

Comparison between two fast threshold strategies: SPARK and SITA in normal subjects

Say Kiang Foo¹, Robert Peter Cubbidge² and Rebekka Heitmar² 

European Journal of Ophthalmology
2021, Vol. 31(4) 1870–1876

© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1120672120926455
journals.sagepub.com/home/ejo



Abstract

Background: Numerous fast threshold strategies have been developed in perimetry which use maximum likelihood approaches to estimate the threshold. A recent approach to threshold estimation has been developed estimating the threshold from a limited number of test points which further reduces examination time. This strategy, SPARK, has not been compared to the SITA strategy. The aim of this study was to compare SPARK with SITA in a normal cohort to evaluate within and between strategy agreement in threshold estimates.

Methods: A total of 83 normal subjects each underwent two visual field examinations with SITA and SPARK on two separate occasions on a randomly selected eye. The eye examined and the order of strategy examined first was randomised but remained constant over the two perimetry visits.

Results: Visual field examination with SPARK Precision was on average 33% faster than SITA Standard. A positive correlation between group mean sensitivities of SITA Standard and SPARK Precision ($\rho = 0.713$, $p < 0.001$) was found. In total, 95% of stimulus locations were located within the 95% limits of agreement and linear regression on the differences in sensitivities showed no statistically significant proportional bias ($t = 1.713$, $p = 0.09$). Pointwise analysis showed SITA Standard had significantly larger variability for individual stimulus locations examined over two visits when compared to SPARK ($t = 9.175$, $p < 0.001$).

Conclusion: The clinical examination of SPARK yields a sensitivity profile similar to SITA but in a faster examination time. The lower threshold variability of SPARK may be as a result of data smoothing in the threshold estimation process.

Keywords

Clinical tests, glaucoma, lens, cataract, optics, refraction, instruments, psychophysical testing, retina, techniques of retinal examination

Date received: 22 October 2019; accepted: 24 April 2020

Background

Standard automated perimetry (SAP) remains an essential clinical assessment of visual function in patients suffering from glaucoma.¹ 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 fatigue^{2–9} 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

¹SEGi University, Faculty of Optometry & Vision Sciences, Petaling Jaya, Malaysia

²School of Life and Health Sciences, Aston University, Birmingham, UK

Corresponding author:

Rebekka Heitmar, School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK.

Email: R.Heitmar1@aston.ac.uk

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 de la Rosa and Gonzalez-Hernandez.¹⁰ 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 point.¹⁰ 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 3 min have been achieved.¹⁴ The initial phase of the SPARK Precision algorithm (also referred to as SPARK training) can be completed in less than 40 s¹⁵ and SPARK Precision has exhibited good diagnostic sensitivity and specificity in glaucoma patients even when utilising the first phase results only.¹⁵ Moreover, the agreement of SPARK Precision with morphologic indices¹⁵ enabled its threshold sensitivity to be used to predict the thickness of retinal nerve fibre layer.¹⁶ The objective of this study was to evaluate the mean sensitivity (MS) and pointwise 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 males; mean age: 39.6 years; standard deviation (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) ≤ 21 mmHg; a previously normal visual field; normal optic disc appearance, that is, no localised neuroretinal rim loss; an absence of optic disc haemorrhage, notches or pallor; a cup/disc ratio ≤ 0.6 or cup/disc asymmetry ≤ 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 ± 6.00 -DS spherical error and less than 2.50-DC 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 2 h 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 (OCULUS Optikgeraete GmbH, Wetzlar, Germany) using the SPARK Precision algorithm and with the Humphrey field analyzer (HFA) (Carl Zeiss Meditec, Dublin, CA, United States) using the 30-2 SITA Standard algorithm. Both instruments use a background luminance of 10 cd/m² 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° field with a square grid of 6° separation.¹⁷ The SPARK Precision algorithm deploys a similar grid with a total of 66 test points (30° × 24°) with the uppermost and bottommost rows and two points located in the blind spots missing compared to the SITA 30-2 test grid.¹⁰ 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 region.¹⁵ Thresholds for all other locations are determined by interpolation, but ultimately each of the 66 points is assessed at least once. In the following three 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 min 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

conducted to determine proportional bias. All correlations were determined using Spearman’s correlation coefficient. A one-way analysis of variance (ANOVA) was used for comparison among the test zones followed with Tukey’s honestly significant difference (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 Armstrong.¹⁹ Statistical significance was set at a level of $p < 0.05$.

