficient for large population screening. Test times of 2.5 minutes per eye, while still being longer than suprathreshold tests with the same number of test locations,[20–22] make routine screening with threshold strategies much more viable with the added advantage that in screening failures there is a baseline 24-2 threshold test result (with extension) without having to reexamine the patient. Although this study has used a dataset obtained with the SITA Standard test algorithm, similar gains would be expected with alternative threshold algorithms (e.g., SITA Fast). Differences between SITA Standard and Fast relate to their terminating criteria[23] and reducing the number of test locations will have a proportional benefit on test times of both algorithms.

The measures of diagnostic accuracy derived in this study used two large population samples, those with Brusini stage 0 and those with Brusini stages 2 and 3. The Brusini staging system was used in this study because it combines the MD and PSD values, making it sensitive to both early localized loss and advanced loss. Excluding the stages of Borderline and Stage 1 from the analysis, although justifiable on the basis that it is very difficult with cross-sectional data to establish whether or not a test location missed at a pattern deviation value of less than 0.01 is the result of normal variability or an early defect, will lead to overestimates of the diagnostic accuracy of the optimized patterns when used on populations including these stages. Nevertheless, because the classification was based on

TABLE 2. Test Time Estimated for Variable Subsets of the Optimized Test Pattern

No. of Test Locations	Sensitivity, %	Specificity, %	Test Time Estimated, min*
10	70.2	96.0	0.8 to 0.9
20	91.0	86.2	1.6 to 1.8
30	95.5	76.3	2.4 to 2.7
54	97.4	58.6	4.3 to 4.9

* Values derived from literature.[7–10]

