Combined structure-function analysis in glaucoma screening

Elina Karvonen MD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital;
Department of Ophthalmology, University of Helsinki and Helsinki University Hospital
Katri Stoor MD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital
Marja Luodonpää MD, PhD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital
Pasi Hägg MD, PhD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital
Ilmari Leiviskä BSc, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital
Johanna Liinamaa MD, PhD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital
Anja Tuulonen MD, PhD, Professor, Tays Eye Centre
Ville Saarela MD, PhD, Department of Ophthalmology, PEDEGO research unit and Medical research center, University of Oulu and Oulu University Hospital

Correspondence:
Elina Karvonen
Department of Ophthalmology
Box 21, 90029 Oulu University Hospital
Finland
tel: +358 50 427 4595
Email: elina.karvonen@fimnet.fi

Word count: Main text 3338, Tables 391

Contributorship Statement:

EK was the corresponding author. EK, KS, ML, PH and AT were involved in the diagnostic protocol. IL assisted in data management and analyses. JL and VS supervise the NFBC Eye Study. AT and VS designed the study and contributed to editing the manuscript.

Financial Support:

The Oulu University Hospital Grant no.24301140, Oulu, Finland

The University of Oulu Grant no.24000692, Oulu, Finland

ERDF European Regional Development Fund Grant no.539/2010 A31592, Brussels, Belgium

Competitive State Research Funding of Tampere University Hospital, Tampere, Finland

Glaukooma Tukisäätiö (Glaucoma Support) Lux Foundation, Helsinki, Finland

Sokeain Ystävät (Friends of Blind Persons) Foundation, Helsinki, Finland

Silmäsäätiö (Eye) Foundation, Helsinki, Finland

Association of the Finnish Ophthalmologists, Helsinki, Finland

The sponsor or funding organization had no role in the design or conduct of this research.

No conflicting relationship exists for any author.

Abbreviations and acronyms: SAP standard automated perimetry, ONH optic nerve head, RNFL retinal nerve fibre layer, SD-OCT spectral-domain optical coherence tomography, S-F structure-function, VF visual field, NFBC Northern Finland Birth Cohort, BCVA best corrected visual acuity, IOP intraocular pressure, GAT Goldmann applanation tonometry, CCT central corneal thickness, SITA Swedish interactive threshold algorithm, MLC machine learning classifier, EGS European Glaucoma Society

SYNOPSIS

The applicability of a structure-function report for glaucoma screening was evaluated in 3001 middle-aged subjects of a birth cohort. Diagnostic performance was moderate. Spatially corresponding structure and function abnormalities were significantly correlated to glaucomatous damage.

ABSTRACT

Aims: To assess the applicability of a structure-function (S-F) analysis combining spectral-domain optical coherence tomography (SD-OCT) and standard automated perimetry (SAP) in glaucoma screening in a middle-aged population.

Methods: A randomised sample of 3001 Caucasian 45-49 -year-old participants of the Northern Finland Birth Cohort (NFBC) Eye Study was examined. We performed an eye examination, including 24-2 SAP, optic nerve head (ONH) and retinal nerve fibre layer (RNFL) photography and SD-OCT of the peripapillary RNFL. The S-F report was generated by Forum Glaucoma Workplace software. OCT, SAP and the S-F analysis were evaluated against clinical glaucoma diagnosis, i.e., the positive '2 out of 3' rule based on the clinician's evaluation of ONH and RNFL photographs and visual fields (VF).

Results: At a specificity of 97.5%, the sensitivity for glaucomatous damage was 26% for abnormal OCT, 35% for SAP and 44% for S-F analysis. Estimated AUCs were 0.74, 0.85 and 0.76, and the corresponding positive predictive values were 8 %, 10 % and 12 %, respectively. By applying a classification tree approach combining OCT, SAP and defect localization data, a sensitivity of 77% was achieved at 90% specificity. In a localisation analysis of glaucomatous structural and functional defects, the correlation with glaucoma increased significantly if the abnormal VF test points were located on borderline or abnormal OCT zones.

Conclusion: SAP performs slightly better than OCT in glaucoma screening of middle-aged population. However, the diagnostic capability can be improved significantly by structure-function analysis.

Keywords: glaucoma, population screening, optical coherence tomography (OCT), standard automated perimetry (SAP), structure-function

INTRODUCTION

Glaucomatous optic neuropathy is among the world's leading causes of irreversible blindness.[1] Glaucoma largely meets the criteria for systematic screening of the World Health Organization[2]. However, screening for glaucoma has not been recommended due to the lack of a reliable test set, screening protocol[3] and evidence on the cost-effectiveness of screening among the general population.[4-11]

Although there is still no worldwide consensus for the diagnosis of glaucoma, traditionally the presence of congruent progressive structural and functional damage is required for diagnosis.[10-15] Automated threshold perimetry represents the cornerstone of the functional assessment of glaucoma. The optic nerve head (ONH) and the retinal nerve fibre layer (RNFL) are typically evaluated in a clinical examination and/or using ONH and RNFL photography.[12-16] In addition, a quantitative evaluation, most often by optical coherence tomography (OCT) with peripapillary and macular RNFL analysis, is used in many glaucoma units for assessing structural damage and its progression.[17-18] The fact that OCT has been reported to perform better in moderate than in early glaucoma[17] may pose challenges in screening for glaucoma with automated imaging.

Software applications have been developed to analyse and display the structure-function (S-F) relationship in an integrated format, e.g., a map for an average eye (Figure 1).[19] Until recently, there has been little evidence on the performance of combined reports in detecting or monitoring glaucoma. These easy-to-read reports have been evaluated mainly qualitatively to support or exclude the clinical diagnosis of glaucoma.

In this report, we analysed the automatically produced combined reports in the six zones of the average RNFL thickness (Cirrus SD-OCT) and the corresponding VF (Humphrey Field Analyzer). To our knowledge, the present study is the first to have investigated the performance of an automatically generated S-F analysis as a potential method for systematic glaucoma screening.

METHODS

Study population

In this screening trial, we report the results of 3001 participants (5964 eyes) of the Northern Finland Birth Cohort (NFBC) Eye Study examined at a single visit in 2012-2015.[20-21] The Eye Study is a satellite of the principal NFBC study evaluating the general health and the quality of life since the birth of the 1966-born cohort in the two northernmost Finnish provinces (Oulu and Lapland). Of the 10 321 NFBC participants living in Finland in 2011, 50 % (5155) were randomised to the Eye Study with the participation rate of 60% (3070). Of them, 3001 subjects had inclusive eye examination data for this report (Supplementary Figure 1). Written informed consent was obtained from all participants. The study was approved by the Ethical Committee of Northern Ostrobothnia Hospital District and adhered to the principles of the Declaration of Helsinki.

Data collection

We assessed refraction, best corrected visual acuity (BCVA), intraocular pressure (IOP) both with rebound tonometer (Icare Ltd., Vantaa, Finland) and Goldmann applanation tonometer (GAT). We obtained grayscale and color fundus images, including stereoscopic ONH photographs in mydriasis (Canon CF-60DSi Digital Mydriatic Fundus Camera and an attached Canon EOS-1Ds MK III SLR Digital Camera, Canon Inc., Tokyo, Japan). RNFL photographs were taken with a monochromatic blue interference filter (495 nm). Visual fields (VF) were examined with SITA Standard 24-2 test pattern of the Humphrey Field Analyzer II-*i* (Carl Zeiss Meditec AG, Oberkochen, Germany). Cirrus SD-OCT 4000 (software version 6.0.0; Carl Zeiss Meditec AG, Oberkochen, Germany) was used for the assessment of the anterior segment and the RNFL thickness. We evaluated the combined S-F report of the circumpapillary OCT and SAP automatically formed by Forum Glaucoma Workplace software (Carl Zeiss Meditec AG, Oberkochen, Germany).

