Optical Coherence Tomography Analysis Based Prediction of Humphrey 24-2 Visual Field Thresholds in Patients With Glaucoma

Zhihui Guo,[1] Young H. Kwon,[2,3] Kyungmoo Lee,[4] Kai Wang,[5] Andreas Wahle,[4] Wallace L. M. Alward,[2,3] John H. Fingert,[2,3] Daniel I. Bettis,[3] Chris A. Johnson,[3] Mona K. Garvin,[4,6] Milan Sonka,[3,4] and Michael D. Abràmoff[1–4,6]

[1]Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa, United States
[2]Stephen A. Wynn Institute for Vision Research, University of Iowa, Iowa City, Iowa, United States
[3]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa, United States
[4]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa, United States
[5]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, Iowa, United States
[6]Iowa City VA Health Care System, Iowa City, Iowa, United States

Correspondence: Michael D. Abràmoff, The University of Iowa, 11205 Pomerantz Family Pavilion, Iowa City, IA 52242, USA; michael-abramoff@uiowa.edu.

Submitted: March 10, 2017
Accepted: June 22, 2017

Citation: Guo Z, Kwon YH, Lee K, et al. Optical coherence tomography analysis based prediction of Humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci*. 2017;58:3975–3985. DOI:10.1167/iovs.17-21832

Purpose. A pilot study showed that prediction of individual Humphrey 24-2 visual field (HVF 24-2) sensitivity thresholds from optical coherence tomography (OCT) image analysis is possible. We evaluate performance of an improved approach as well as 3 other predictive algorithms on a new, fully independent set of glaucoma subjects.

Methods. Subjects underwent HVF 24-2 and 9-field OCT (Heidelberg Spectralis) testing. Nerve fiber (NFL), and ganglion cell and inner plexiform (GCL+IPL) layers were cosegmented and partitioned into 52 sectors matching HVF 24-2 test locations. The Wilcoxon rank sum test was applied to test correlation R, root mean square error (RMSE), and limits of agreement (LoA) between actual and predicted thresholds for four prediction models. The training data consisted of the 9-field OCT and HVF 24-2 thresholds of 111 glaucoma patients from our pilot study.

Results. We studied 112 subjects (112 eyes) with early, moderate, or advanced primary and secondary open angle glaucoma. Subjects with less than 9 scans (15/112) or insufficient quality segmentations (11/97) were excluded. Retinal ganglion cell axonal complex (RGC-AC) optimized had superior average $R = 0.74$ (95% confidence interval [CI], 0.67–0.76) and RMSE = 5.42 (95% CI, 5.1–5.7) dB, which was significantly better ($P < 0.05/3$) than the other three models: Naïve ($R = 0.49$; 95% CI, 0.44–0.54; RMSE = 7.24 dB; 95% CI, 6.6–7.8 dB), Garway-Heath ($R = 0.66$; 95% CI, 0.60–0.68; RMSE = 6.07 dB; 95% CI, 5.7–6.5 dB), and Donut ($R = 0.67$; 95% CI, 0.61–0.69; RMSE = 6.08 dB, 95% CI, 5.8–6.4 dB).

Conclusions. The proposed RGC-AC optimized predictive algorithm based on 9-field OCT image analysis and the RGC-AC concept is superior to previous methods and its performance is close to the reproducibility of HVF 24-2.

Keywords: perimetry, visual field, image analysis, OCT, ganglion cell

The limitations of reliability and reproducibility of visual field (VF) testing, as the main parameter in assessing glaucoma damage, inhibit optimal patient care and research for improved outcome. The clinical standard for VF testing in glaucoma is automated perimetry, and the Humphrey 24-2 SITA Standard VF (HVF 24-2) is the most widely used method. However, once moderate VF loss occurs, 12 to 15 dB mean deviation (MD) loss or more, VF test–retest variability rises substantially[1–4] and limits a reliable determination of change.

