OPEN ACCESS

Clinical science

Prediction of repeatable glaucomatous visual field defects based on cluster characteristics

Jeremy C K Tan ,[1,2] Jack Phu,[1,3] Katharina Bell,[4,5] Ashish Agar,[6] Jonathan Crowston,[2] Giovanni Montesano [7]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bjo-2025-327416).

[1]Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia
[2]The University of Sydney Save Sight Institute, Sydney, New South Wales, Australia
[3]Optometry and Vision Science, University of New South Wales, Sydney, New South Wales, Australia
[4]NHMRC Clinical Trials Centre, Camperdown, New South Wales, Australia
[5]Singapore National Eye Centre, Singapore
[6]Department of Ophthalmology, Prince of Wales Hospital and Community Health Services, Sydney, New South Wales, Australia
[7]Optometry and Vision Sciences, City University of London, London, UK

Correspondence to
Dr Jeremy C K Tan;
jeremy.c.tan@unsw.edu.au

Received 19 February 2025
Accepted 24 June 2025

To cite: Tan JCK, Phu J, Bell K, *et al*. *Br J Ophthalmol* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/bjo-2025-327416

ABSTRACT

Aim This study evaluates if characteristics (eg, location, size, volume) of clusters of defects on an initial visual field (VF) test were predictive of a repeatable defect in the subsequent two tests.

Methods Retrospective cohort study of 197 eyes of 103 patients with healthy, suspect or early glaucoma. Using the initial VF pattern deviation probability grid, we defined the number of clusters (≥1 location of p<5%) and associated size (number of adjoining defect locations) and volume (sum of corresponding total deviation values) for each cluster stratified by the four probability levels (ie, p<5%; p<2%; p<1% and p<0.5%).

Results Of 4424 locations with a defect of p<5%, only 1189 (26.9%) were repeatable. The size [area under the receiver operating characteristic curve (AUC) 0.80, CI 0.76 to 0.85)] and volume (AUC 0.80, CI 0.76 to 0.85) of clusters were predictive of a repeatable defect within the cluster. The optimal thresholds for predicting a repeatable location within each cluster at 95% specificity based on initial cluster size were >6 locations at p<5%, >4 locations at p<2%, >3 locations at p<1% and >2 locations at p<0.5%. Defining cluster defects by involvement of central or peripheral rim locations improved the predictive value compared with the entire 24–2 grid.

Conclusion The location, size and volume of clusters of defects on an initial VF test may be predictive of subsequent repeatability. This may help distinguish eyes with a higher risk of repeatable defects.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Visual field (VF) testing in suspected and early glaucoma is often characterised by non-repeatable defects, making it difficult to distinguish variability from true functional loss.

WHAT THIS STUDY ADDS

⇒ The location, size and volume of clusters of defects on an initial VF test may be predictive of subsequent repeatability. This may help distinguish the eyes of subjects with a higher risk of repeatable defects.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Distinguishing between patients with a higher risk of repeatable defects may be helpful in risk stratification and resource allocation, especially given the lack of capacity in hospital eye services for patients with glaucoma.

INTRODUCTION

The assessment of visual field (VF) function using standard automated perimetry (SAP) is an essential component in the diagnosis and monitoring of patients with suspected or manifest glaucoma.[1 2] SAP is, however, a subjective assessment and is influenced by the reliability of the patient's performance.[3] Despite results being potentially reliable, the inherent fluctuations of VF testing can make it difficult to distinguish between true functional loss and variability.[4] A common definition used to define a glaucomatous VF defect is the identification of clusters of contiguous test points exhibiting statistically significant sensitivity loss—typically defined as one or more adjacent VF locations with a probability level of 5% or less on pattern deviation plots. Cluster defects have demonstrated varying degrees of specificity of diagnosing glaucoma.[5–8]

Another outcome of interest is whether a cluster defect on an initial VF test will be repeatable in subsequent tests. Can certain characteristics of the cluster, for instance, the eccentricity of locations on the grid involved, size and depth of the defect be used to predict a subsequent repeatable defect? If so, these features may help differentiate between variability and true sensitivity loss. This is especially relevant in glaucoma suspects or early glaucoma, since VFs in these patients are more likely to display cluster defects[5] or non-repeatable changes.[9] In this study, we evaluated if characteristics of cluster defects on an initial VF test could be used to predict subsequent repeatable defects in a cohort of patients with healthy, suspect or manifest early glaucoma.