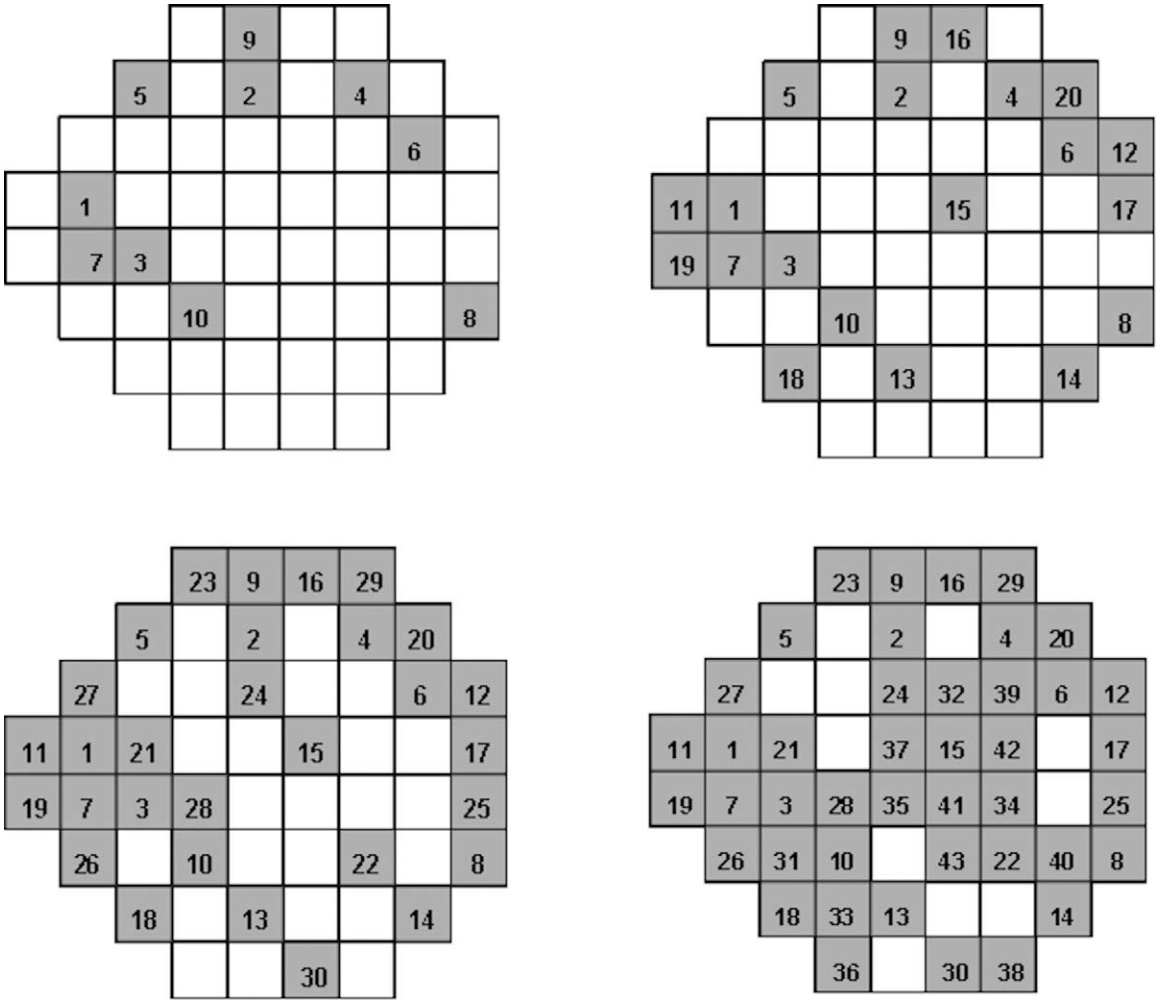

Figure 3. Optimized distributions consisting of 10, 20, 30, and 43 test locations. Numbers in the *box* indicate the *i*th test location involved in the pattern.

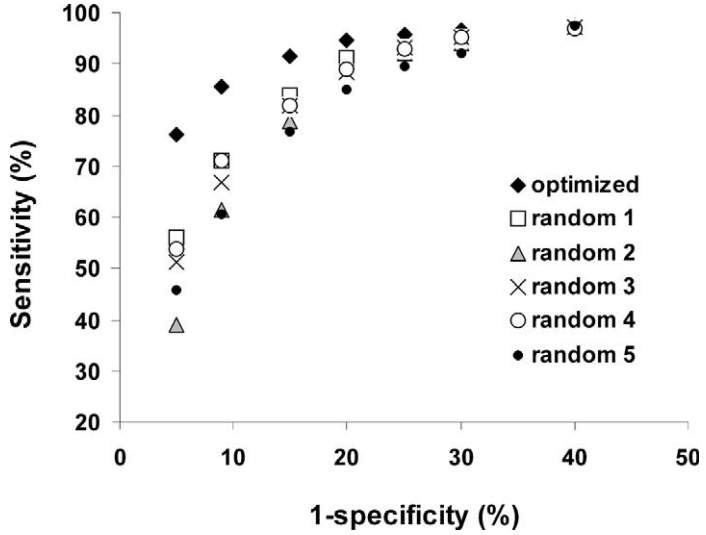

Figure 4. ROC curves of visual field test with the optimized and five randomized test location patterns.