Optical coherence tomography (OCT) can quantify glaucomatous damage through nerve fiber layer (NFL) thickness and cup-to-disc estimates in a patient-friendly, robust, and reproducible fashion.[5–7] However, OCT-derived measurements of glaucoma damage correlate poorly with VF thresholds.[8–11] We have shown previously that OCT-based image analysis in glaucoma patients allows loss to be quantified from the retinal ganglion cell body to the optic nerve head (ONH),[12] suggesting that damage to the retinal ganglion cell–axonal complex occurs simultaneously along this entire path. We proposed the term retinal ganglion cell axonal complex (RGC-AC) to stress the distributed nature of this loss in multiple neighboring ganglion cells and corresponding axons, which leads to characteristic glaucomatous VF loss, while damage to the part of the RGC-AC within the ONH leads to characteristic cupping.

We also demonstrated that retinal anatomy-based analysis of multifield spectral domain (SD)–OCT can predict the thresholds at all 52 test locations of the most widely used Humphrey 24-2 VF with an average correlation of 0.68,[13] in a leave-one-out test design. In a pilot study, average correlation between repeat Humphrey 24-2 thresholds in established glaucoma patients proficient in perimetry is 0.83 (Abramoff MD. *IOVS* 2015;56:ARVO E-abstract 1696), which allows additional improvements in

the image analysis and predictive algorithm up to that performance limit to be measured. Improvements in the predictive algorithm make it of interest to compare these to our previously published approach, as well as the widely used peripapillary NFL thickness assessment. In addition, an independent test population of patients with a wide range of glaucoma severity allows a better assessment of external validity.

Thus, the purpose of this study was to demonstrate the performance improvement by more sophisticated OCT image analysis and validate this performance on an independent test set of glaucoma subjects.

Methods

Subjects

In this prospective study, inclusion criteria were as reported previously and summarized here[13]: age 18 to 85, diagnosed with glaucoma suspect or open-angle glaucoma according to the following definitions: (1) Glaucoma suspect: suspicious optic nerve appearance (enlarged cupping on clinical examination) with normal VF and IOP $\leq$ 21 mm Hg or normal optic disc appearance on biomicroscopy and normal VF, but with elevated IOP $>$ 21 mm Hg. (2) Open-angle glaucoma: primary or secondary open-angle glaucoma (e.g., exfoliative or pigmentary) with an open iridocorneal angle, glaucomatous optic disc, and/or NFL defects on biomicroscopy, and VF changes (regardless of IOP level). Glaucomatous optic discs were identified as those with either diffuse or focal thinning of the neuroretinal rim. VF abnormalities were considered to be glaucomatous if they were consistent with the optic nerve examination and had either (1) a typical NFL distribution, or (2) a glaucoma hemifield test outside the normal limits. This diagnosis was made by fellowship trained glaucoma specialists according to the above definitions.

Inclusion criteria were subject's fundus visualization sufficient on indirect ophthalmoscopy to allow OCT; able to undergo perimetry Humphrey 24-2 VF SITA Standard with sufficient reliability (false-positive [FP] error $<$ 15% and false-negative [FN] error $<$ 25%, fixation loss [FL] $<$ 33%); perimetry obtained within a 3-month period of SD-OCT imaging; and perimetry free of artifacts, such as lens rim effects. Exclusion criteria were a history of angle closure or combined mechanism glaucoma, or any nonglaucomatous optic neuropathy, corneal, or retinal diseases that can affect VF, cataracts or any other disease with visual acuity $<$ 20/40, and OCT of unsuitable quality determined by visual observation. Subjects were recruited matching age and disease severity in one of three approximately equally sized severity groups, based on the mean deviation of the 24-2 HVF threshold testing: early glaucoma (including glaucoma suspects) $<$ 6 dB loss, moderate glaucoma 6 to 12 dB loss, and advanced glaucoma $>$ 12 dB loss.

One eye of each subject was studied. When both eyes were eligible, the study eye was chosen to reflect adequate representation of each of the three severity groups.

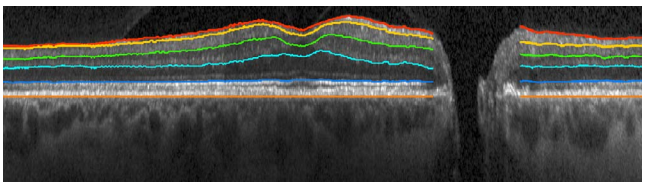

Figure 1. Wide field composite OCT obtained from 9-field OCTs after coregistration, showing the central B-scan overlaid with cosegmented surfaces for a subject with advanced glaucoma. NFL and GCL+IPL thicknesses were not measured within the ONH region.