METHODS

This was a retrospective study conducted at the Prince of Wales Hospital Department of Ophthalmology, a tertiary referral eye unit in Sydney, Australia and the Centre for Eye Health, University of New South Wales, Sydney, Australia with institutional ethics approval obtained. We have described our methodology in previous publications.[9 10] In brief, VF testing was performed using the 24–2 Swedish Interactive Thresholding Algorithm (SITA)-Fast or Faster strategy on the Humphrey Field Analyzer 3. Only patients with a

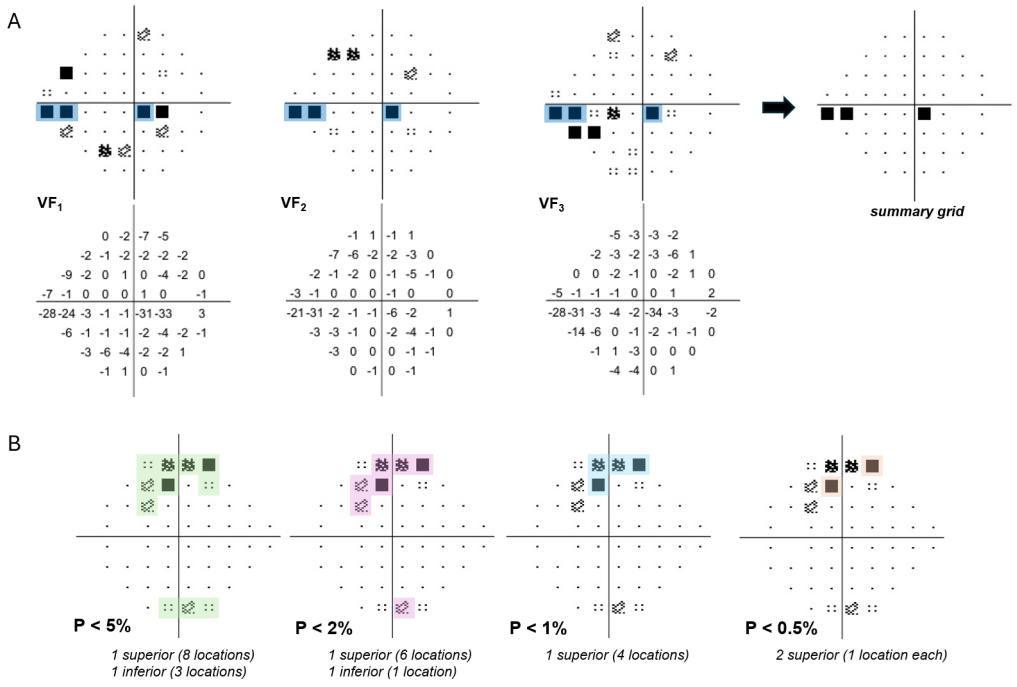

Figure 1 (A) A series of three consecutive VF tests $VF_{1\text{-}3}$, with a peripheral cluster defect and a centre-involving cluster defect. The blue boxes highlight locations in each test which were repeatable over the three consecutive tests. This was used to form a summary grid of repeatable defects for each eye. The corresponding total deviation values of each cluster were used to calculate the volume. Characteristics of clusters (location, size, volume) on the initial test (VF_1) were thereafter used to predict firstly the repeatability of the entire defect, and secondly the presence of at least one subsequently abnormal location within the defect. (B) Using the initial VF test for each eye (VF_1), we defined the number of clusters and associated size and volume (using corresponding total deviation values) of each cluster using the four probability defect levels as minimum thresholds (p<5%; p<2%; p<1% and p<0.5%). Note that at higher probability thresholds (p<5%, p<2% and p<1%), the defect in the superior hemifield comprises only one cluster of multiple locations. When the minimum probability threshold of p<0.5% is applied however, the superior defect now becomes two separate clusters of one location each. VF, visual field.

mean deviation (MD) of better than −6 dB who received three VF tests performed on three separate visits (VF_1, VF_2 and VF_3) within a 12-month period were included, since non-repeatable defect clusters were more likely in this cohort. Subjects included in this study were already under routine monitoring; they therefore already had some perimetric experience prior to the first VF test.

VF analysis

The following parameters were extracted using a script written in Python: VF global indices (MD, Visual Field Index and glaucoma hemifield test), false-positive rate (FP), pointwise pattern deviation probability (PDP) and total deviation values directly from the VF printout. We then defined clusters of defects in the initial VF of each eye based on the PDP grid data. A defect 'cluster' was defined as a defect comprised of at least one location on the PDP grid with a probability score of at least p<5% (ie. this definition also encompasses all stricter probability levels). The PDP grid was used to define defects since it corrects the total deviation probability grid for overall sensitivity loss, highlighting patterns of defect seen in conditions such as glaucoma. VFs with FP above 15% were not excluded from the initial analysis.[11–14] We first generated a *summary grid* of repeatable defects for each eye by defining locations of the PDP grid which had a probability score of at least p<5% on each of the three consecutive tests ($VF_{1\text{–}3}$ in figure 1A, locations defined by blue box). The highest probability score over the three tests at each location was assigned as the final value for that location in the summary grid (ie. a location with PDP scores of 'P < 5%', 'P < 2%' and 'P < 1%' from tests 1 to 3 will have a final PDP score of 'P < 5%' on the summary grid).