Results

MS

All visual field test results were within our inclusion criteria. Analysing 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. FLs 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 MS (on average 0.93 dB 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 10 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 MS 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, further linear regression testing on the differences showed no statistically significant proportional bias ($t = 1.713$, $p = 0.09$).

Table 1. The mean sensitivity and test duration of SITA and SPARK in normal subjects.

	Mean	SD	Median	Range	
				Min	Max
Mean sensitivity (dB)					
SITA	30.81	1.32	30.94	27.32	33.61
SPARK	31.74	1.50	31.94	28.06	34.32
Test duration (min)					
SITA	6.05	0.67	5.87	4.92	8.18
SPARK	3.50	0.13	3.50	3.23	3.82

SD: standard deviation.

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

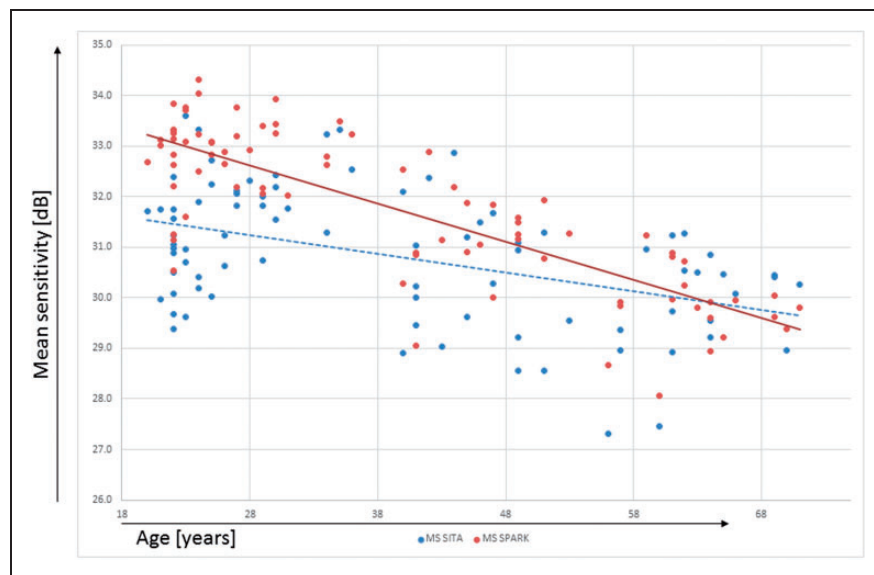


Figure 2. Correlation between age and mean sensitivity or SPARK Precision and SITA.

Table 2. Mean and standard deviation of pointwise mean differences between-visit using SITA and SPARK.

	Type of strategy	Mean	SE
Pointwise mean difference (dB)	SITA	0.37	0.03
	SPARK	0.04	0.02

SE: standard error.

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 MS of SITA at the second visit was significantly higher compared to the first visit (mean of visit 1 = 30.58 dB, mean of visit 2 = 30.96 dB; paired t-test: $p = 0.003$), although these differences are not clinically significant. No statistically significant difference in group MS was found between visits for the SPARK Precision strategy (median of visit 1 = 32.20 dB, 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 pointwise variations between visits of all 66 test points showed that 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 strategy comparisons were made by calculating the mean difference between the measured threshold of SPARK Precision and SITA on a pointwise 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

Table 3. Mean and standard deviation of bias in MS between strategies according to six functional regions described by de la Rosa et al.¹⁷

Region	Bias in MS (dB)	
	Mean	SD
Superior	1.37	0.69
Superior nasal	1.06	0.37
Inferior nasal	1.10	0.46
Inferior	0.43	0.37
Temporal	0.11	0.46
Central	0.92	0.41

MS: mean sensitivity; SD: standard deviation.

Table 4. Mean and standard deviation of bias in MS between strategies according to eccentricity.

Location	Bias of MS (dB)	
	Mean	SD
Central ($<10^\circ$)	0.82	0.32
Mid-peripheral (10° – 20°)	0.85	0.49
Peripheral ($>20^\circ$)	0.91	0.76

MS: mean sensitivity; SD: standard deviation.

threshold estimates decreased with age in agreement with previous reports^{17,20–24} by approximately -0.4 and -0.8 dB/decade for SITA and SPARK Precision, respectively, which is close to the rates reported by others.^{17,22,25,26}

The MS of SPARK Precision was statistically but not clinically higher than that of SITA (0.92 dB). 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.,²⁷ a bias of less than 1 dB indicates a good agreement. However, LoA greater than 4 dB can be classed as clinically acceptable. The most likely explanations for the higher threshold estimates produced by SPARK Precision are first reduced fatigue and second 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 6 min compared with 3.5 min 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 MS at a rate of 0.08–0.1 dB has been attributed to the fatigue effect.⁶ Thus, the faster examination time of SPARK Precision could partly explain the higher group MS. 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 test.⁸ 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 utilise 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.²⁸ 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.