Data Collection

As stated previously,[13] standardized automated perimetry based on the SITA Standard 24-2 VF protocol was performed with the Humphrey Field Analyzer (HFA II; Carl Zeiss Meditec, Inc., Dublin, CA, USA), which evaluates the VF as threshold assessments at 52 different retinal locations (the two locations corresponding to the blind spot were subtracted from the total of 54 locations). For OCT image acquisition, a 9-field per eye protocol was used, where a subject sequentially fixates on a spot 12.5° apart in a 3 $\times$ 3 grid pattern. This protocol takes approximately 5 minutes per eye and covers 60° on the retina, sufficiently large enough to include the 60° area probed with 24-2 VF test. Each field is imaged with SD-OCT (768 $\times$ 61 $\times$ 496 voxels, 9.53 $\times$ 8.07 $\times$ 1.92 mm^3, with a voxel size of 12.41 $\times$ 132.22 $\times$ 3.87 μm^3; Spectralis; Heidelberg Engineering, Inc., Heidelberg, Germany) using eye tracking mode. The device additionally acquires a 2D scanning laser ophthalmoscopy (SLO) fundus image (768 $\times$ 768 pixels, 9.5 $\times$ 9.5 mm^2 with a pixel size of 12.41 $\times$ 12.41 μm^2), automatically coregistered with the OCT image by the device. The raw VF data, exported from the Humphrey Field Analyzer as integer threshold data, as well as the raw OCT volumes, exported as .vol format, were de-identified, and stored in our XNAT ophthalmology research database.[14] The study protocol was approved by the institutional review board of the University of Iowa and adhered to the tenets of the Declaration of Helsinki; written informed

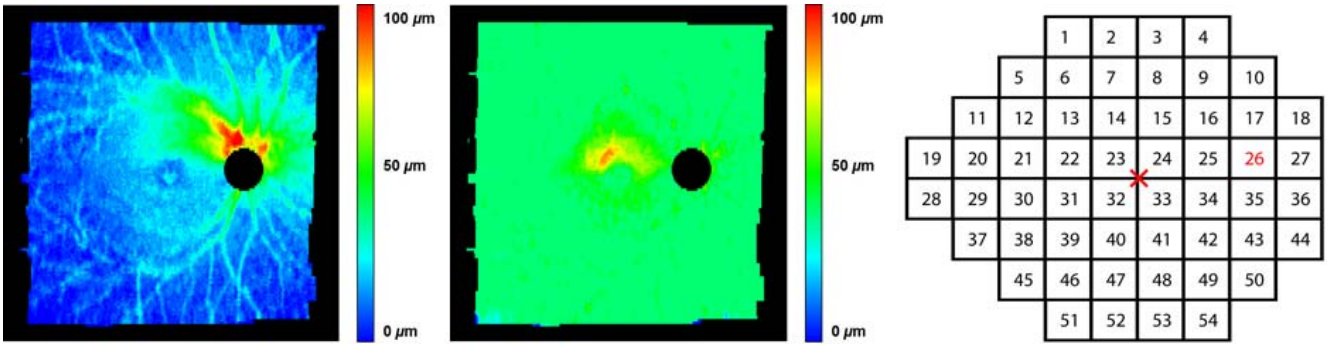

Figure 2. Wide field composite OCT cosegmented NFL thickness (*left*) and GCL+IPL thickness (*middle*) of the same subject with advanced glaucoma, as in Figure 1. The HVF 24-2 derived S-Grid (*right*) illustrates the numbered sectors for all test locations, the *cross marker* indicates the fixation center, and *sector 26* the blind spot. This S-grid is aligned with the wide field OCT and used to identify the OCT sectors for which NFL and GCL+IPL thickness are calculated.