Defining characteristics of clusters at different minimum probability levels

The size of a cluster as defined using the PDP grid is related to the PDP score of each location; intuitively clusters with higher (ie. less strict) probability values (p<5%) tend to be larger and may be less repeatable than clusters with lower (ie. stricter) probability values (p<0.05%). We therefore defined the number of clusters and associated size of each cluster using the four probability levels as minimum thresholds (p<5%; p<2%; p<1% and p<0.5%) on the initial VF test (figure 1B). For each cluster, we used the corresponding total deviation values for each location to calculate the 'volume' of each cluster, which represented the 'depth' of the cluster defect. For instance, if a particular cluster of three adjoining adjacent locations had total deviation values of '−5', '−8' and '−6', the volume of the cluster would be '−19'. Hence, we defined the size and volume of each cluster using the four probability levels in all eyes.

After documenting characteristics of all clusters in all eyes, we then examined what proportion of locations in each cluster on the initial test was repeatable over the subsequent two tests—that is, what percentage of locations was defective below the 5%

Table 1 Baseline global sensitivity and reliability characteristics of study cohort (N=197 eyes)

Parameter	Visit 1	Visit 2	Visit 3	P value
MD (SD), dB	−1.03 (1.82)	−1.08 (3.02)	−0.99 (2.37)	0.61
PSD (SD), dB	2.39 (1.21)	2.35 (1.35)	2.35 (1.34)	0.38
VFI (SD), %	97.3 (2.6)	96.8 (4.6)	97.1 (3.4)	0.63
GHT				0.95
Within normal limits	65 (33.0%)	72 (36.5%)	72 (36.5%)	
Borderline	39 (19.8%)	40 (20.3%)	37 (18.8%)	
Outside normal limits	85 (43.1%)	76 (38.6%)	81 (41.1%)	
Abnormally high sensitivity	8 (4.1%)	9 (4.6%)	7 (3.6%)	
FP (SD), %	9.3 (12.7)	8.3 (12.8)	8.7 (13.8)	0.65
Time since first visit (SD), years	0	0.3 (0.1)	0.8 (0.2)	

FP, false positive rate; GHT, glaucoma hemifield test; MD, mean deviation; PSD, pattern standard deviation; VFI, visual field index.

probability level over the next two tests. This was to explore the relationship between cluster characteristics (size, volume) and repeatability at each probability depth level. This was evaluated using all clusters for each eye as well as using only the largest cluster for each eye.

Prediction of repeatable location within cluster

Next, we evaluated the ability of cluster characteristics (size and volume) to predict the presence of at least one repeatable location within the cluster across the three tests, at each of the four minimum defect depth levels, using Area under the receiver operating characteristic curves (ROC). We stratified this analysis by location by evaluating the clusters across the 24–2 grid irrespective of location, clusters in contact with the 20 locations forming the peripheral rim (peripheral rim clusters) and clusters in contact with any of the four central locations (centre-involving clusters). Finally, we performed subanalyses examining cluster defects by superior and inferior hemifields only and by excluding tests with an FP above 15%. All analyses were performed using Python and R.

RESULTS

Baseline characteristics of study cohort

197 eyes of 103 subjects were included, with a mean baseline MD of −1.03 (SD 1.82, IQR −2.17–0.11). The mean age was 62.9 (SD 10.4) years, and the cohort comprised 102 eyes with manifest glaucoma (51.8%), 54 (27.4%) suspect and 41 (20.8%) healthy eyes. There was no significant difference in mean MD in each eye between the three tests (MD −1.03 vs −1.08 vs −0.99 dB, repeated measures Analysis of Variance, p=0.61) or for the other sensitivity and reliability measures (table 1), which likely demonstrated stability of VF function over the study period. The mean follow-up duration of the second and third visit was 0.3 (0.1) and 0.8 (0.2) years from baseline, respectively.