Clinical glaucoma diagnosis as the reference standard

In the assessment of the performance of the combined S-F display, we used clinical glaucoma diagnosis as the reference standard. According to the Finnish Evidence-Based Guideline for Glaucoma[14-15], we analysed the stereoscopic ONH photographs, RNFL photographs and SAP to define glaucoma following the ‘2 out of 3’ rule (Supplementary Table 1). In the ONH stereo photographs, the disc, the appearance of the neuroretinal rim and potential hemorrhages were assessed. In the RNFL photographs, a local dark wedge-shaped area or a diffuse poorly visible RNFL pattern were considered as defects, paying attention to possible asymmetry

between the eyes and the hemispheres. A glaucomatous VF defect was defined as 1) at least three adjacent test points at p<5% and at least one point of these at p<1% on the pattern deviation plot (the '5-5-1 cluster'), 2) the glaucoma hemifield test (GHT) graded as 'outside normal limits', and/or 3) the pattern standard deviation abnormal at p<5% level. Reliability indices less than 20% for fixation losses and 15% for false positive responses were demanded for a reliable VF.

An ONH notch, an RNFL damage or a typical VF defect were each considered as one out of three glaucomatous findings (Supplementary Table 1). The reference standard of this report, i.e., definite glaucoma, required congruent glaucomatous changes in at least two out of these three findings. IOP was not included in glaucoma diagnosis.

A total of 3001 persons (5964 eyes) were included. The demographics are presented in Supplementary Table 2. The comprehensive evaluation protocol and the results of the clinical assessment have been described in detail earlier.[20-21] Definite glaucoma was diagnosed in 43 eyes of 33 persons (1.1% [95% CI 0.8-1.5]). Of them, only four participants (12%) had a prior glaucoma diagnosis. The majority (73%) of the glaucomatous subjects had mild glaucoma (SAP mean deviation [MD] equal to or higher than -6.0 dB). Their median (range) MD and pattern standard deviation (PSD) were -2.0 dB (-22.2 to 0.9) and 2.7 dB (1.2 to 14.1). The mean IOP in the glaucomatous eyes was 16.6 ± 5.4 mmHg (range 9-35) and 88% of them were normotensive at the time of the eye examination. The glaucomatous eyes had a spherical equivalent of -1.9 ± 3.5 D. The mean angle opening distance (AOD) assessed 750 μm from the anterior chamber angle was 965 ± 336 μm in glaucomatous eyes and 890 ± 320 μm in healthy eyes (p=0.295). The smallest AOD in a glaucomatous eye was 346 μm indicating an open angle in all measured eyes.

Structure analysis with OCT

The circumpapillary RNFL thickness was analysed in the six zones of the combined report (Figure 1, Supplementary Table 3).[19] Compared to the normative database, RNFL thickness over the 95th percentile and between the 5th and the 95th percentile were regarded as normal (color-coded as white or green,

respectively). Measures below the 5th percentile were rated as borderline (yellow) and those below the 1st percentile as abnormal (red). We explored the color-coded zones in a gradually worsening order starting from one at least borderline (yellow or red) zone to six abnormal (red) zones (Supplementary Table 3) and calculated statistical measures of their performance (Figure 2 A-B, Table 1).

Approx. cut-off specificities	Specificity (%) (95% CI)	Sensitivity (%) (95% CI)	PPV (%)	Number of defects	Combination of defective locations[1] for cut-off
OCT defect					
97.5%	97.7 (97.3 – 98.1)	25.6 (13.5 – 41.2)	7.5	≥ 3	2 yellow/red zones, 1 red zone
95%	94.5 (93.9 – 95.0)	34.9 (21.0 – 50.9)	4.4	≥ 2	1 yellow/red zone, 1 red zone
90%	91.7 (91.0 – 92.4)	44.2 (29.1 – 60.1)	3.7	≥ 2	2 yellow/red zones
80%[2]	-	-	-	-	-
VF defect					
97.5%	97.6 (97.2 – 97.6)	34.9 (21.0 – 50.9)	9.6	≥ 9	<5%, <2%, <1%, <0.5% (×6)
95%	95.0 (94.3 – 95.5)	53.5 (37.7 – 68.8)	7.1	≥ 7	<5%, <2%, <1% (×2), <0.5% (×3)
90%	90.8 (90.0 – 91.5)	58.1 (42.1 – 73.0)	4.4	≥ 5	<5%, <2%, <1%, <0.5% (×2)
80%	82.7 (81.6 – 83.6)	72.1 (56.3 – 84.7)	2.9	≥ 4	<5%, <2%, <1%, <0.5%
VF cluster analysis					
97.5%	97.6 (97.2 – 98.0)	34.9 (21.0 – 50.9)	9.6	≥ 9	<5%, <2%, <1%, <0.5% (×6)
95%	94.7 (94.1 – 95.3)	55.8 (39.9 – 70.9)	7.1	≥ 7	<5%, <2% (×2), <1%, <0.5% (×3)
90%	91.5 (90.7 – 92.2)	58.1 (42.1 – 73.0)	4.7	≥ 5	<5%, <2%, <1%, <0.5% (×2)
80%	79.5 (78.4 – 80.5)	76.7 (61.4 – 88.2)	2.4	≥ 4	<5%, <2% (×2), <1%
Structure-function defect					
97.5%	97.7 (97.3 – 98.1)	44.2 (29.1 – 60.1)	12.3	≥ 7 + 1	<5% (×2), <2%, <1%, <0.5% (×3) and 1 yellow/red zone
95%	95.0 (94.5 – 95.6)	51.2 (35.5 – 66.7)	7.0	≥ 5 + 1	<5%, <2% (×2), 1%, 0.5% and 1 yellow/red zone
90%	89.2 (88.4 – 90.0)	62.8 (46.7 – 77.0)	4.0	≥ 4 + 1	<5% (×2), <2% <1% and 1 yellow/red zone
80%	81.5 (80.5 – 82.5)	65.1 (49.1 – 79.0)	2.5	≥ 2 + 1	<5%, <2% and 1 yellow/red zone

Table 1. Sensitivities at the cut-off specificities of 80%, 90%, 95% and 97.5%. These cut-offs were rounded to the nearest applicable figure due to the non-continuous nature of the data. 'Visual field defects' refer to non-clustered defective VF test points whereas 'cluster analysis' refers to at least three tangential defective VF test points. The number and severity of defective test locations are presented.

[1]Combination of defective test points on the SAP pattern deviation plot at p<5%, 2%, 1% or 0.5% and/or number and depth of defective OCT zones.

[2]For OCT defects, no suitable combination was found with the specificity of 80%.

Function analysis with SAP

We evaluated sensitivity values of all 52 test locations of the SAP pattern deviation plot expressed on a logarithmic scale (dB). In the function analysis, the test points were automatically grouped into six sectors: superotemporal, superonasal, nasal, inferonasal, inferotemporal and temporal, following the S-F correspondence map (Figure 1).[19] We assessed the number, location and depth of the defective points on a gradually worsening scale (Supplementary Table 4), ranging from a single test point at $p<5\%$ to different combinations of at least eleven abnormal points. Sensitivity, specificity and AUC for these VF defects were calculated (Figure 2 C-D, Table 1).

Cluster analysis

We also explored glaucomatous VF defects with at least three tangentially clustered points on the pattern deviation plot (Figure 2 E-F, Table 1) to assess the value of cluster criterion in detecting glaucoma with SAP.

Combined structure-function analysis

We evaluated the peripapillary RNFL thickness in the six zones superimposed with the 24-2 SAP in a combined report provided by Forum Glaucoma Workplace 2.0 (Figure 1). We used the statistical grading of the RNFL provided by the Cirrus software. We evaluated whether defective VF test points located on borderline or abnormal OCT zones would be associated with glaucomatous damage more accurately than a defective point on any colored zone (Figure 2 G-H, Figure 3, Table 1, Supplementary Figure 2, Supplementary Table 5). The significance of the location and S-F correlation of the defective points was also assessed. Moreover, in order to demonstrate the glaucoma screening process, a classification tree was created (Figure 4).