Investigative Ophthalmology & Visual Science

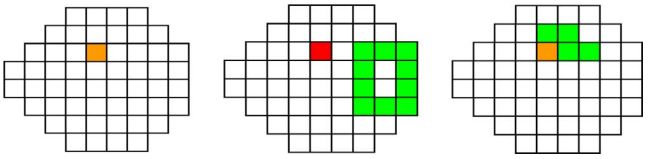

Figure 3. Sectors used for Naive, Donut, and Garway-Heath predictive models, to predict the HVF 24-2 threshold for Sector 14. The predicted sector 14 is shown in *red*, the sectors of which NFL thickness is used are in *green*. If a sector's NFL and GCL+IPL are both used, it is shown in *orange*.

consent was obtained from all participants, and HIPAA compliance was adhered to.

Multifield Registration and Intraretinal Layer Segmentation

We previously have described how the 9-field OCT volumes are registered, and regional NFL and ganglion cell layer (GCL) thickness is quantified.[13] In summary, all 9 individual SLO images were registered automatically, and the resulting affine (i.e., only including scaling, rotation, and translation) transformations then were applied to the corresponding OCT volumes, so that their relative positions, scales, and rotations were known. The retinal layers (NFL, and GCL and inner plexiform layer [GCL+IPL]) of all OCT volumes thus aligned, were cosegmented, taking into account the possible mutual displacements along the *z*-axis, using an extension of the Iowa Reference Algorithms.[15–17] After cosegmentation, the volumes and segmented surfaces were stitched together to obtain a wide-field composite OCT (Fig. 1) and the corresponding layer

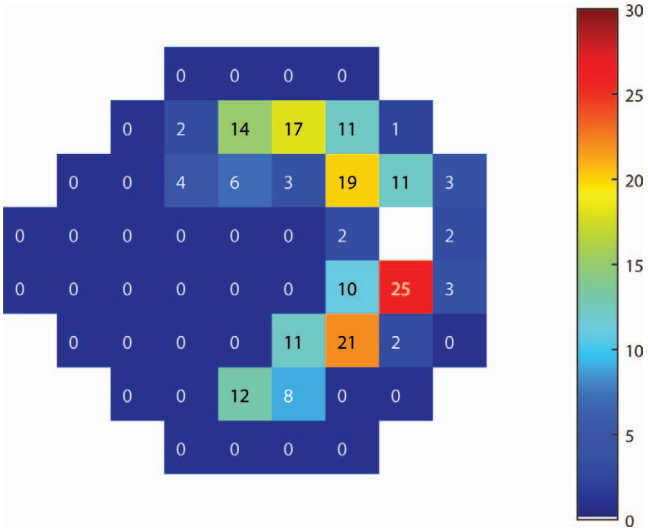

Figure 5. Map of the most relevant sectors' NFL thickness for prediction of function using RGC-AC Optimized. The *number* in each sector indicates the total number of sector-specific predictions that sector's NFL thickness is used for. In other words, if a sector S has "19," there are 19 sectors that have sector S's NFL thickness in their feature vector.

thicknesses (Fig. 2). As we did previously, we partitioned the wide field composite OCT into 54 sectors that corresponded to the HVF 24-2 SAP matrix, called Structure-Grid (S-Grid) where the automatically identified fovea and ONH center were coregistered to the fixation and the center of sector 26,