Pointwise comparison of group mean sensitivities between strategies showed that 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 al.,²⁹ 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 characterised by visual field loss will need to be assessed in a separate trial including patients with established visual field defects.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Say Kiang Foo received an institutional financial grant to complete this work (SEGiRP/2014-41/FOVS-2/51).

ORCID iD

Rebekka Heitmar  <https://orcid.org/0000-0002-7657-1788>

References

1. Wu Z and 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, et al. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand* 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 Scand* 1998; 76(2): 165–169.
4. Bengtsson B and Heijl A. Evaluation of a new perimetric strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand* 1998; 76: 368–375.
5. Olsson J, Bengtsson B, Heijl A, et al. An improved method to estimate frequency of false positive answers in computerized perimetry. *Acta Ophthalmol Scand* 1997; 75(2): 181–183.
6. 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. *Vision Res* 1994; 34: 885–912.
7. Vingrys AJ and Pianta M. A new look at threshold estimation algorithms for automated static perimetry. *Optom Vis Sci* 1999; 76(8): 588–595.
8. Turpin A, McKendrick AM, Johnson CA, et al. 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, 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(2): 322–331.

10. de la Rosa MG and Gonzalez-Hernandez M. A strategy for averaged estimates of visual field threshold: spark. *J Glaucoma* 2013; 22(4): 284–289.
11. Garway-Heath DF, Poinoosawmy D, Fitzke FW, et al. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000; 107(10): 1809–1815.
12. de la Rosa MG, Gonzalez Hernandez M, Aguilar Estevez J, et al. Clasificacion topografica del campo visual glaucomatoso. *Arch Soc Esp Oftalmol* 2002; 77: 87–94.
13. de la Rosa MG 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(1): 81–86.
15. de la Rosa MG, Gonzalez-Hernandez M, Sanchez-Garcia M, et al. Oculus-spark perimetry compared with 3 procedures of glaucoma morphologic analysis (GDx, HRT, and OCT). *Eur J Ophthalmol* 2013; 23(3): 316–323.
16. de la Rosa MG, Gonzalez-Hernandez M and 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. Hodapp E, Parrish RKII and Anderson DR. *Clinical decisions in glaucoma*. St. Louis, MO: C.V. Mosby Company, 1993.
19. Armstrong RA. When to use the Bonferroni correction. *Ophthalm Physiol Opt* 2014; 34: 502–508.
20. Haas A, Flammer J and Schneider U. Influence of age on the visual fields of normal subjects. *Am J Ophthalmol* 1986; 101(2): 199–203.
21. Jaffe GJ, Alvarado JA and Juster RP. Age-related changes of the normal visual field. *Arch Ophthalmol* 1986; 104: 1021–1025.
22. Heijl A, Lindgren G and Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1987; 105(11): 1544–1549.
23. Iwase A, Kitazawa Y and Ohno Y. On age-related norms of the visual field. *Jpn J Ophthalmol* 1987; 32: 429–437.
24. Spry PG and Johnson CA. Senescent changes of the normal visual field: an age-old problem. *Optom Vis Sci* 2001; 78: 436–441.
25. Weijland A, Fankhauser F, Bebie H, et al. *Automated perimetry visual field digest*. Köniz: Haag-Streit AG, 1983.
26. Wild JM, Pacey IE, Hancock SA, et al. Between-algorithm, between-individual differences in normal perimetric sensitivity: full threshold, FASTPAC, and SITA. *Invest Ophthalmol Vis Sci* 1999; 40(6): 1152–1161.
27. Luithardt AF, Meisner C, Monhart M, et al. Validation of a new static perimetric thresholding strategy (GATE). *Br J Ophthalmol* 2015; 99(1): 11–15.
28. Anderson AJ. Spatial resolution of the tendency-oriented perimetry algorithm. *Invest Ophthalmol Vis Sci* 2003; 44(5): 1962–1968.
29. de la Rosa MG, Gonzalez-Hernandez M, Abraldes M, et al. Quantification of inter-point topographic correlations of threshold values in glaucomatous visual fields. *J Glaucoma* 2002; 11: 30–34.