Characteristics of initial and repeatable defects

Defects defined as p<5% on the PDP grid on the initial VF test in our cohort were observed to occur in a superior and inferior arcuate distribution, with a frequency of occurrence between 4.6% and 22.8% across locations (figure 2A). The frequency of *repeatable defects* was generally lower across the grid and ranged from 1% to 13.7%, with a similar greater involvement of the superior and inferior arcuate distribution (figure 2B). The mean frequency for the four central locations was 3.4% versus 4.5% for non-central locations (p=0.33). Of 4424 locations with a defect of at least p<5%, only 1189 (26.9%) were repeatable over three tests. 128 (65.0%) eyes had a cluster defect composed of at least three contiguous locations on the initial VF test, of which only 41 eyes (20.8%) displayed this as a repeatable defect in the subsequent two tests.

Repeatability of defects by cluster size and sum

196 (99.5%) of the 197 eyes had at least one cluster defect comprised of at least one location at the p<5% probability level; one eye with a healthy diagnosis had no defect. 177 (89.8%), 159 (80.7%) and 106 (53.8%) eyes had at least one cluster at the p<2%, p<1% and p<0.5% probability levels, respectively (table 2). A total of 1446 clusters were observed on the initial VF test from the 196 eyes. The mean size of each cluster was 3.1 (SD 4.3) locations, while the mean volume (sum of total deviation values) for each cluster was −17.7 (SD 27.2). Using the largest cluster for each eye only, the mean size and volume of each cluster were 4.4 (SD 5.4) locations and −25.5 (SD 33.6).

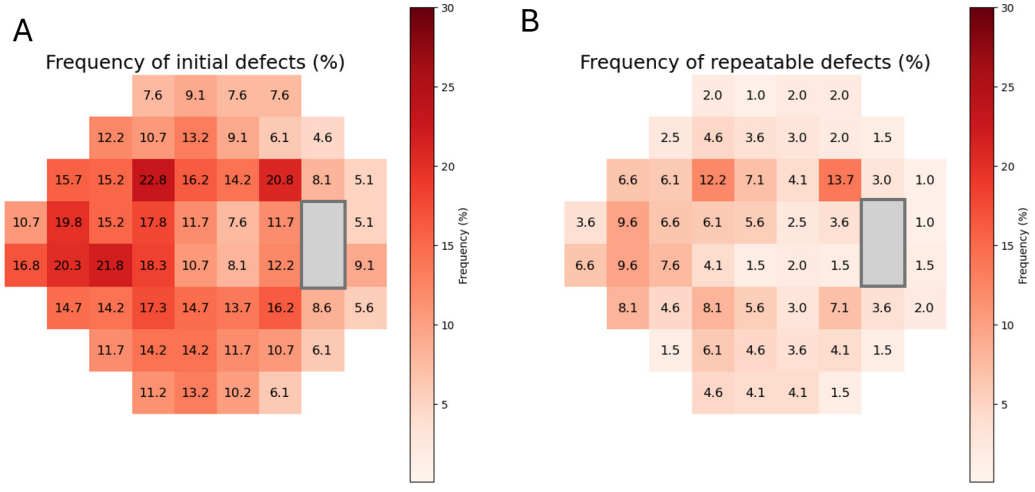

Figure 2 Frequency of defects (%) by location on the pattern deviation probability (PDP) grid of the initial visual field (VF) test (A) vs in defects that were repeatable over the three consecutive VF tests (B) in 197 eyes. There was a greater frequency of initial and repeatable defects in the superior and inferior arcuate distribution.

Table 2 Number and proportion of eyes with 1, 2, 3 or ≥4 clusters at each probability defect depth level (N=196 eyes)

Probability depth level	Number of clusters	Number of eyes	Proportion of eyes
p<5%	1	33	17%
	2	59	30%
	3	47	24%
	≥4	57	29%
p<2%	1	49	28%
	2	61	34%
	3	38	21%
	≥4	29	16%
p<1%	1	63	40%
	2	64	40%
	3	16	10%
	≥4	16	10%
p<0.5%	1	65	61%
	2	27	25%
	3	10	9%
	≥4	4	4%

Online supplemental figurea 1 and 2 are heatmaps showing what proportion of the extent of each cluster on VF_1 was repeatable, based on the size and volume of each distinct cluster on VF_1.

We then assessed the repeatability of at least one location of each cluster using initial cluster characteristics. Figure 3 displays ROC curves of cluster size (panel A) and volume (panel B) to predict the presence of at least one repeatable location within the cluster in the subsequent two tests, stratified by location of involvement (entire grid, centre-involving or peripheral rim clusters) and at each of the four probability levels. Using cluster size for all locations, the highest area under the ROC curve (AUC) was achieved by a threshold of >2 locations at the probability level of p<5% (AUC 0.80, CI 0.76 to 0.85). We found that defining clusters by involvement of central or peripheral rim locations improved the predictive value. For instance, size of clusters with central involvement and at the p<1% level achieved an AUC of 0.93 (CI 0.86 to 1.0), with a threshold of >3 locations. Central clusters displayed higher sensitivity but lower specificity than clusters involving the peripheral rim at the p<5% and p<2% levels. At the p<1% and p<0.5% levels, central clusters displayed superior sensitivity and specificity than peripheral rim-involving clusters. In addition, the thresholds for central were consistently larger than peripheral rim clusters. The