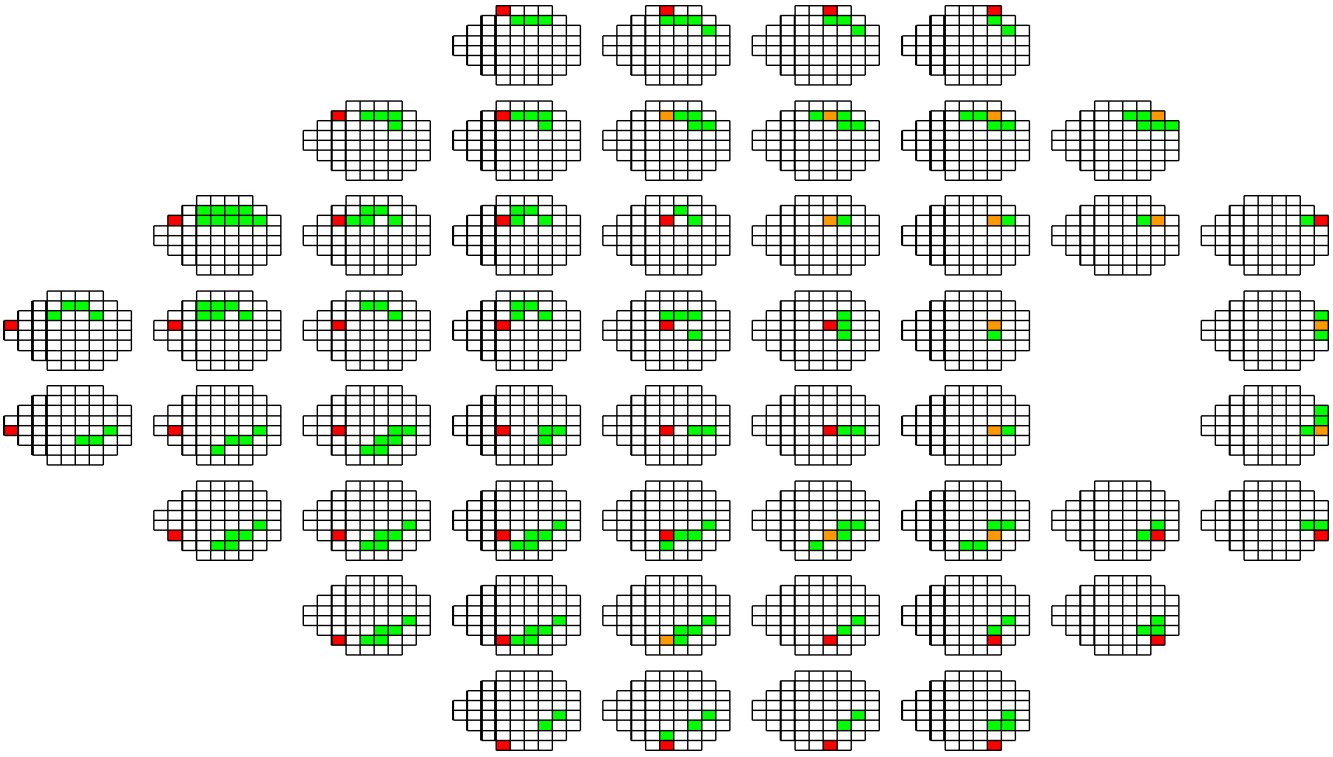

Figure 4. All sectors included in each sector-specific feature vector that is used to predict the HVF 24-2 threshold for that sector using RGC-AC optimized. The predicted sector is *red*—this also is the sector whose GCL+IPL thickness is used. Sectors whose NFL thickness is used are *green*. If NFL and GCL+IPL thicknesses are used for a sector, that sector is *orange*. Notice that in some cases, prediction performance was improved by including NFL sectors temporal of the sector to be predicted—this does not imply that the ganglion cell axons in that sector originate nasally, rather that damage to the axons in the more temporal sector covaries substantially with that of the predicted sector.

Investigative Ophthalmology & Visual Science

Table 1. Averaged *R*, RMSE, and Width of LoA of Four Predictive Approaches for All Glaucoma Severity Groups

Performance	RGC-AC Optimized	Naïve	Garway-Heath	Donut
Average *R*	0.74	0.49	0.66	0.67
Range of *R*	[0.50, 0.85]	[0.08, 0.74]	[0.40, 0.83]	[0.48, 0.75]
P Value for difference in *R* to RGC-AC optimized	–	<0.001	<0.001	<0.001
Average RMSE, dB	5.42	7.24	6.07	6.08
P Value for difference in RMSE to RGC-AC optimized	–	<0.001	0.010	0.001
Average width of LoA, dB	21.2	28.2	23.8	23.8
P Value for difference in width of LoA to RGC-AC optimized	–	<0.001	0.009	<0.001

respectively.[13,18,19]) Thus, essentially all A-scans in the wide field composite OCT were assigned to a corresponding S-Grid sector, and the average GCL+IPL and NFL thickness values were computed as the mean layer thickness from all A-scans in that sector, from a total of approximately 2000 A-scans per sector. Any missing thickness information of a sector was bilinearly interpolated from the four neighboring sectors.

