

Background

Standard automated perimetry (SAP) remains an essential clinical assessment of visual function in patients suffering from glaucoma¹. A number of threshold estimation algorithms have been developed to achieve accurate estimation of the differential light threshold in an acceptable test time, thereby reducing the influence of fatigue²⁻⁹ on the results. Such algorithms exhibit greater efficiency in threshold estimation since they are based on maximum likelihood theory which results in fewer stimulus presentations compared with predecessor algorithms. The SITA algorithm used in the Humphrey Field Analyser can be considered as a 'gold standard' reference to other perimeters as extensive visual field research has been carried out using this instrumentation. Threshold variability, defined by both short-term and long-term fluctuations can influence the statistical and clinical analysis of visual fields, particularly in measures of focal visual field loss.

A different approach to achieving efficient threshold estimation has been developed by Gonzalez de la Rosa and Gonzalez-Hernandez¹⁰. The SPARK Precision algorithm is a fast threshold strategy intended to use the information from a limited number of test points to estimate threshold values in the central visual field within a short duration of examination. It uses the statistical relationships between neighbouring test points of the visual field which are in sectors which are thought to contain the same ganglion cell fibre bundles which correspond to morphologically and functionally defined regions^{11, 12}. In brief, the model was derived from a sample of more than 90,000 visual field examinations (originating from 14,923 patients) obtained using the G1 program (TOP strategy, Octopus 1-2-3 perimeter). Four threshold estimates are determined in four different phases and a final threshold estimate is then averaged for each testing point¹⁰. By averaging the thresholds, a reduction in threshold fluctuation of approximately 40% has been reported in patients in the early stages of glaucoma¹³ and test durations of less than three minutes have been achieved¹⁴. The initial phase of the SPARK Precision algorithm (also referred to as SPARK training) can be completed in less than 40 seconds¹⁵ and SPARK Precision has exhibited good diagnostic

sensitivity and specificity in glaucoma patients even when utilising the first phase results only¹⁵. Moreover, the agreement of SPARK Precision with morphologic indices¹⁵ enabled its threshold sensitivity to be used to predict the thickness of retinal nerve fibre layer¹⁶. The objective of this study was to evaluate the mean sensitivity (MS) and point-wise analysis of threshold values of SPARK Precision in comparison with the Humphrey Field Analyser in a normal population.

Methods

A total of 83 normal subjects (49 male, mean age 39.6 years, SD 16.1; range 20-71) were recruited from the optometry clinics at SEGi University, Malaysia. Inclusion criteria were absence of any ocular disease, any impairment of the transparency of ocular media, intraocular pressure (IOP) \leq 21mmHg, a previously normal visual field, normal optic disc appearance i.e. no localised neuroretinal rim loss, an absence of optic disc haemorrhage, notches or pallor, a cup/disc ratio \leq 0.6 or cup/disc asymmetry \leq 0.2, and no nerve fibre layer defects or a positive family history of glaucoma. Additional inclusion criteria were best corrected visual acuity (BCVA) of 6/6 or better; refractive errors below \pm 6.00DS spherical error and less than 2.50DC astigmatism, no history of intraocular surgery or other ocular diseases that could affect the visual field. All participants were free from any systemic disease that could affect ocular health such as diabetes mellitus (DM) and hypertension (HT). Participants were asked to refrain from caffeine, alcohol and nicotine use for a minimum of two hours prior to their examination. The study was approved by the Aston University Research Ethics Committee (ID 755) and adhered to the tenants of the Declaration of Helsinki.