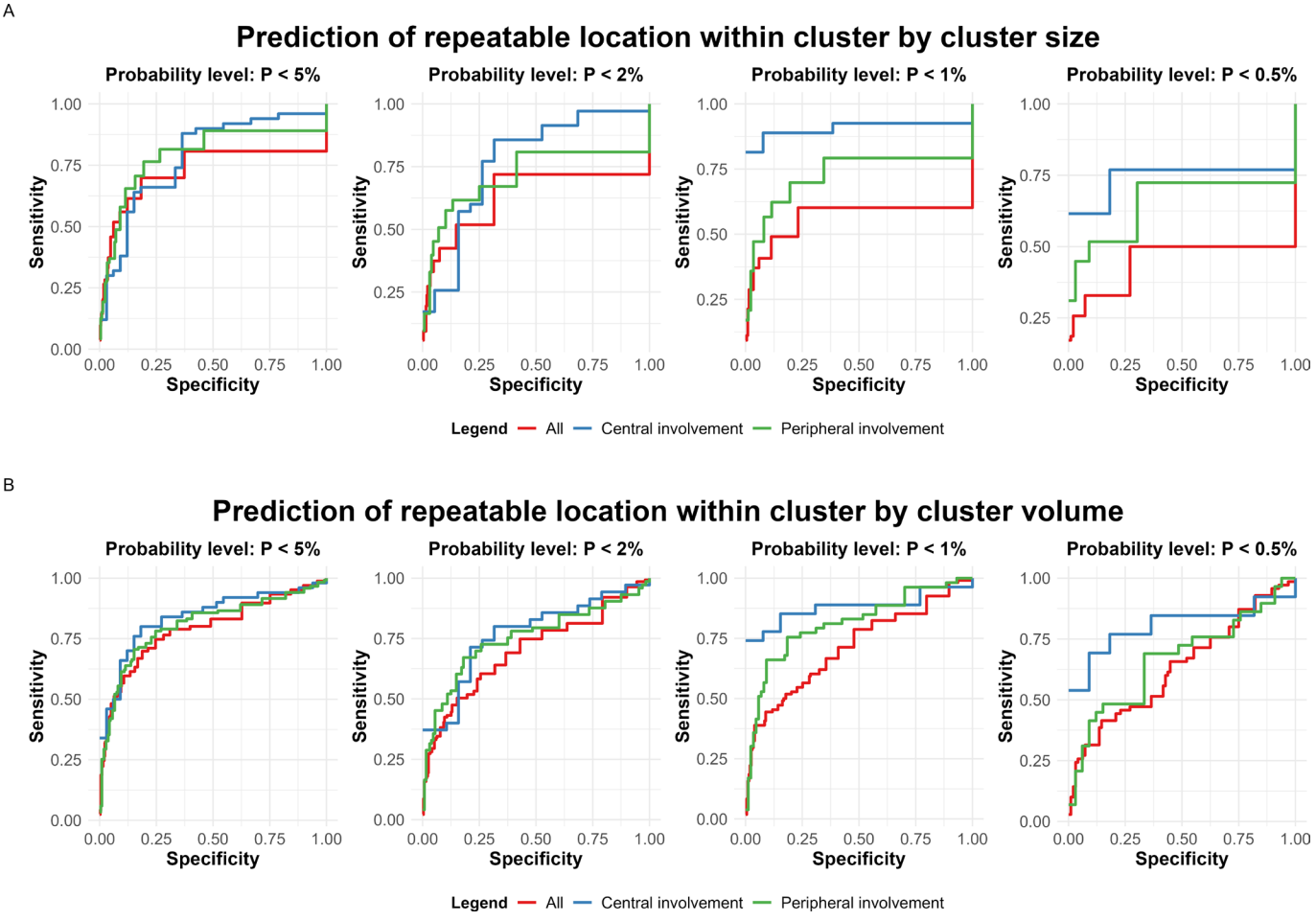

Figure 3 Receiver operating characteristic curves of cluster size (A) and cluster volume (B) to predict the presence of at least one repeatable location within the cluster across the three tests at each of the four probability defect depth levels, stratified by location (all: entire grid; central: involvement of central four locations; peripheral: involvement of any of the 20 locations that make up the peripheral rim). Centre-involving clusters generally displayed higher AUC values than peripheral or all clusters. AUC, area under the receiver operating characteristic curve.