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics (version 25 SPSS Inc., Armonk, NY, USA) and R programming language (version 4.0.1). All eyes with successful examinations were included in the analyses. We used a log-linear approach, i.e., VF values were evaluated on a logarithmic scale and RNFL data on a linear scale. Normally distributed data were reported as mean and standard deviation (SD), whereas non-

normally distributed data were expressed as median and range. We used the chi-square test to compare categorical data, the t-test for normally distributed data and the Mann-Whitney test for non-normally distributed data. The alpha level (type 1 error) was set at 0.05. The sensitivity, specificity and estimated AUCs for the different potential screening parameters were calculated using the clinical glaucoma diagnosis as a reference standard (Figure 2, Table 1). Due to the categorical nature of the data, an estimation of the AUC was calculated by interrupted sensitivity-specificity pairs and therefore rounded cut-off specificities are reported (Figure 2, Table 1, Supplementary Tables 3, 4 and 5). A defect localization analysis was conducted to illustrate the correlation of the functional defect with the corresponding OCT zone exhibiting glaucomatous damage (Figure 3, Supplementary Figure 2). A decision tree was created to demonstrate the most relevant steps in detecting glaucomatous damage in this population when RNFL thickness and VF sensitivity were combined (Figure 4).

RESULTS

Structure analysis (OCT)

There was a statistically significant difference in the RNFL thicknesses between glaucomatous and non-glaucomatous eyes in the five sectors not including the temporal zone 6 (Figure 1, Supplementary Table 2). The majority of the abnormal OCT findings had high specificity of 98-99% with poor sensitivity of 26% or less (Figure 2 A-B, Supplementary Table 3). The highest sensitivity was 67% with a corresponding specificity of 74% if at least one yellow zone or worse was observed. Sensitivities at fixed specificities are presented in Table 1. Most commonly, zone 2 (inferotemporal RNFL) and least frequently, zones 5 and 6 (temporal and nasal RNFL, respectively) were affected.

Function analysis (SAP)

We included a total of 38 options for VF defects when evaluating findings from one to at least eleven defective points observed on the pattern deviation plot (Supplementary Table 4). VF defects were also arranged on a ROC curve in a worsening order, starting from a single defective point with P <5% (Figure 2 C-F). We explored both randomly organized (Figure 2 C-D) and clustered (Figure 2 E-F) defective test points. In the

case of few abnormal points, the adoption of a cluster criterion improved the specificity. Nonetheless, in the case of six or more defective points, there was no difference in accuracy between these two methods. Sensitivities at fixed specificities are presented in Table 1. The '5-5-1 cluster', a common VF criterion for a glaucomatous defect, resulted in a sensitivity of 77% and a specificity of 75%.

Combined structure-function analysis

The S-F association of damage increased significantly if defective VF test points were located on a zone with at least borderline RNFL thickness (yellow or red). The localization of the VF defect on the abnormal (red) zone did not increase the association with glaucomatous damage any further (Figure 3, Supplementary Figure 2). Sensitivities at fixed specificities are presented in Table 1.

Classification tree

We analysed and chose five cut-offs in the form of a classification tree (Figure 4). The number, depth and clustering of the defective VF test points, and the number, severity and location of the abnormal OCT zones along with their combinations were considered. The specificity of the five-step process was 90% with a sensitivity of 77%. By applying this approach, 33 of 43 glaucomatous eyes were detected, whereas 610 of 5921 healthy eyes were misdiagnosed as having glaucoma.

DISCUSSION

We report the performance of a structure-function analysis in a middle-aged population-based cohort to evaluate the potential advantage of an easy-to-read combined report in glaucoma screening. To the best of our knowledge, this is the first study evaluating the applicability of the automated analysis by Cirrus SD-OCT and Humphrey SAP in glaucoma screening.

Firstly, we assessed several combinations of borderline and abnormal OCT zones to obtain the best accuracy in detecting glaucomatous damage (Figure 2 A-B, Table 1). In agreement with the previous literature[18, 22-24], inferotemporal RNFL defects were most frequent. With a specificity of 95% or more, involving at least

two defective zones and including at least one red zone, the sensitivity was poor (35%). In our recent publication, abnormal (red) peripapillary or macular RNFL analysis or their combination resulted in sensitivities of 53%, 50% or 61% with corresponding specificities of 95%, 92% or 90%, respectively.[25] Therefore, this evidence does not support the use of sole automated imaging in general population screening. The Choosing Wisely Recommendations by the Finnish Guidelines and the European Glaucoma Society (EGS) similarly warn against defining a diagnosis or progression of glaucoma based only on OCT, which provides mostly a statistical deviation from a reference database.[11-12]

Secondly, we analysed the number and depth of the defective test points on the SAP pattern deviation plot with two approaches: 1) regardless of their localization, or 2) using the cluster criterion (Figure 2 C-F, Table 1). When we applied a specificity of 95%, the sensitivities were 54% and 56%, respectively. Surprisingly, at least three contiguous defective VF points did not perform any better than randomly organized defective points in detecting glaucoma. In the previous literature, combining structure and function performed significantly better than the best independent structural and functional measurements[12-13, 26] even if VF parameters (e.g., PSD) without any spatial correspondence to the structural measures were included.[26-27] The early diagnosis of glaucoma is still a challenge, as even up to 35% of the retinal ganglion cells (RGCs) may be lost before there are apparent signs of damage (PSD at $p<0.5\%$) in the SAP.[28] Nonetheless, in our study, SAP surpassed OCT in detecting early glaucoma.

Thirdly, the main aim of this report was to evaluate the performance of the automated structure-function report. The S-F analysis outperformed the single measures by only a relatively narrow margin (Figure 2, Table 1). However, a localization analysis (Figure 3) showed the importance of corresponding VF and OCT defects. The surprising failure to detect a difference between at least borderline (yellow or red) and abnormal (red) OCT zones in the combined analysis may reflect the mild glaucomatous damage detected in most cases. We included both borderline and abnormal OCT measurements to find optimal sensitivity and specificity on the gradually worsening OCT defect curve (Figure 2, Supplementary Tables 2 and 4). To the contrary, in a recent

worldwide consensus on objective VF and OCT criteria for glaucomatous optic neuropathy, only abnormal (red) RNFL zones were regarded as glaucomatous.[18]

A classification tree was created to demonstrate the structure-function analysis in a real-life scenario (Figure 4). This kind of decision aid could be applied in everyday use as an inexpensive, automated and simple glaucoma screening tool utilized e.g., by ophthalmic technicians in glaucoma detection.[6] An interesting approach would be to apply machine learning classifiers (MLCs) for analyzing OCT and SAP data for screening purposes. MLCs have been reported to outperform general ophthalmologists in detecting early to moderate glaucoma with AUC values of 0.93 and 0.88, respectively.[29] Although our classification tree with five branches was able to separate the glaucomatous and non-glaucomatous eyes the specificity of 90% and sensitivity of 77% would still be way too modest to establish a screening program for the general population.

The poor structure-function correlation may be explained by the fact that mapping is based on a standard eye. In particular, a high axial length, ocular torsion and angle of temporal raphé (the line dividing the superior and inferior RNFL) have been reported to influence the mapping of the RNFL thickness and the corresponding VF area.[19, 30] The S-F agreement has been suggested to be improved by including disc and macular OCT scans along with 24-2 and 10-2 VFs but convincing evidence is missing.[27, 31-32]

Combined maps have been available for almost two decades. Nonetheless, S-F mapping has been difficult to validate, especially without an objective reference standard. Many studies have focused on measuring the overall correlation between RNFL thickness and VF sensitivity.[19, 22-23, 26, 33] Cui et al.[34] reported an agreement rate of 75% and kappa statistic of 0.62 between glaucoma specialists and automated S-F analysis combining Spectralis SD-OCT and Heidelberg Edge Perimeter data. Similarly, Hood et al. reported an excellent inter-rater agreement of two experienced non-ophthalmologists assessing one-page reports including data from Topcon OCT macular and ONH scans along with Humphrey 24-2 SAP.[35] However, the above-mentioned study populations of 45 and 50 glaucomatous subjects, respectively, differ from our general population.