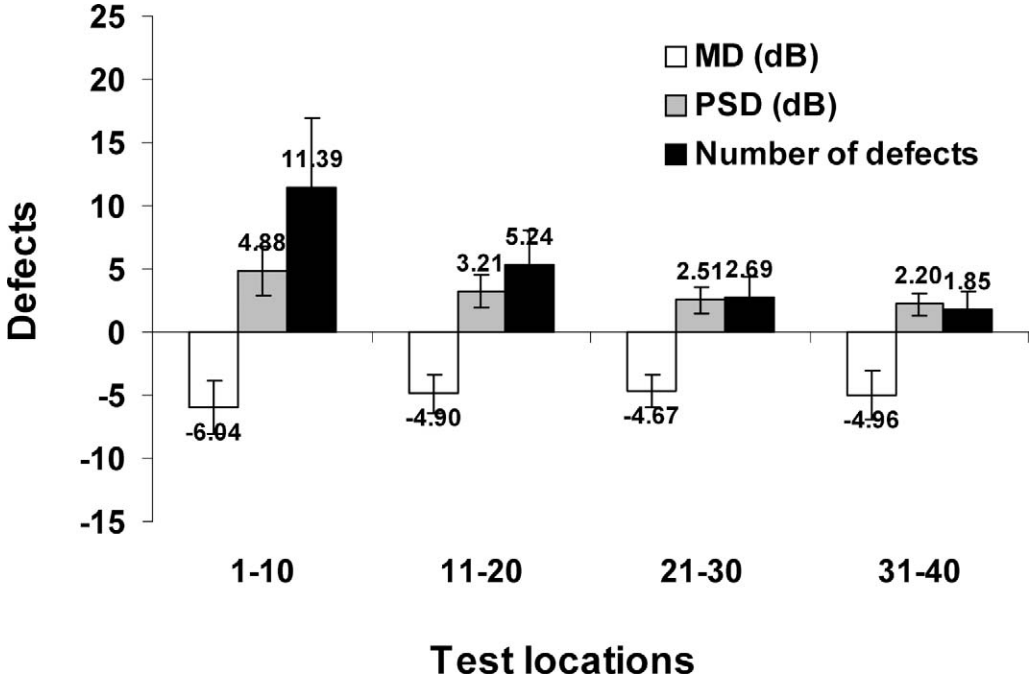

Figure 5. The mean deviation (MD), pattern SD (PSD), and number of defective locations detected with every 10 locations following the optimized test location patterns.

single visual field test results, the eyes defined as "normal" in this study cannot fully exclude those that could be abnormal on subsequent testing or have an early undetectable functional loss. Excluding eyes with Brusini stages 4 and 5 will lead to an underestimate of the optimized pattern sensitivity when testing populations that include these stages.

In our study, the relationship between the sensitivity and the number of test locations was well fitted with a logarithmic curve. Simultaneously, the specificity was linearly related to the number of test stimuli. These findings are consistent with a previous study,[24] as were the findings of good performance with comparatively few test locations.[24–27]

The use of optimized test locations based on the PPV of each test location and an experimental design that accounted for interpoint correlations (removing cases detected with prior distribution) was found to be superior to the use of random distributions of less than approximately 35 stimuli. Once above this number there was little difference between the two designs, a finding consistent with that of Henson et al.[24] using optimized versus an even distribution of stimuli. Examination on the residual PPV maps reveals that after 35 stimuli have been chosen, the variance of the PPV values is small and hence there is no real benefit in selecting locations on the basis of this map or randomly.

The finding that some locations in the arcuate areas contribute little to the diagnosis performance reflects the strong interpoint correlations within this region of the visual field. With such correlations, adding neighboring test points provides little additional value.

The nature of current threshold tests places high demands on the patient's ability to maintain vigilance. In a recent study using pupil dynamics to derive an estimate of vigilance during a perimetric test, it was found that most patients start to lose vigilance after approximately 3 minutes of testing.[28] After this time, their response variability increases. Using subsets of the 24-2 pattern with reduced test times is, therefore, likely to have the added advantage of reducing variability in responses.

In conclusion, good diagnostic performance can be obtained with optimized subsets of the current 24-2 stimulus pattern. Such patterns could be valuable when testing suspect eyes (those with no known visual field loss) and for screening large populations in which test times with the complete 24-2 pattern are not cost effective.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90:262–267.
2. Musch DC, Lichter PR, Guire KE, Standardi CL. The collaborative initial glaucoma treatment study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. 1999;106:653–662.
3. The AGIS investigators. The advanced glaucoma intervention study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials*. 1994;15:299–325.
4. Leske MC, Heijl A, Hyman L, Bengtsson B. Early manifest glaucoma trial: design and baseline data. *Ophthalmology*. 1999;106:2144–2153.
5. CNTGS Group. Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. *Am J Ophthalmol*. 1998;126:487–497.
6. Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120:701–713.
7. Wall M, Punke SG, Stickney TL, et al. SITA standard in optic neuropathies and hemianopias: a comparison with full threshold testing. *Invest Ophthalmol Vis Sci*. 2001;42:528–537.
8. Budenz DL, Rhee P, Feuer WJ, et al. Sensitivity and specificity of the Swedish interactive threshold algorithm for glaucoma-

tous visual field defects. *Ophthalmology*. 2002;109:1052–1058.