optimal thresholds for predicting a repeatable location within each cluster at 95% specificity based on initial cluster size were >6 locations at p<5%, >4 locations at p<2%, >3 locations at p<1% and >2 locations at p<0.5%. A similar trend was also observed for cluster volume. The highest AUC when considering all locations was achieved using a threshold of <−10 at the probability level of p<5% (AUC 0.80, CI 0.76 to 0.85). The thresholds and AUC values for central-involving clusters were also consistently larger than peripheral clusters. The optimal thresholds for predicting a repeatable location within each cluster at 95% specificity based on initial cluster volume were <−25 dB at p<5%, <−33 dB at p<2%, <−29 dB at p<1% and <−38 dB at p<0.5%. Online supplemental table 1 summarises the best thresholds, sensitivity, specificity and AUC for prediction of a repeatable location within each cluster based on size and volume, stratified by location (also see online supplemental tables 2-7).

Repeatable defect by hemifield and test reliability

We divided cluster defects by hemifield and examined predictive values of the size and sum characteristics. The AUC was observed to be similar when examining cluster defects in the superior or inferior hemifield compared with using the entire grid. For instance, the highest AUC for cluster size in the superior hemifield was achieved by the probability depth level of p<5% (AUC 0.78, CI 0.72 to 0.84). This was also the case for the inferior hemifield, where AUC values were comparable to using the entire grid; the highest AUC for cluster size in the inferior hemifield was achieved by the probability level of p<5% (AUC 0.80, CI 0.74 to 0.85). Finally, we performed a subanalysis to assess if excluding tests with FP >15% could improve the predictive value. For cluster size and volume, the highest AUC was similarly achieved by the probability level of p<5%, with similar AUC values to including all tests. This finding demonstrates that excluding unreliable tests using the FP >15% criterion did not improve the predictive value of classifying eyes with a repeatable defect.

DISCUSSION

This study evaluated if characteristics of cluster defects on the PDP grid on an initial VF test can be used to predict repeatable defects in subsequent tests. We found that cluster size and volume were predictive of repeatable defects, especially when defining cluster defects by location of involvement.

Variability of VF testing

In our cohort of patients with healthy, suspect or mild glaucoma, only 26.9% of locations with probability of 5% or below on the initial VF test were repeatable over the subsequent two tests. Defects that were repeatable tended to fall within the superior arcuate or inferior arcuate distribution. Of note, central locations were not more repeatable than non-central locations. We have previously found variability to be lesser for central compared with peripheral locations on the SITA-Faster strategy,[15] but this could be related to inherently higher sensitivity centrally.

Use of cluster characteristics to predict future repeatable defects

We sought to leverage cluster characteristics by assessing their predictive value in classifying clusters with at least one repeatable location. We found that both properties yielded moderate AUC values, with the highest AUC achieved with a probability depth minimum level of p<5%. While defects with stricter probability scores (p<1% and p<0.5%) generally displayed greater repeatability than similar sized defects at p<5%, the superior predictive performance of the latter using cluster size may be due to the greater range of sizes at this probability level. Conversely, defects at a minimum probability level of p<1% or p<0.5% tended to be much smaller, with cluster size therefore less useful in distinguishing those more likely or not to be repeatable.

We observed that classifying clusters by central- or peripheral rim locations of involvement improved the predictive value. At stricter probability levels (p<1% and p<0.5%), centre-involving clusters tended to display significantly higher AUC values than peripheral rim or all clusters. This could be explained by test–retest variability in SITA-Faster, which we have found to be greater for peripheral compared with central test locations.[15] The thresholds for size and volume for centre-involving clusters were also consistently larger than peripheral rim clusters. This either indicated that centre-involving clusters needed to be larger to achieve similar or better sensitivity and specificity compared with peripheral rim clusters or could simply reflect an inherent tendency for centre-involving clusters to display larger size and volumes than peripheral rim clusters. We did not notice an improvement in predictive value by dividing clusters into hemifields. While most initial glaucomatous VF defects tend to develop in the superior hemifield,[16 17] we found that analysing clusters by hemifield did not improve the predictive value of cluster characteristics for repeatable defects.

Clinical implications

Examining cluster characteristics on the PDP grid may be an objective method to help identify eyes which are more likely to display a subsequent repeatable defect. This can assist in the interpretation of VF outputs, which is often subjective and with interobserver agreement among even expert observers moderate at best.[18–20] The implementation of such a system will, however, likely require automation through for instance a computer application,[21 22] given the tedious nature of calculating these characteristics manually which may be prone to human error or bias.