Data collection:

Following written informed consent, all subjects underwent two visual field assessments on two separate days. At each visit, one randomly selected eye was tested in succession with

the Oculus Twinfield 2 (Haag-Streit, Wetzlar, Germany) using the SPARK Precision algorithm and with the HFA (Carl Zeiss Meditec, Dublin, CA) using the 30-2 SITA Standard algorithm. Both instruments use a background luminance of 10cdm^{-2} and have a maximum stimulus luminance of 10,000 asb. The 30-2 test pattern used in the HFA has a total of 76 test points covering the central 30 degree field with a square grid of 6 degree separation¹⁷. The SPARK Precision algorithm deploys a similar grid with a total of 66 test points ($30^\circ \times 24^\circ$) with the uppermost and bottommost rows and two points located in the blind spots missing compared to the SITA 30-2 test grid¹⁰. For the purpose of this study all participants were examined with the four phases of the SPARK Precision algorithm. In the first phase the 66 threshold values corresponding to the 66 test points are estimated by directly examining only six points, one in each functional region¹⁵. Thresholds for all other locations are determined by interpolation, but ultimately each of the 66 points is assessed at least once. In the following 3 phases the threshold values of the 66 point grid get refined by testing 21 points (per phase) of the six regions (see figure 1) and subsequent adjusting of neighbouring points through further interpolation. The final threshold value at each location is determined by calculating the median of the four threshold values derived at each of the four phases.

Participants were given a break of 10 minutes between the two tests. The order of SITA and SPARK Precision testing was randomised among subjects to minimise learning effects but remained constant between the two visits of a given individual. Consideration of the first and second visits enabled comparison of threshold variability within the SITA and SPARK Precision strategies. The results of the second visit only were used to make a between strategy comparison because they are less influenced by the learning effect^{14, 15}.

Data analysis:

Test results with poor reliability criteria, defined as false positives (FP) or negatives (FN) > 20% and fixation losses (FL) > 30% with SITA and FP>20% and FL>30% based on SPARK Precision; since it does not measure FN, were excluded from data analysis, as were any

fields which exhibited a defect in consecutive visits according to Hodapp-Parrish-Anderson classification of glaucoma¹⁹.

A comparison of threshold variability within each of the SITA and SPARK Precision strategies was made by carrying out a Bland-Altman analysis for the 66 stimulus locations in SPARK Precision and the 66 stimulus locations in SITA which were coincident with the SPARK Precision test pattern. To make a between strategy comparison, the mean sensitivities of the 66 points test grid of SPARK Precision and matching stimulus locations in the SITA test grid from the second visit were calculated for each patient and a point-wise analysis between-strategies was carried out to determine the threshold agreement between the two strategies. The results from the test points of the uppermost and bottommost rows and two points located at the blind spot for SITA were excluded. The central threshold in SPARK Precision was also excluded in the calculation of mean sensitivity. The left eye results were transposed into right eye format for the purpose of the analysis. The association between the mean sensitivity and age was determined for each strategy and test duration of each strategy was recorded and compared between the strategies.

Differences in threshold estimation between visits were completed for each method separately. Differences between the two methods were evaluated using point-wise analysis for each stimulus location (using only the 66 corresponding data points and only VF data obtained at visit 2 of each participant). A comparison between the two methods of the six regions according to the functional map described by Gonzalez de la Rosa et al.¹⁷ was made and additionally in terms of three eccentric test zones defined as central ($< 10^\circ$), mid-peripheral ($10^\circ - 20^\circ$) and peripheral ($>20^\circ$).

Statistical Analysis:

Normality of the threshold data was determined by the Shapiro-Wilk test and the Wilcoxon Signed Ranks test which was used for between-strategy comparisons. Bland-Altman plots were used to determine the agreement of the mean sensitivities within and between both

testing algorithms and regression testing was conducted to determine proportional bias. All correlations were determined using Spearman's correlation coefficient. A one-way ANOVA was used for comparison among the test zones followed with Tukey's HSD post-hoc analysis. While this constitutes multiple comparisons, we did not apply any correction on the basis of the exploratory nature of these particular analyses as recommended by Armstrong²⁰. Statistical significance was set at a level of $p < 0.05$.

Results

Mean Sensitivity

All visual field test results were within our inclusion criteria. Analyzing individual error responses however exhibited that all participants of the included VF data exhibited FPs and FLs <20% for both test strategies and FN<20% for SITA. Absolute values for FPs were 2.1 +/-3.0 (SITA) and 2.7 +/- 6.0 (SPARK Precision), FNs are only measured for SITA (0.8 +/- 0.5) but not for SPARK Precision. Fixation losses are logged for both test strategies and absolute values were as follows, SITA: 3.4 +/-3.8 and SPARK Precision: 1.0 +/-3.6. Between strategy analyses were carried out in all 83 subjects. SITA exhibited a significantly lower mean sensitivity (on average 0.93dB less than SPARK Precision) ,Wilcoxon Signed Rank test: $Z = -6.118$, $p < 0.001$. The examination with SPARK was, when adjusting for the fact that it uses ten test locations less than SITA, on average 33% faster to complete than SITA ($p < 0.01$) (Table 1 shows the raw data duration times for both methods).