9. Bengtsson B, Heijl A, Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. *Acta Ophthalmol Scand*. 1998;76:165–169.
10. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997;75:368–375.
11. Interzeag AG. *OCTOPUS: Visual Field Digest*. 4th ed. Koeniz, Switzerland. Interzeag AG; 1998;22–31.
12. Henson DB. Visual field screening and the development of a new screening program. *J Am Optom Assoc*. 1989;60:893–897.
13. Henson DB, Dix SM, Oborne AC. Evaluation of the Friedmann Visual Field Analyser Mark II. Part 1. Results from a normal population. *Br J Ophthalmol*. 1984;68:458–462.
14. Henson DB, Chauhan BC. Informational content of visual field location in glaucoma. *Doc Ophthalmol*. 1985;59:341–352.
15. Gonzalez de la Rosa M, Gonzalez-Hernandez M, Abraldes M, Azuara-Blanco A. Quantification of interpoint topographic correlations of threshold values in glaucomatous visual fields. *J Glaucoma*. 2002;11:30–34.
16. Chauhan BC, Henson DB, Hobley AJ. Cluster analysis in visual field quantification. *Doc Ophthalmol*. 1988;69:25–39.
17. Heijl A, Lindgren A, Lindgren G. Inter-point correlations of deviations of threshold values in normal and glaucomatous visual fields. In: Heijl A, ed. *Perimetry Update 1988/1989*. Amsterdam: Kugler & Ghedini; 1988:177–183.
18. Crabb DP, Fitzke FW, McNaught AI, Hitchings RA. A profile of the spatial dependence of pointwise sensitivity across the glaucomatous visual field. In: Wall M, Heijl A, eds. *Perimetry updates 1996/1997*. Amsterdam: Kugler; 1996:301–310.
19. Brusini P, Filacorda S. Enhanced Glaucoma Staging System (GSS 2) for Classifying functional damage in glaucoma. *J Glaucoma*. 2006;15:40–46.
20. Henson DB, Artes PH. New developments in supra-threshold perimetry. *Ophthalmic Physiol Opt*. 2002;22:463–468.
21. Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci*. 1997;38:413–425.
22. Fingeret M, Johnson C. The new Humphrey Matrix FDT. *Optometric Management*. 2003;1–8.
23. Bengtsson B, Heijl A. SITA fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand*. 1998;76:431–437.
24. Henson DB, Chauhan BC, Hobley A. Screening for glaucomatous visual field defects: the relationship between sensitivity, specificity and the number of test locations. *Ophthalmic Physiol Opt*. 1988;8:123–127.
25. Sugimoto K, Schötzau A, Bergamin O, Zulauf M. Optimizing distribution and number of test locations in perimetry. *Graefes Arch Clin Exp Ophthalmol*. 1998;236:103–108.
26. Krakau CET. Visual field testing with reduced sets of test points: a computerized analysis. *Doc Ophthalmol*. 1989;73:71–80.
27. Gonzalez de la Rosa M, Reyes JA, Gonzalez Sierra MA. Rapid assessment of the visual field in glaucoma using an analysis based on multiple correlations. *Graefes Arch Clin Exp Ophthalmol*. 1990;228:387–391.
28. Henson DB, Emuh T. Monitoring vigilance during perimetry with pupillography. *Invest Ophthalmol. Vis Sci*. 2010;51:3540–3543.