Practically, distinguishing between patients with a higher risk of repeatable defects may be helpful in risk stratification and resource allocation, especially given the lack of capacity in hospital eye services for patients with glaucoma.[23] For instance, a cluster size of >6 locations or cluster volume of <−25 dB using a minimum probability level of p<5% may be used by clinicians as a way to distinguish clusters with repeatable defects at a specificity of 95%. Prospective validation of these thresholds in a larger and more diverse dataset is however required.

Importantly, knowledge of a likely non-repeatable cluster defect is useful as it may guide the clinician to extend the subsequent VF test interval. Conversely, a baseline VF that displays a likely repeatable cluster defect may benefit from repetition and closer follow-up, as a treatment modification may be warranted. This system may also complement other baseline variables used in risk stratification tools, such as the risk calculators formulated from the Ocular Hypertension Treatment Study and European Glaucoma Prevention Study used to predict the development of primary open-angle glaucoma.[24]

Limitations

First, the use of both eyes of each patient may be a confounder, given the non-independence of VF test performance of each eye. Furthermore, the test order (right eye first by convention) may subject the second eye to the effects of fatigue.[25] We assessed the ROC metrics of the minimum cluster composition scores

in a subanalysis where only one eye per patient was randomly included, which produced consistent results of composition scores which yielded the highest AUC. Our study cohort was also observed over a time period of 1 year, with the latter chosen to mitigate any chance of progression potentially confounding the analysis. Second, the SITA-Fast and Faster strategy was used in our cohort, which may display different sensitivity outcomes compared with SITA-Standard.[26] Modifications made to SITA-Fast to produce SITA-Faster have been comprehensively described by Heijl and colleagues.[27] SITA-Faster may display sensitivity differences with SITA-Standard, especially in subjects with greater field loss.[27 28] The prospective validation of our findings in a larger cohort of subjects from other centres including with healthy eyes and also using other test strategies or perimeters may therefore be useful. Third, we acknowledge that the use of summed total deviation values as a surrogate for VF loss volume may not account for the clinical impact of central versus peripheral loss or generalised depression from media opacities. Alternative approaches—such as spatial weighting or pattern deviation-based metrics—may provide more refined assessments and will be considered in future work. Finally, while our sample population (MD >-6 dB) was deliberately chosen, given the higher likelihood of non-repeatable VF defects, the predictive value of cluster characteristics should also be assessed in subjects with more advanced disease.

CONCLUSION

The location, size and volume of clusters of defects in the initial VF test may be predictive of subsequent repeatability. These cluster characteristics may help distinguish subjects with a higher risk of repeatable defects.

X Giovanni Montesano @@GiovMontesano

Contributors Research design: JCKT, GM. Data acquisition and/or research execution: JCKT, JP. Data analysis and/or interpretation: JCKT, JP, GM. Manuscript preparation: JCKT, JP, KB, AA, JC, GM. JCKT is the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethics approval for the study was provided by the South-Eastern Sydney Local Health District institutional ethics review board (2024/ETH02244) and the Human Research Ethics Committee of the University of New South Wales (HC210563) and the study adhered to the tenets of the Declaration of Helsinki. The research described adhered to the National Statement on Ethical Conduct in Human Research guidelines. Waiver of consent was granted by the institutional ethics review board due to the retrospective nature of this study.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement No data are available.

ORCID iDs
Jeremy C K Tan http://orcid.org/0000-0003-4117-7446
Giovanni Montesano http://orcid.org/0000-0002-9148-2804