The mean sensitivities of SITA and SPARK Precision were both shown to be negatively correlated to the subject's age (Spearman correlation coefficient: $\rho = -0.382$, $p < 0.001$ for SITA; $\rho = -0.752$, $p < 0.001$ for SPARK; see figure 2).

A positive correlation between group mean sensitivities of SITA and SPARK Precision was found (Spearman correlation coefficient: $\rho = 0.713$, $p < 0.001$). The bias of the mean sensitivity difference between SITA and SPARK Precision and the 95% limits of agreement (LoA) was 0.92 dB (LoA -1.12, 2.97 dB). As 95% of the data points were located within the LoA and further linear regression testing on the differences showed no statistically significant proportional bias ($t = 1.713$, $p = 0.09$).

Pointwise Analysis

Fourteen subjects showed poor reliability in the first visual field visit. These subjects were excluded from the within threshold test strategy comparisons which was carried out on 69 subjects (mean age 37.6 years, SD 16.0; range 20-71). Within test strategy comparisons

were made by calculating the mean difference in threshold for each visit for a given test algorithm (i.e. sensitivity at visit 2 – sensitivity at visit 1). Table 2 shows that there was much lower test-retest variability in SPARK Precision compared to SITA. The group mean sensitivity of SITA at the second visit was significantly higher compared to the first visit (mean of visit 1 = 30.58dB, mean of visit 2 = 30.96dB; paired t-test: $p = 0.003$), although these differences are not clinically significant. No statistically significant difference in group mean sensitivity was found between visits for the SPARK Precision strategy (median of visit 1 = 32.20dB, median of visit 2 = 32.18 dB; Wilcoxon Signed Rank test: $p = 0.522$).

The comparison between the two strategies (69 subjects) for the average of point-wise variations between visits of all 66 test points showed SITA had statistically significant larger variability (Paired t-test: $t = 9.175$, $df = 65$, $p < 0.001$) compared to using SPARK Precision in normal subjects.

Between test strategies comparisons were made by calculating the mean difference between the measured threshold of SPARK Precision and SITA on a point-wise basis (Figure 1). At 92% of the 66 stimulus locations, SPARK Precision recorded a higher sensitivity than SITA.

A statistical difference was found across the six regions (Table 3; one-way ANOVA: $F = 11.192$, $p < 0.001$). Whereas, the largest bias of mean sensitivities between SPARK Precision and SITA was found in the superior region. Post-hoc analysis using Tukey's HSD showed that the temporal region had the lowest bias compared to all other regions ($p < 0.05$) except for the inferior region ($p = 0.633$). The same analysis in terms of central, mid-peripheral and peripheral rings of eccentricity yielded no statistically significant difference among these three regions (one-way ANOVA: $F = 0.146$, $p = 0.865$) (Table 4).

Discussion

In the presented sample, the threshold estimates produced by SPARK Precision and SITA were found to have significant associations with age. The mean threshold estimates decreased with age in agreement with previous reports^{17, 21-25} by approximately -0.4dB/decade and -0.8dB/decade for SITA and SPARK Precision respectively which is close to the rates reported by others^{17,23,26,27}.