Prediction of VF Threshold for Each Sector From NFL and GCL+IPL Thicknesses

As we did previously,[13] we built independent predictive models for each sector threshold, except for sectors 26 and 35, which covered the ONH area. These models only used the NFL and GCL+IPL thicknesses for one or more (structural) sectors, so no functional information was used as input to the model. To study the effect of the contribution of a structural sector, that is, the contribution of a sector's regional NFL and GCL+IPL thickness, four models were compared for each sector threshold prediction, using four different approaches for model inputs, with their descriptors as follows:

1. Naive: GCL+IPL and NFL thickness for the predicted sector only.
2. Donut: NFL thickness of 10 sectors that form a donut centered on the ONH. This approach most closely approximates the use of peripapillary NFL thickness

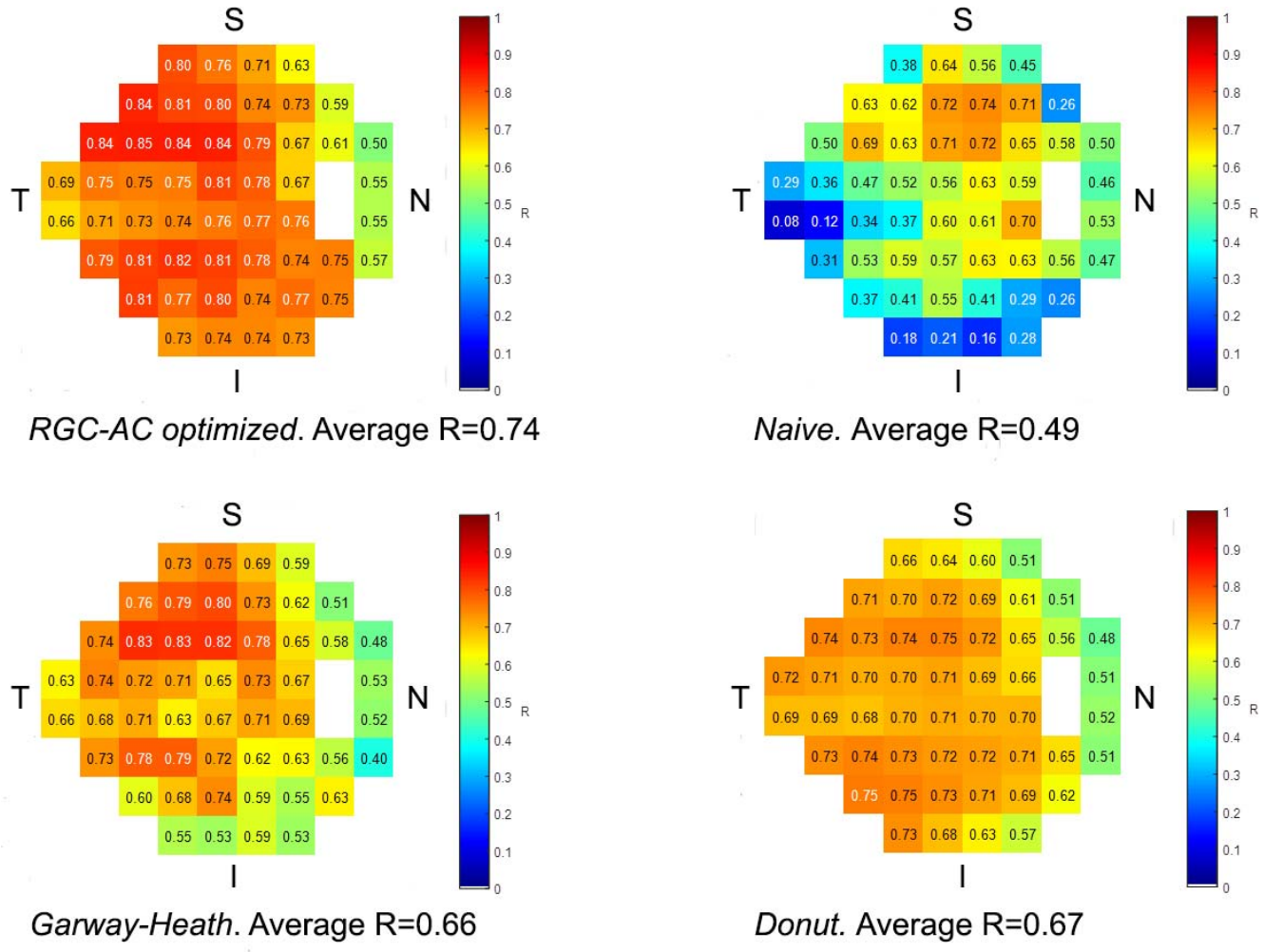

Figure 6. Sector-specific predictive performance *R* for each approach as well as average *R* over all sectors. The differences between *R* for RGC-AC optimized and the other approaches are all significant.

Investigative Ophthalmology & Visual Science

Table 2. The Comparison of the Bias Between the *RGC-AC* Optimized Model and the Other Three Models Across 52 Sectors

	RGC-AC Optimized		Naïve		Garway-Heath		Donut	
Sensitivity, dB	Average Bias, dB	*P* Value to RGC-AC Optimized	Average Bias, dB	*P* Value to RGC-AC Optimized	Average Bias, dB	*P* Value to RGC-AC Optimized	Average Bias, dB	*P* Value to RGC-AC Optimized
10	2.8	–	8.3	<0.001	3.8	0.15	2.3	0.15
20	−0.01	–	1.0	<0.001	0.3	0.39	−0.4	0.12
30	−2.8	–	−6.3	<0.001	−3.4	0.17	−3.1	0.13

measurements that is widely available in, for example, the Zeiss Cirrus SD-OCT.[20]

3. Garway-Heath: GCL+IPL thickness for the predicted sector, as well as all NFL sectors that fall within the Garway-Heath nerve fiber bundle distribution.[21] This is the approach we have reported previously[13] (see Fig. 3 for additional insight).
4. RGC-AC optimized: GCL+IPL thickness for the predicted sector, as well as NFL thicknesses for a set of between 1 and 10 sectors following the so-called RGC-AC optimized regional path. These sets are optimized for performance on the training set and approximate the nerve fiber bundles as much as possible, by iteratively adding a sector using sequential floating forward sector search,[22] if the marginal improvement in correlation $\Delta R > 0.01$ (Figs. 4, 5).