We found relatively poor sensitivities for all our diagnostic parameters. Although the sensitivity of a screening test is important, we additionally expect reasonable specificity to minimize subsequent over-diagnostics, unwarranted burden and costs to society, especially in the case of a non-fatal and, in many cases, non-aggressive disease (e.g., open-angle glaucoma). When resources are limited, the 14-fold false positives compared to the true positives (Figure 4) is alarming and would burden both patients and the system. Prevent Blindness America has recommended that a screening test should have at least 85% sensitivity and 95% specificity to identify all moderate and advanced glaucoma cases and be able to detect the majority of the early cases.[36]

According to our study, the accuracy of the combined S-F analysis is poor to assess the general population. Nevertheless, the diagnostic accuracy is always dependent on the reference standard. As the prevalence of definite glaucoma in this cross-sectional analysis was relatively low, in the future, we will be able to use the long-term follow-up data of this cohort as a golden standard for re-assessing the long-term value of various screening tests.

Screening particular risk groups, e.g., individuals with a family history of glaucoma, African descent, persons with myopia or diabetes, could be more useful. In a study conducted in the UK, Burr et al. estimated that screening of the above-mentioned high-risk individuals would include 37% of the country's population.[5] The prevalence of glaucoma would still be too low to support cost-effectiveness of screening compared to the opportunistic case finding. In a Finnish simulation model, screening was reported as being cost saving among the elderly, i.e., aged 75-79 years, the main explanation being that in the model treatment was discontinued in non-glaucomatous subjects.[7] The cost-effectiveness of screening in the NFBC Eye Study will be evaluated in the future when also the non-screened group will be examined and the visual disability compared with initially screened group.

Since the purpose of this study was to assess a potential screening tool among the population, some compromises in the diagnostic protocol were inevitable. Although repeated testing should be performed to verify a glaucomatous VF defect, the study protocol included only one VF examination to mimic a real-life glaucoma screening. We included all cohort members attending the eye examinations with test results enabling glaucoma diagnostics. For example, we did not exclude eyes with profound refractive errors. VF data was evaluated as a part of the reference standard and as an index test. However, the index tests were evaluated by a person masked to the reference standard. Similarly, the definition of reference standard was masked to the index test analyses. Our study population was a geographically defined, middle-aged Caucasian cohort and the results may not be generalized to other more diverse populations.

The NFBC Eye Study is the first randomised population-based screening trial for glaucoma. In this cross-sectional analysis, we did not find an acceptable diagnostic cut-off by using only OCT or SAP in the glaucoma screening. Nonetheless, this report suggests that their combination may aid in the screening and diagnosis of glaucoma. However, further studies will be needed to assess the applicability of the structure-function analysis, especially if it is applied among the elderly as well as in glaucoma risk groups.

ACKNOWLEDGEMENTS

We express our warmest thanks to the late Professor Paula Rantakallio (founder of the NFBC 1966), the NFBC project center, the study staffs in Oulu University Hospital and Tampere University Hospital and the cohort members participating the current 46-year study. NFBC 1966 has received financial support from the Oulu University Hospital Grant no.24301140, the University of Oulu Grant no.24000692, ERDF European Regional Development Fund Grant no.539/2010 A31592 and the Competitive State Research Funding of Tampere University Hospital. NFBC Eye Study has been supported by Glaukooma Tukisäätiö (Glaucoma Support) Lux Foundation, Sokeain Ystävät (Friends of Blind Persons) Foundation, Silmäsäätiö (Eye) Foundation and the Association of the Finnish Ophthalmologists. The authors do not have any financial disclosures related to this study.

REFERENCES

1. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–90. doi:10.1016/j.ophtha.2014.05.013
2. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*. 2008;86:317–9. doi:10.2471/blt.07.050112
3. Mowatt G, Burr JM, Cook JA, et al. Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis. *Invest Ophthalmol Vis Sci*. 2008;49:5373-85. doi:10.1167/iovs.07-1501
4. Hatt S, Wormald R, Burr J. Screening for prevention of optic nerve damage due to chronic open angle glaucoma. *Cochrane Database Syst Rev*. 2006;18:CD006129. doi:10.1002/14651858.CD006129.pub2
5. Burr JM, Mowatt G, Hernández R, et al. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technol Assess.* 2007;11:1-190. doi:10.3310/hta11410
6. Burr J, Hernández R, Ramsay C, et al. Is it worthwhile to conduct a randomized controlled trial of glaucoma screening in the United Kingdom? *J Health Serv Res Policy* 2014;19:42-51. doi:10.1177/1355819613499748
7. Vaahtoranta-Lehtonen H, Tuulonen A, Aronen P, et al. Cost effectiveness and cost utility of an organized screening programme for glaucoma. *Acta Ophthalmol Scand*. 2007;85:508–18. doi:10.1111/j.1600-0420.2007.00947.x
8. Hernández R, Rabindranath K, Fraser C, et al. Screening for open angle glaucoma: systematic review of cost-effectiveness studies. *J Glaucoma.* 2008;17:159–68. doi:10.1097/IJG.0b013e31814b9693
9. Tuulonen A. Cost-effectiveness of screening for open-angle glaucoma in developed countries. *Indian J Ophthalmol*. 2011;59:S24–30. doi:10.4103/0301-4738.73684

10. Weinreb R, Healey P, Topouzis F (ed.). Glaucoma screening. In: *World Glaucoma Association Consensus Series 5*. First edition. Amsterdam, The Netherlands: Kugler Publications, 2018. https://wga.one/wga/consensus-5/ (Accessed May 9, 2021).
11. Tuuminen R, Sipilä R, Komulainen J, et al. The first ophthalmic Choosing Wisely recommendations in Finland for glaucoma and wet age-related macular degeneration. *Acta Ophthalmol*. 2019;97:e808-10. doi:10.1111/aos.14031
12. European Glaucoma Society. Terminology and Guidelines for Glaucoma. 5th edition. ISBN 978-88-98320-37-0. PubliComm, Savona, Italy, GECA Srl, 2020. https://www.eugs.org/eng/guidelines.asp (Accessed May 9, 2021).
13. Foster PJ, Buhrmann R, Quigley HA, et al. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol.* 2002;86:238–42. doi:10.1136/bjo.86.2.238
14. Tuulonen A, Airaksinen PJ, Erola E, et al. The Finnish evidence-based guideline for open-angle glaucoma. *Acta Ophthalmol Scand.* 2003;81:3-18. doi:10.1034/j.1600-0420.2003.00021.x
15. Tuulonen A, Forsman E, Hagman J, et al. [Update on current care guideline: glaucoma]. *Duodecim.* 2015:131:356-8 (in Finnish). English version: www.kaypahoito.fi/wp-content/uploads/sites/15/2019/01/Glaucoma-Current-Care-Guideline-2014-1.pdf (Accessed May 9, 2021).
16. Gedde S, Vinod K, Wright M, et al. Primary Open-Angle Glaucoma Preferred Practice Pattern. *Ophthalmology* 2021;128:P71-150. org/10.1016/j.ophtha.2020.10.022 (Accessed May 9, 2021)
17. Michelessi M, Li T, Miele A, et al. Accuracy of optical coherence tomography for diagnosing glaucoma: an overview of systematic reviews. *Br J Ophthalmol.* 2021;105:490-5. doi: 10.1136/bjophthalmol-2020-316152.
18. Iyer JV, Boland MV, Jefferys J, et al. Defining glaucomatous optic neuropathy using objective criteria from structural and functional testing [published online ahead of print, 2020 Jul 22]. *Br J Ophthalmol*. 2020. doi:10.1136/bjophthalmol-2020-316237
19. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107:1809–15. doi:10.1016/s0161-6420(00)00284-0