REFERENCES

1 Tan JCK, Yohannan J, Ramulu PY, *et al*. Visual field testing in glaucoma using the Swedish Interactive Thresholding Algorithm (SITA). *Surv Ophthalmol* 2025;70:141–52.
2 Phu J, Khuu SK, Nivison-Smith L, *et al*. Standard automated perimetry for glaucoma and diseases of the retina and visual pathways: Current and future perspectives. *Prog Retin Eye Res* 2025;104:101307.
3 Tan NYQ, Tham Y-C, Koh V, *et al*. The Effect of Testing Reliability on Visual Field Sensitivity in Normal Eyes: The Singapore Chinese Eye Study. *Ophthalmology* 2018;125:15–21.
4 Boeglin RJ, Caprioli J, Zulauf M. Long-term fluctuation of the visual field in glaucoma. *Am J Ophthalmol* 1992;113:396–400.
5 Wu Z, Medeiros FA, Weinreb RN, *et al*. Specificity of various cluster criteria used for the detection of glaucomatous visual field abnormalities. *Br J Ophthalmol* 2020;104:822–6.
6 Katz J, Sommer A, Gaasterland DE, *et al*. Comparison of analytic algorithms for detecting glaucomatous visual field loss. *Arch Ophthalmol* 1991;109:1684–9.
7 Bizios D, Heijl A, Bengtsson B. Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms. *J Glaucoma* 2007;16:20–8.
8 Budenz DL, *et al*. Sensitivity and specificity of the Swedish interactive threshold algorithm for glaucomatous visual field defects. *Ophthalmology* 2002;109:1052–8.
9 Tan JCK, Phu J, Go D, *et al*. Evaluation of the Consistency of Glaucomatous Visual Field Defects Using a Clustered SITA-Faster Protocol. *Ophthalmology* 2023;130:1138–48.
10 Tan JCK, Kalloniatis M, Phu J. Frontloading SITA-Faster Can Increase Frequency and Reliability of Visual Field Testing at Minimal Time Cost. *Ophthalmol Glaucoma* 2023;6:445–56.
11 Camp AS, Long CP, Patella VM, *et al*. Standard Reliability and Gaze Tracking Metrics in Glaucoma and Glaucoma Suspects. *Am J Ophthalmol* 2022;234:91–8.
12 Phu J, Kalloniatis M. The Frontloading Fields Study: The Impact of False Positives and Seeding Point Errors on Visual Field Reliability When Using SITA-Faster. *Transl Vis Sci Technol* 2022;11:20.
13 Heijl A, Patella VM, Flanagan JG, *et al*. False Positive Responses in Standard Automated Perimetry. *Am J Ophthalmol* 2022;233:180–8.
14 Phu J, Kalloniatis M. Gaze tracker parameters have little association with visual field metrics of intrasession frontloaded SITA-Faster 24-2 visual field results. *Ophthalmic Physiol Opt* 2022;42:973–85.
15 Tan JCK, Agar A, Kalloniatis M, *et al*. Quantification and Predictors of Visual Field Variability in Healthy, Glaucoma Suspect, and Glaucomatous Eyes Using SITA-Faster. *Ophthalmology* 2024;131:658–66.
16 Hood DC, Raza AS, de Moraes CGV, *et al*. Initial arcuate defects within the central 10 degrees in glaucoma. *Invest Ophthalmol Vis Sci* 2011;52:940–6.
17 Kim JM, Kyung H, Shim SH, *et al*. Location of Initial Visual Field Defects in Glaucoma and Their Modes of Deterioration. *Invest Ophthalmol Vis Sci* 2015;56:7956–62.
18 Werner EB, Bishop KI, Koelle J, *et al*. A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. *Arch Ophthalmol* 1988;106:619–23.
19 Tanna AP, Bandi JR, Budenz DL, *et al*. Interobserver agreement and intraobserver reproducibility of the subjective determination of glaucomatous visual field progression. *Ophthalmology* 2011;118:60–5.
20 Viswanathan AC, Crabb DP, McNaught AI, *et al*. Interobserver agreement on visual field progression in glaucoma: a comparison of methods. *Br J Ophthalmol* 2003;87:726–30.
21 Tan JCK. Coherent Interpretation of Entire Visual Field Test Reports Using a Multimodal Large Language Model (ChatGPT). *Vision (Basel)* 2025;9:33.
22 Tan JCK. n.d. Using a large language model to process biometry reports and select intraocular lens for cataract surgery. *J Cataract Refract Surg*10.
23 Mushtaq Y, Panchasara B, Nassehzadehtabriz N, *et al*. Evaluating multidisciplinary glaucoma care: visual field progression and loss of sight year analysis in the community vs hospital setting. *Eye (Lond)* 2022;36:555–63.
24 Mansberger SL, Medeiros FA, Gordon M. Diagnostic tools for calculation of glaucoma risk. *Surv Ophthalmol* 2008;53 Suppl1:S11–6.
25 Kelly SR, Bryan SR, Crabb DP. Does eye examination order for standard automated perimetry matter? *Acta Ophthalmol* 2019;97:e833–8.
26 Phu J, Khuu SK, Agar A, *et al*. Clinical Evaluation of Swedish Interactive Thresholding Algorithm-Faster Compared With Swedish Interactive Thresholding Algorithm-Standard in Normal Subjects, Glaucoma Suspects, and Patients With Glaucoma. *Am J Ophthalmol* 2019;208:251–64.
27 Heijl A, Patella VM, Chong LX, *et al*. A New SITA Perimetric Threshold Testing Algorithm: Construction and a Multicenter Clinical Study. *Am J Ophthalmol* 2019;198:154–65.
28 Pham AT, Ramulu PY, Boland MV, *et al*. The Effect of Transitioning from SITA Standard to SITA Faster on Visual Field Performance. *Ophthalmology* 2021;128:1417–25.