All sectors of which NFL and/or GCL+IPL are used to predict the HVF 24-2 threshold for sector 14 (as an example) for RGC-AC optimized are shown in Figure 4, and examples of the sectors used for a single sector prediction for the other three approaches are shown in Figure 3.

Each of these four approaches creates a feature vector that then is used to train the predictive model, implemented as a support vector regression machine (SVM)[23] with a radial basis function kernel. As previously, to account for the slight rotation between the OCT imaging and actual perimetry, the S-Grid and the 24-2 VF grid are aligned by similarity transform between two pairs of points, the center of the fovea on OCT and the fixation center and the ONH center and blind spot center. Then, a thin plate spline transform is used to interpolate the actual measured thresholds on the S-Grid. Obviously, this interpolation is required only when the predicted sensitivity threshold must be associated with the measured sensitivity in a specific location. Each sector-specific SVM then is trained using the sector NFL and GCL+IPL thicknesses in the corresponding feature vector as described above, as well as using as the reference standard the corresponding interpolated VF thresholds for the sector.

For training the predictive algorithms we used the data collected for our previous study,[13] as follows. We had collected HVF 24-2 and 9-field OCT with exactly the same protocol and inclusion and exclusion criteria on 142 subjects from the Glaucoma Service at the University of Iowa. Among these 142 subjects, 20 had incomplete imaging, 4 had no composite OCT volume and 7 had layer segmentation failure, and thus the data of 111 subjects could be used. A random eye from each of the remaining 111 subjects was selected as the training set (111 eyes, 999 scans), of which 59 were right and 52 were left eyes; 39 had early, 36 moderate, and 36 advanced glaucoma. Thus, there were 111 (subjects) × 52 (sectors) training vectors used for training the four approaches.