20. Saarela V, Karvonen E, Stoor K, et al. The Northern Finland birth cohort eye study: design and baseline characteristics. *BMC Ophthalmol.* 2013;13:51. doi:10.1186/1471-2415-13-51
21. Karvonen E, Stoor K, Luodonpää M, et al. Prevalence of glaucoma in the Northern Finland birth cohort eye study. *Acta Ophthalmol.* 2019;97:200-7. doi:10.1111/aos.13912
22. Nilforushan N, Nassiri N, Moghimi S, et al. Structure-function relationships between spectral-domain OCT and standard achromatic perimetry. *Invest Ophthalmol Vis Sci*. 2012;53:2740–8. doi:10.1167/iovs.11-8320
23. Wu H, de Boer JF, Chen L, Chen TC. Correlation of localized glaucomatous visual field defects and spectral domain optical coherence tomography retinal nerve fiber layer thinning using a modified structure-function map for OCT. *Eye (Lond)*. 2015;29:525–33. doi:10.1038/eye.2014.317
24. Virgili G, Michelessi M, Cook J, et al. Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the GATE study. *Br J Ophthalmol*. 2018;102:604–10. doi:10.1136/bjophthalmol-2017-310642
25. Karvonen E, Stoor K, Luodonpää M, et al. Diagnostic performance of modern imaging instruments in glaucoma screening. *Br J Ophthalmol.* 2020 Oct;104:1399-1405. doi:10.1136/bjophthalmol-2019-314795.
26. Shah NN, Bowd C, Medeiros FA, et al. Combining structural and functional testing for detection of glaucoma. *Ophthalmology*. 2006;113:1593–1602. doi:10.1016/j.ophtha.2006.06.004
27. Wu Z, Medeiros FA, Weinreb RN, Zangwill LM. Performance of the 10-2 and 24-2 Visual Field Tests for Detecting Central Visual Field Abnormalities in Glaucoma. *Am J Ophthalmol*. 2018;196:10–7. doi:10.1016/j.ajo.2018.08.010
28. Kerrigan-Baumrind LA, Quigley HA, Pease ME, et al. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci*. 2000;41(3):741–8. PMID: 10711689
29. Shigueoka LS, Vasconcellos JPC, Schimiti RB, et al. Automated algorithms combining structure and function outperform general ophthalmologists in diagnosing glaucoma. *PLoS One*. 2018;13:e0207784. doi:10.1371/journal.pone.0207784

30. McKendrick AM, Denniss J, Wang YX, et al. The Proportion of Individuals Likely to Benefit from Customized Optic Nerve Head Structure-Function Mapping. *Ophthalmology*. 2017;124:554–61. doi:10.1016/j.ophtha.2016.12.016
31. Hood DC, Tsamis E, Bommakanti NK, et al. Structure-Function Agreement Is Better Than Commonly Thought in Eyes With Early Glaucoma. *Invest Ophthalmol Vis Sci*. 2019;60:4241-8. doi:10.1167/iovs.19-27920.
32. West ME, Sharpe GP, Hutchison DM, et al. Utility of 10-2 Visual Field Testing in Glaucoma Patients with Early 24-2 Visual Field Loss. *Ophthalmology* 2021;128:545-53. doi:10.1016/j.ophtha.2020.08.033
33. Ballae Ganeshrao S, Turpin A, Denniss J, McKendrick AM. Enhancing Structure-Function Correlations in Glaucoma with Customized Spatial Mapping. *Ophthalmology*. 2015;122:1695–1705. doi:10.1016/j.ophtha.2015.04.021
34. Cui QN, Fudemberg SJ, Resende AF, et al. Validation of the structure-function correlation report from the heidelberg edge perimeter and spectral-domain optical coherence tomography. *Int Ophthalmol*. 2019;39:533–40. doi:10.1007/s10792-018-0836-z
35. Hood DC, Raza AS, De Moraes CG, et al. Evaluation of a One-Page Report to Aid in Detecting Glaucomatous Damage. *Transl Vis Sci Technol*. 2014;3:8. doi:10.1167/tvst.3.6.8.
36. Sponsel WE, Ritch R, Stamper R, et al. Prevent Blindness America visual field screening study. The Prevent Blindness America Glaucoma Advisory Committee. *Am J Ophthalmol.* 1995;120:699-708. doi:10.1016/s0002-9394(14)72723-0

FIGURE LEGENDS

Figure 1. The six zones of the peripapillary RNFL thickness are demonstrated in the fundus view of a left eye (A and B). In the combined structure-function display (E), OCT data is inverted around the horizontal axis (C) to match the corresponding SAP 24-2 data (D), i.e., inferior nerve fibres are presented in the superior part of the S-F display. The measured RNFL zones are as follows: superotemporal (41° – 80°), superonasal (81° – 120°), nasal (121° – 230°), inferonasal (231° – 270°), inferotemporal (271° – 310°) and temporal (311° - 40°).

This correspondence map was introduced by Garway-Heath et al. in 2000.[19] Due to the greater density of the nerve fibres on the superior and inferior poles of the optic disc, these ONH sectors are outlined as being narrower, although they represent a larger VF area. Thus, the narrower superotemporal, superonasal, inferonasal and inferotemporal sectors each comprise a 40° rim area whereas the nasal and temporal sectors consist of a 110° and 90° area, respectively.

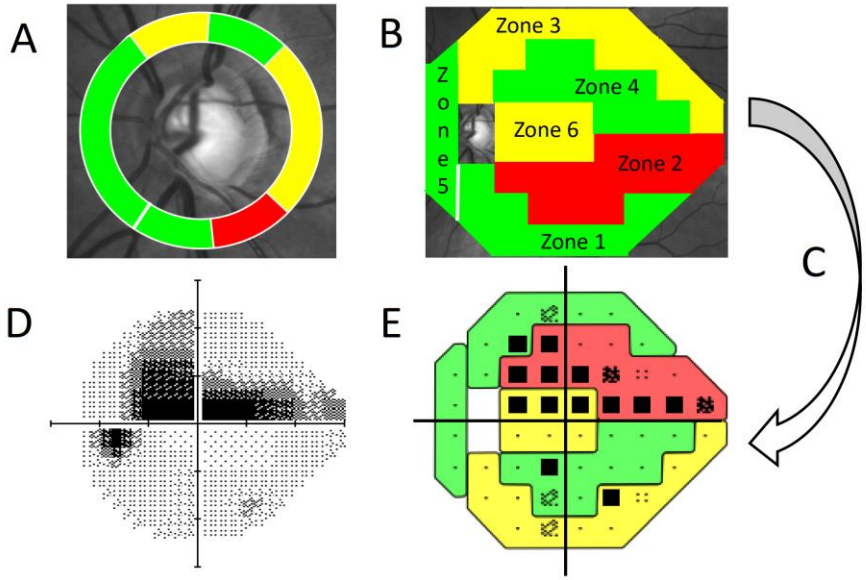
Figure 2. The performance of the various observed OCT abnormalities (A and B), visual field defects (C-F) and structure-function analysis (G-H) detecting definite glaucoma are presented in a gradually worsening order. For VF data, both randomly organized (C-D) and clustered (E-F) defective points are introduced. In the left column, sensitivities and specificities are expressed from 0% to 100% and in the right column, especially the highest specificities of 90-100% are displayed. The explanations for the single points on a ROC curve are described in Supplementary Tables 3, 4 and 5.

Figure 3. Defect localization analysis combining visual field defective points with a pattern deviation at $p<5\%$, $<2\%$, $<1\%$ or $<0.5\%$ and a simultaneous OCT finding. Correlation coefficients are presented on a grayscale here and in numeric values in Supplementary Figure 2. The data of the right eyes were evaluated as a mirror image to present all of the data uniformly as left eyes in this analysis.

Figure 4. The classification tree for screening of glaucoma in this study population. Demonstrative structure-function images of glaucomatous eyes in this cohort are presented.

[1] Zone 2 (inferotemporal) and zone 3 (superonasal) are presented in Figure 1.