The mean sensitivity of SPARK Precision was statistically, but not clinically higher than that of SITA (0.92dB). Bland-Altman analysis showed a high level of agreement between the two strategies in this group of normal subjects. According to the evaluation criteria suggested by Luithardt et al.²⁸ a bias of less than 1dB indicates a good agreement. However, limits of agreement greater than 4dB can be classed as clinically acceptable. The most likely explanations for the higher threshold estimates produced by SPARK Precision are firstly reduced fatigue and secondly that threshold responses in SPARK Precision are interpolated from acquired thresholds at limited stimulus locations whereas each stimulus location in SITA is derived from independent assays at each stimulus location and therefore is influenced more by subject variability. The test duration of SITA, which uses 10 more test locations than SPARK, was on average six minutes compared with 3.5 minutes in SPARK Precision. While this time reduction is significant it is not an absolute time reduction as SPARK Precision tests only 66 points whereas SITA is assessing 76 points. Nevertheless any reduction in testing time which does not sacrifice accuracy is desirable to avoid fatigue, in particular when testing older subjects. A reduction in mean sensitivity at a rate of 0.08 to 0.1 dB has been attributed to the fatigue effect⁶. Thus the faster examination time of SPARK Precision could partly explain the higher group mean sensitivity. SITA utilises two likelihood functions developed from prior knowledge of normal and glaucomatous models which continue to be adjusted according to the patients' response during the test⁸. The final threshold estimate of SITA is derived according to the pre-determined precision in the mathematical model selected (Standard or Fast). Whereas SPARK Precision uses the first phase stimulus responses to determine the threshold estimates for the other three phases.

In these phases, the threshold estimates appear to be predominantly determined by interpolation and multiple regression equations which could be considered as a smoothing of the acquired data, reducing threshold variability. This smoothing effect can lead to reduced accuracy in detecting more sharply demarked areas of VF loss as well as a lack of precision in evaluating the depth of the VF loss. As SPARK Precision can be seen a logical progression from the TOP algorithm, it has distinct differences to its “predecessor”. While SPARK Precision has built on the basic ideas of TOP which are to utilize the correlation between neighbouring locations and is carrying out the test in four phases, there are distinct differences between the two test methods. TOP has been shown to lack accuracy in regards to estimating the spatial extent and absolute sensitivity loss of VF defects ²⁹. While both employ four phases to generate their threshold values, TOP measures fewer points than SPARK Precision which presents stimuli at each test location at least once. This may appear as a relatively small difference between methods but has a significant impact on the accuracy of the final thresholds obtained. This is because if a patient does not respond to a stimulus presented in TOP which could be detected, then all its neighbouring points will be influenced by this, in SPARK Precision, where each test location is at least measured once, the effect of such an error is less pronounced.

Point-wise comparison of group mean sensitivities between strategies showed most of the test points had higher threshold estimate when using SPARK Precision compared to SITA. A higher sensitivity using SITA was only present at five stimulus locations in the temporal field. This region also yielded the lowest mean bias when assessing the six regions described by de la Rosa et al³⁰ but we are unable to provide an explanation for this finding. In conclusion, the clinical examination of normal subjects using SPARK Precision yields a sensitivity profile across the visual field which is similar to SITA Standard but in a relatively faster examination time. Our test results are not generalizable in that the clinical validity of SPARK Precision in patients with a variety of systemic and ocular conditions characterized by visual field loss will need to be assessed in a separate trial including patients with established visual field defects.

References:

1. Wu Z, Medeiros FA. Recent developments in visual field testing for glaucoma. *Curr Opin Ophthalmol*. 2018;29(2):141-146.
2. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scandinavica*. 1997;75(4):368-375.
3. Bengtsson B, Heijl A and Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. *Acta Ophthalmol Scandinavica*. 1998;76(2):165-169.
4. Bengtsson B, Heijl A. Evaluation of a new perimetric strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scandinavica*. 1998;76:368-375.
5. Olsson J, Bengtsson B, Heijl A, Rootzen H. An improved method to estimate frequency of false positive answers in computerized perimetry. *Acta Ophthalmol Scandinavica*. 1997;75:181-183.
6. King-Smith PE, Grigsby SS, Vingrys AJ, Venes SC, Supowit A. Efficient and unbiased modifications of the Qest Threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res*. 1994;34:885-912.
7. Vingrys AJ, Pianta M. A new look at threshold estimation algorithms for automated static perimetry. *Optom Vis Sci*. 1999;76:588-595.
8. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Performance of efficient test procedures for frequency doubling technology in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2002;43:709-715.
9. Turpin A, Mc Kendrick AM, Johnson CA, Vingrys AJ. Development of efficient threshold strategies for frequency doubling technology perimetry using computer simulation. *Invest Ophthalmol Vis Sci*. 2002;43:322-331.
10. Gonzalez de la Rosa M and Gonzalez-Hernandez M. A strategy for averaged estimates of visual field threshold: Spark. *J Glaucoma*. 2013;22:284-289.
11. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107(10):1809-15.
12. Gonzalez de la Rosa M, Gonzalez Hernandez M, Aguilar Estevez J, Abreu Teyes A, Areja Tios A. Clasificación topográfica del campo visual glaucomatoso. *Arch Soc Esp Oftalmol*. 2002;77:87-94.
13. Gonzalez de la Rosa M and Gonzalez-Hernandez M. Monitoring visual field progression. *Brit J Ophthalmol*. 2011;95:157-158.
14. Heijl A, Lindgren G and Olsson J. The effect of perimetric experience in normal subjects. *Arch Ophthalmol*. 1989;107:81-86.
15. Gonzalez de la Rosa M, Gonzalez-Hernandez M, Sanchez-Garcia M, Rodriguez De la Vega R, Diaz-Aleman T, Rios AP. Oculus-Spark perimetry compared with 3

- procedures of glaucoma morphologic analysis (GDx, HRT, and OCT). *Eur J Ophthalmol.* 2013;23(3):316–323.
16. Gonzalez de la Rosa M, Gonzalez-Hernandez M, Alayon S. Glaucoma morphologic damage estimated from functional tests. *Eur J Ophthalmol.* 2015;25(6):496-502
 17. Brenton RS and Phelps CD. The normal visual field on the Humphrey Field Analyzer. *Ophthalmologica.* 1986;193: 56-74.
 18. Searle AET, Wild JM, Shaw DE and O'Neill EC. Time related variation in normal automated static perimetry. *Ophthalmology.* 1991;98: 701-707.
 19. Hodapp E, Parrish RK II, Anderson DR. *Clinical decisions in glaucoma.* St. Louis: The CV Mosby Co. 1993.
 20. Armstrong RA. When to use the Bonferroni correction. *Ophthalm Physiol Opt.* 2014;34: 502-508.
 21. Haas A, Flammer J and Schneider U. Influence of age on the visual fields of normal subjects. *Am J Ophthalmol.* 1986;101: 199-203.
 22. Jaffe GJ, Alvarado JA and Juster RP. Age-related changes of the normal visual field. *Arch Ophthalmol.* 1986;104: 1021-1025.
 23. Heijl A, Lindgren G and Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol.* 1987;105: 1544-1549.
 24. Iwase A, Kitazawa Y and Ohno Y. On age-related norms of the visual field. *Jpn J Ophthalmol.* 1987;32: 429-437.
 25. Spry PG and Johnson CA. Senescent changes of the normal visual field: an age-old problem *Optom Vis Sci.* 2001;78: 436-441.
 26. Weijland A, Fankhauser F, Bebie H and Flammer J. *Automated Perimetry. Visual Field Digest.* CH-Köniz: Haag-Streit AG. 1983.
 27. Wild JM, Pacey IE, Hancock SA and Cunliffe IA. Between-algorithm, between-individual differences in normal perimetric sensitivity: Full Threshold, FASTPAC, and SITA. *Invest Ophthalmol Vis Sci.* 1999;40: 1152-1161.
 28. Luithardt AF, Meisner C, Monhart M, Krapp E, Mast A and Schiefer U. Validation of a new static perimetric thresholding strategy (GATE). *Br J Ophthalmol.* 2015;99: 11-15.
 29. Anderson AJ. Spatial resolution of the tendency-oriented perimetry algorithm. *Invest Ophthalmol Vis Sci.* 2003; 44(5):1962-8.
 30. Gonzalez De La Rosa M, Gonzalez Hernandez M, Abrales M and Azuara-Blanco A. Quantification of inter-point topographic correlations of threshold values in glaucomatous visual fields. *J Glaucoma.* 2002;11: 30-34 , 11, p.30-34.