502 Figure 1.



503

504

505

506

507

508

509

510

511

512

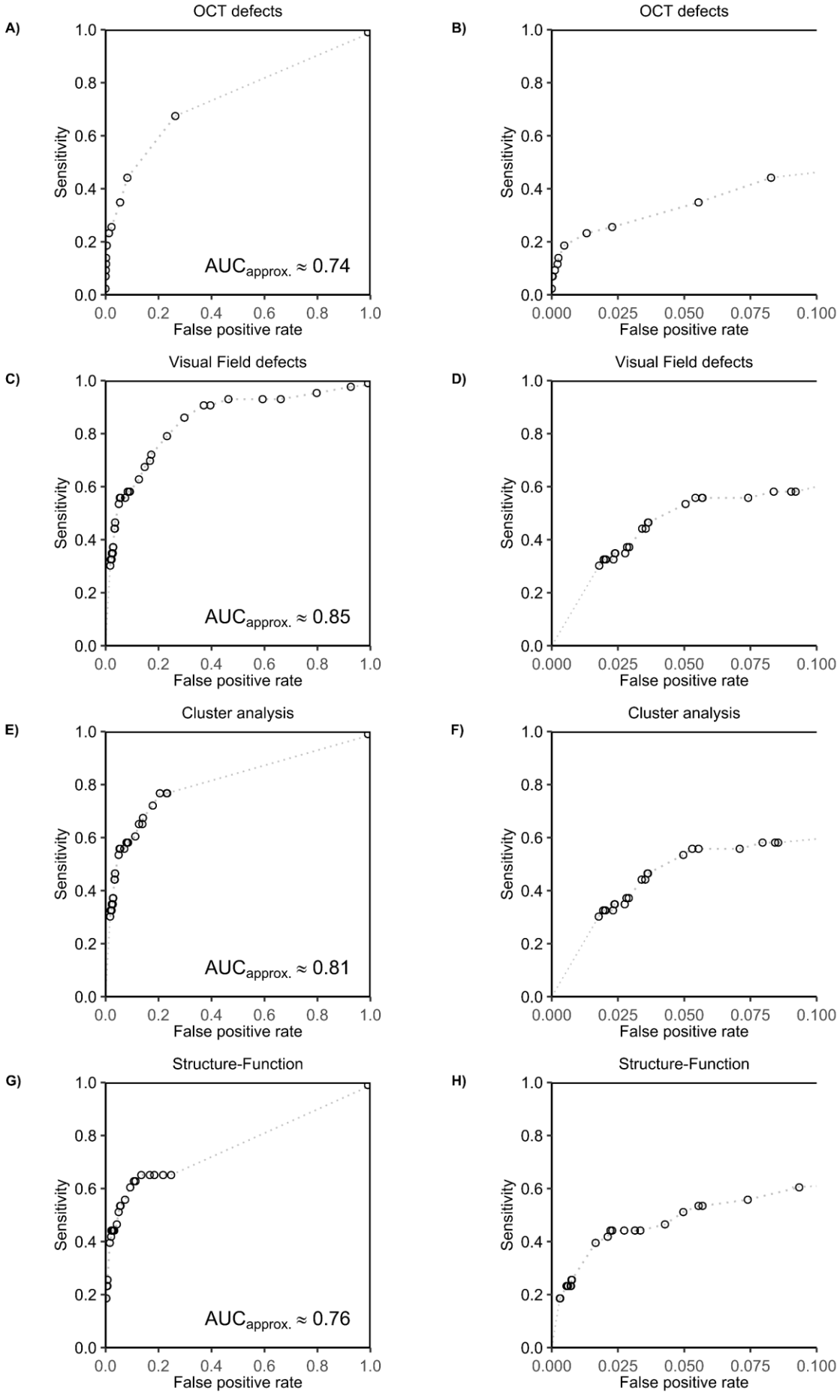
513

514

515

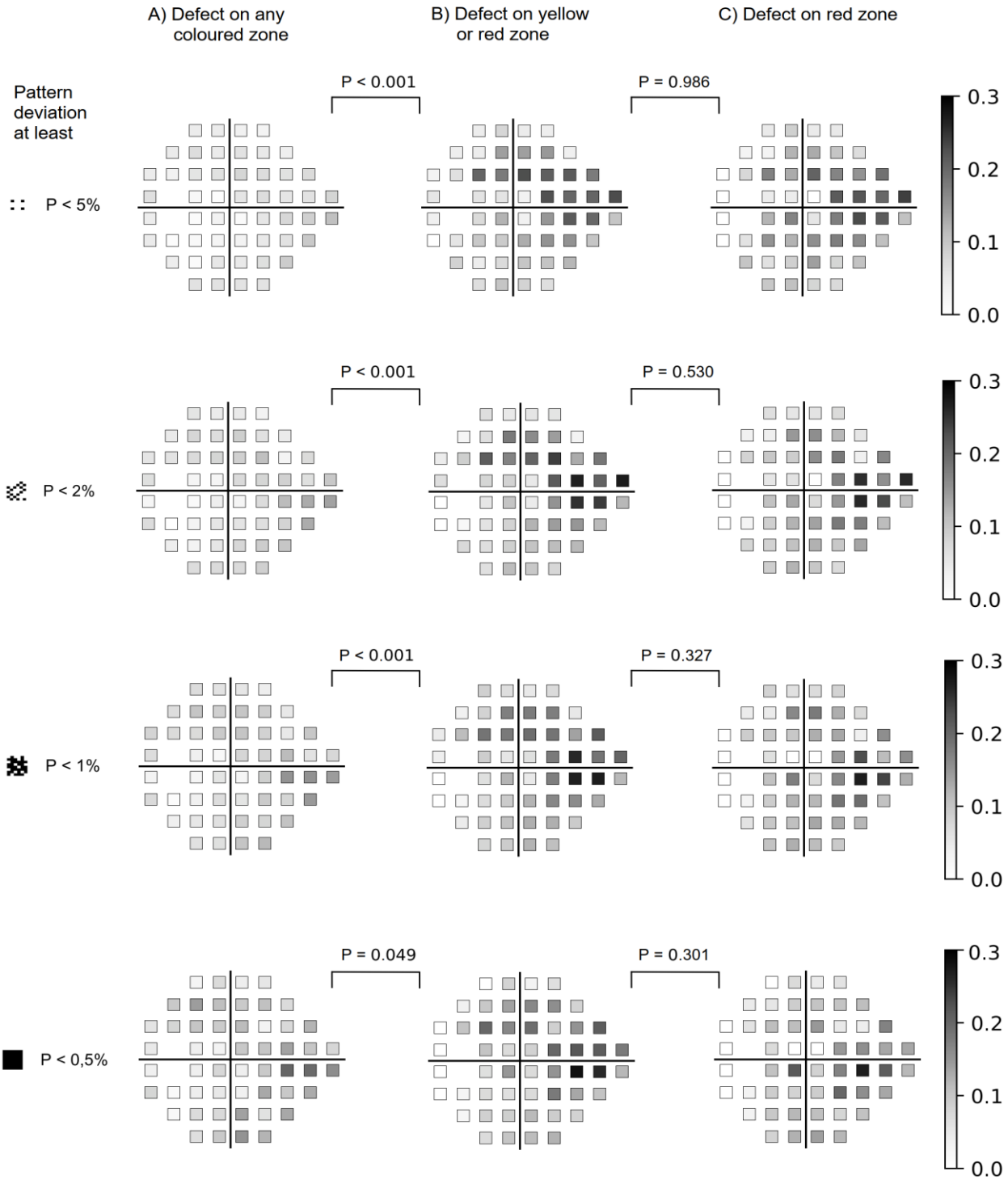
516

517 Figure 2.

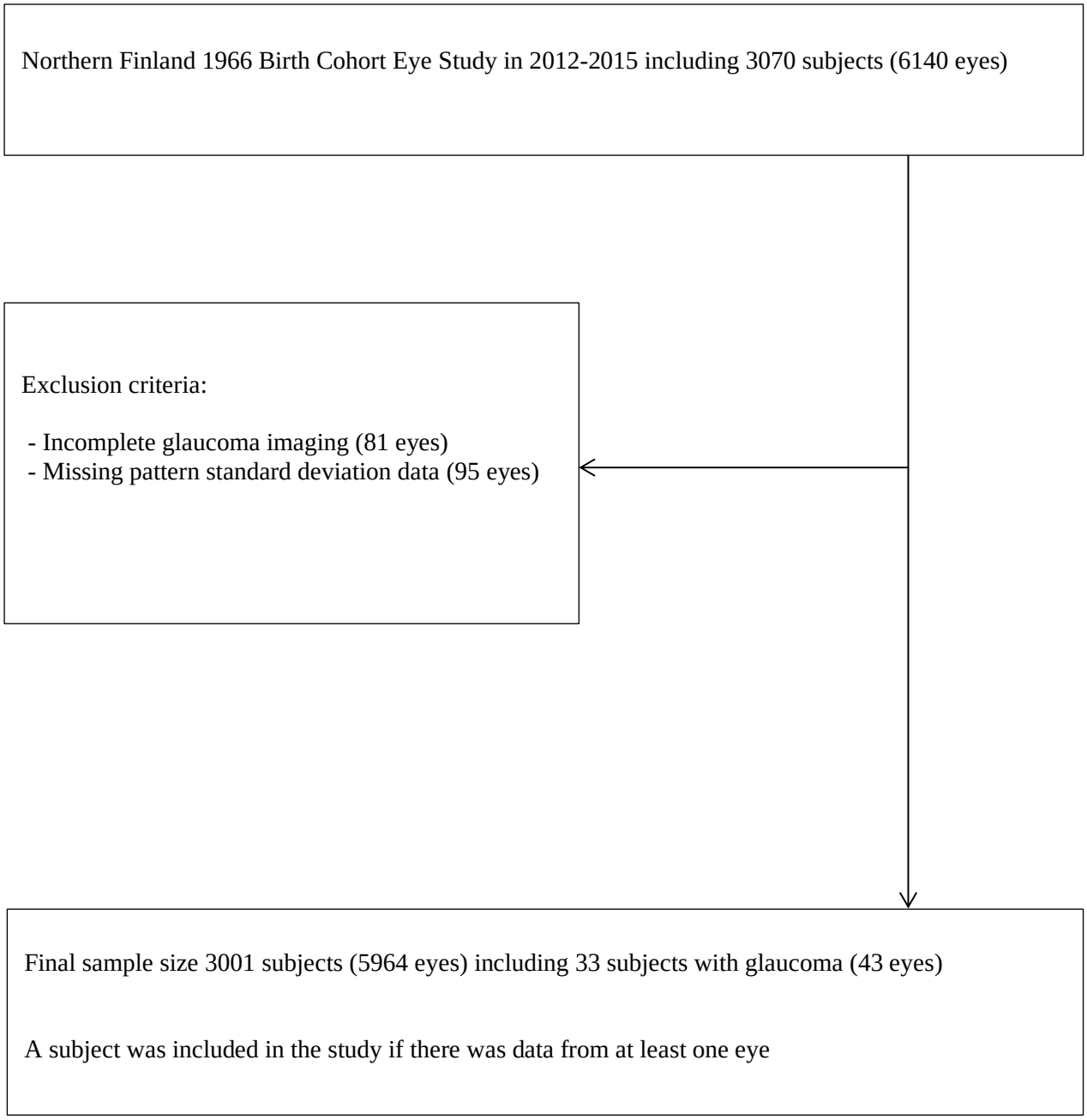

A)
OCT defects
Sensitivity
False positive rate
AUC approx. ≈ 0.74
B)
OCT defects
Sensitivity
False positive rate
C)
Visual Field defects
Sensitivity
False positive rate
AUC approx. ≈ 0.85
D)
Visual Field defects
Sensitivity
False positive rate
E)
Cluster analysis
Sensitivity
False positive rate
AUC approx. ≈ 0.81
F)
Cluster analysis
Sensitivity
False positive rate
G)
Structure-Function
Sensitivity
False positive rate
AUC approx. ≈ 0.76
H)
Structure-Function
Sensitivity
False positive rate

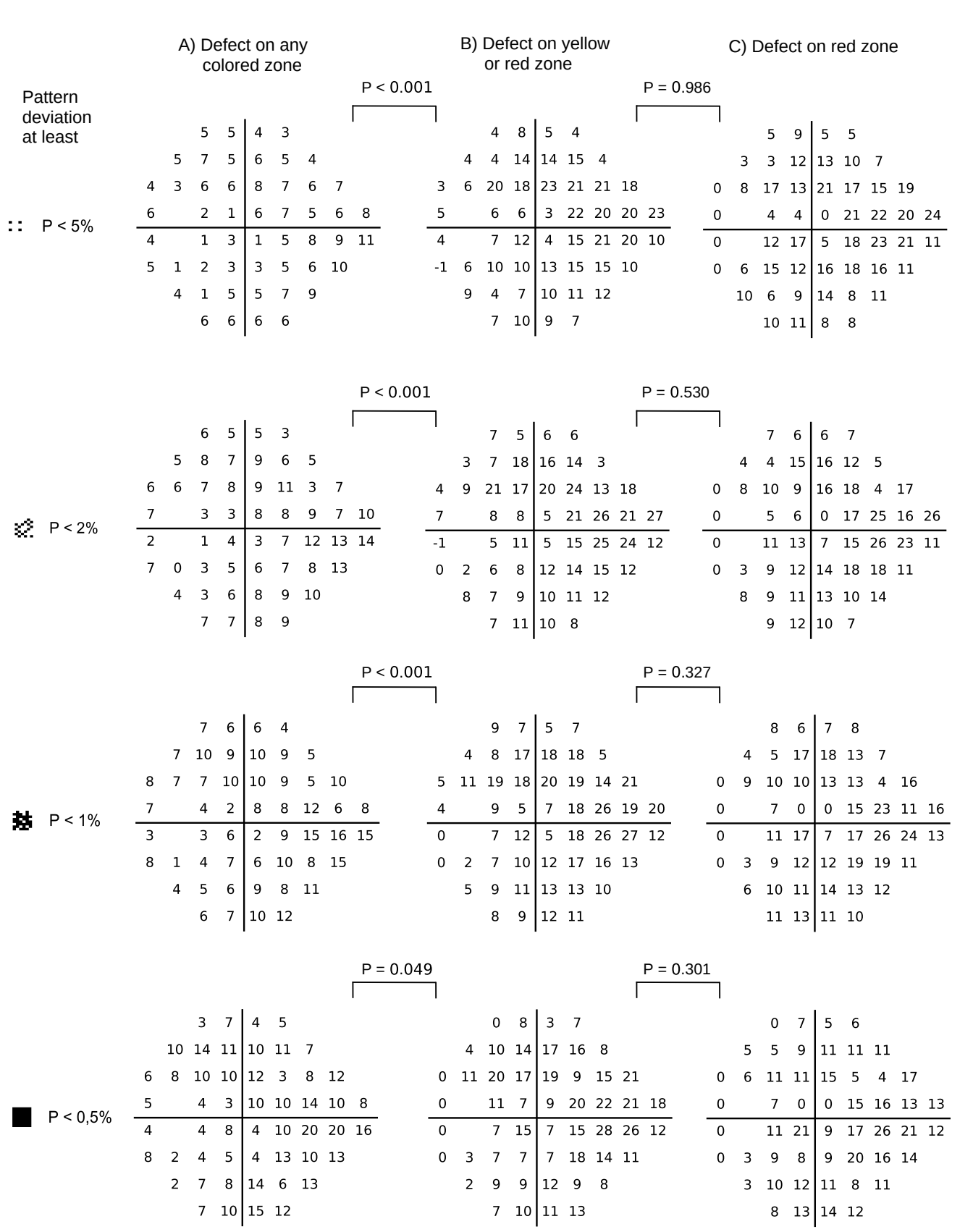
518

519 Figure3. 3



520


Northern Finland 1966 Birth Cohort Eye Study in 2012-2015 including 3070 subjects (6140 eyes)
Exclusion criteria:
- Incomplete glaucoma imaging (81 eyes)
- Missing pattern standard deviation data (95 eyes)
Final sample size 3001 subjects (5964 eyes) including 33 subjects with glaucoma (43 eyes)
A subject was included in the study if there was data from at least one eye



Supplementary Figure 2. Defect localization analysis combining visual field defective points with a pattern deviation at p<5%, <2%, <1% or <0.5% and a simultaneous OCT finding. Correlation coefficients (x100) are presented in numeric values on a 0 to 100 scale. Corresponding results are displayed on a grayscale in Figure 3. The data of the right eyes were evaluated as a mirror image to present all of the data consistently as left eyes in this analysis.

Supplementary Table 1. The Finnish evidence-based guideline for the diagnosis of glaucoma. The bolded values refer to the 'at least 2 out of 3' rule. Modified with permission from Tuulonen et al. (2015).

Glaucomatous	Normal	Diagnosis	Note
RNFL **ONH** **VF**	-	**Glaucoma**	Diagnosis conclusive
RNFL **VF**	ONH	**Glaucoma**	Apparently small disc
RNFL **ONH**	VF	**Glaucoma**	Preperimetric glaucoma
ONH **VF**	RNFL	**Glaucoma**[1]	Exclusion of neurological cause
RNFL	ONH VF	Possible glaucoma	Consider follow-up
ONH	RNFL VF	Possible glaucoma	Consider follow-up
VF	RNFL ONH	Possible glaucoma	Consider controlling SAP

[1]Rare finding in glaucoma if RNFL image is of good quality.

Supplementary Table 2. Demographics of the NFBC study population. Comparisons were performed with Mann-Whitney U-test (age, MD and PSD) and Chi-square test (gender).

	Glaucomatous	Non-glaucomatous	Total	p
Persons (eyes)	33 (43)	2968 (5921)	3001 (5964)	
Age (SD)	47.2 (0.8)	47.0 (0.9)	47.0 (0.9)	0.449
Gender: Female (%)	16 (48%)	1634 (55%)	1650 (55%)	0.451
IOP mmHg (median, range)	16 (9 to 35)	15 (7 to 24)	15 (7 to 35)	0.037
VF MD dB (median, range)	-2.0 (-22.2 to 0.9)	0.2 (-32.8 to 9.0)	0.2 (-32.8 to 9.0)	<0.001
VF PSD dB (median, range)	2.7 (1.2 to 14.1)	1.5 (0.8 to 15.9)	1.5 (0.8 to 15.9)	<0.001
RNFL thickness µm (mean, SD)				
Average (360 degrees)	78.8 (15.6)	90.9 (9.3)	90.8 (9.4)	<0.001
Inferonasal (zone 1)	90.2 (28.4)	103.2 (24.0)	103.2 (24.0)	<0.001
Inferotemporal (zone 2)	109.6 (32.5)	137.1 (19.1)	136.9 (19.3)	<0.001
Superonasal (zone 3)	80.0 (27.8)	100.9 (22.6)	100.8 (18.9)	<0.001
Superotemporal, (zone 4)	100.1 (33.8)	127.7 (18.6)	127.5 (11.7)	<0.001
Nasal (zone 5)	67.6 (13.8)	73.6 (11.7)	73.6 (11.7)	0.007
Temporal (zone 6)	63.6 (17.0)	65.1 (12.2)	65.1 (12.2)	0.559

Supplementary Table 3. The options of the observed OCT abnormalities in worsening order presented on the ROC plot (Figure 2A,B). In the different options, a given defect or worse, was detected.

Order of the step	Number of defective zones					
	1	2	3	4	5	6
1	≤ yellow					
2	red					
3	red	≤ yellow				
4	red	red				
5	red	red	≤ yellow			
6	red	red	red			
7	red	red	red	≤ yellow		
8	red	red	red	red		
9	red	red	red	red	≤ yellow	
10	red	red	red	red	red	
11	red	red	red	red	red	≤ yellow
12	red	red	red	red	red	red

Supplementary Table 4. The included visual field defects in worsening order presented on the ROC plot (Figure 2C,D). In the different options, a given defect or worse, was detected.

Order of the step	Number of defective points in the visual field										
	1	2	3	4	5	6	7	8	9	10	11
1	5										
2	5	5									
3	5	2									
4	5	2	5								
5	5	2	2								
6	5	2	1								
7	5	2	1	5							
8	5	2	1	2							
9	5	2	1	1							
10	5	2	1	0.5							
11	5	2	1	0.5	5						
12	5	2	1	0.5	2						
13	5	2	1	0.5	1						
14	5	2	1	0.5	0.5						
15	5	2	1	0.5	0.5	5					
16	5	2	1	0.5	0.5	2					
17	5	2	1	0.5	0.5	1					
18	5	2	1	0.5	0.5	0.5					
19	5	2	1	0.5	0.5	0.5	5				
20	5	2	1	0.5	0.5	0.5	2				
21	5	2	1	0.5	0.5	0.5	1				
22	5	2	1	0.5	0.5	0.5	0.5				
23	5	2	1	0.5	0.5	0.5	0.5	5			
24	5	2	1	0.5	0.5	0.5	0.5	2			
25	5	2	1	0.5	0.5	0.5	0.5	1			
26	5	2	1	0.5	0.5	0.5	0.5	0.5			
27	5	2	1	0.5	0.5	0.5	0.5	0.5	5		
28	5	2	1	0.5	0.5	0.5	0.5	0.5	2		
29	5	2	1	0.5	0.5	0.5	0.5	0.5	1		
30	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5		
31	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	5	
32	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	2	
33	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	1	
34	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	
35	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5
36	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	2
37	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
38	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Supplementary Table 5. The included structure-function combinations with an OCT abnormality and VF defects in worsening order presented on the ROC plot (Figure 2G,H). In the different options, a given defect or worse, was detected.

Order of the step	Number of defective zones[1]	Number of defective points in the visual field[2]										
		1	2	3	4	5	6	7	8	9	10	11
1	1	5										
2	1	5	5									
3	1	5	2									
4	1	5	2	5								
5	1	5	2	2								
6	1	5	2	1								
7	1	5	2	1	5							
8	1	5	2	1	2							
9	1	5	2	1	1							
10	1	5	2	1	0.5							
11	1	5	2	1	0.5	5						
12	1	5	2	1	0.5	2						
13	1	5	2	1	0.5	1						
14	1	5	2	1	0.5	0.5						
15	1	5	2	1	0.5	0.5	5					
16	1	5	2	1	0.5	0.5	2					
17	1	5	2	1	0.5	0.5	1					
18	1	5	2	1	0.5	0.5	0.5					
19	1	5	2	1	0.5	0.5	0.5	5				
20	1	5	2	1	0.5	0.5	0.5	2				
21	1	5	2	1	0.5	0.5	0.5	1				
22	1	5	2	1	0.5	0.5	0.5	0.5				
23	2	5	2	1	0.5	0.5	0.5	0.5				
24	2	5	2	1	0.5	0.5	0.5	0.5	5			
25	2	5	2	1	0.5	0.5	0.5	0.5	2			
26	2	5	2	1	0.5	0.5	0.5	0.5	1			
27	2	5	2	1	0.5	0.5	0.5	0.5	0.5			
28	2	5	2	1	0.5	0.5	0.5	0.5	0.5	5		
29	2	5	2	1	0.5	0.5	0.5	0.5	0.5	2		
30	2	5	2	1	0.5	0.5	0.5	0.5	0.5	1		
31	2	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5		
32	2	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	5	
33	2	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	2	
34	2	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	1	
35	2	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	
36	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	
37	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5
38	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5
39	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	2
40	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
41	3	5	2	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

[1]Yellow or red zones. [2] On the pattern deviation plot, each abnormal test point is depressed below this level or worse.