All other parameters for the SVM were the same for all sectors, and no other training data were used. Once trained, each sector's predictive model, given a corresponding previously unseen feature vector, produces a predicted threshold at that sector. To make comparison to the familiar HVF 24-2 printout easier, we simulated our result output as a grayscale map.

Statistical Analysis

All left eye scans were mirrored to conform to the scans of the right eye. Primary outcome was the performance improvement, measured by Pearson correlation *R* between the actual and predicted HVF 24-2 thresholds, averaged over all 52 sectors, of the RGC-AC optimized approach over the other three approaches, on the independent test set of newly recruited subjects with glaucoma. Average *R* and root mean square error (RMSE) were calculated by averaging the Pearson correlation coefficient (*R*) and RMSE between predicted and actual thresholds for all subjects for each sector, and hypotheses were tested using the Wilcoxon rank sum test. Bland-Altman plots[24] were used to show the agreement between the predicted and measured HVF 24-2 thresholds, with the *y*-axis representing the difference and the *x*-axis the mean of predicted and measured HVF 24-2 thresholds. The 95% limits of agreement (LoA) were computed for each sector.[25] The average width of LoA was obtained by averaging the width of LoA for each sector and testing between the different models with the Wilcoxon rank sum test. For stratified analysis, a linear regression and its 95% confidence limits (CI)[26] were calculated on the Bland-Altman plot for each sector. At 10, 20, and 30 dB, we applied the Wilcoxon test to 2 categories, the prediction error on the regression line (called bias) to test the bias from 0, and the width of the 95% CI to test the range of agreements for the prediction error. These also

Table 3. The Comparison of the Width of 95% CI Between the RGC-AC Optimized Model and the Other Three Models Across 52 Sectors

	RGC-AC Optimized		Naïve		Garway-Heath		Donut	
Sensitivity, dB	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized
10	4.1	–	5.3	0.001	4.7	0.002	4.7	0.001
20	2.4	–	2.9	<0.001	2.7	0.005	2.8	<0.001
30	3.7	–	5.0	0.003	4.3	0.040	4.4	0.001

Table 4. The Comparison of the Bias and the Width of 95% CI Between RGC-AC Optimized Model and Repeat HVF at Different Sensitivities Across 52 Sectors

	RGC-AC Optimized		Repeat HVF		RGC-AC Optimized		Repeat HVF	
Sensitivity, dB	Average Bias, dB	*P* Value to RGC-AC Optimized	Average Bias, dB	*P* Value to RGC-AC Optimized	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized	Average Width of 95% CI, dB	*P* Value to RGC-AC Optimized
10	2.8	–	−0.39	<0.001	4.7	–	3.2	<0.001
20	−0.01	–	0.04	0.9	2.7	–	1.7	<0.001
30	−2.8	–	0.5	<0.001	4.3	–	2.5	<0.001

were calculated for the repeat HVF 24-2. Significance was set at the 0.05/3 level, with Bonferroni correction. For qualitative evaluation, grayscale maps of the actual and predicted HVF 24-2 were created.

Results

We recruited 112 new consecutive participants from the Glaucoma Service at the University of Iowa. Of these 112 participants, 15 were excluded due to incomplete imaging, and 11/97 (11%) subjects were excluded because of complete layer segmentation failure (i.e., undetectable NFL/GCL layer and/or shift of NFL lower boundary to GCL lower boundary), leaving 86 subjects for the study (we used 48 right and 38 left eyes). Thus, approximately 15% of the total number of potential subjects had to be excluded because of segmentation failures. One eye from each of the remaining subjects formed the independent test set to evaluate the performance of the four models. Thus, an independent test set was collected. Demographics of the 86 subjects were mean age, 65.3 years and 38 (44.2%) were male. A total of 30 patients had early, 24 moderate, and 32 advanced glaucoma. The cohort included 79 (self-identified) Caucasian, 3 African-American, 1 Asian-American, 0 Native American, and 1 Native Hawaiian/Pacific Islander subjects (race was unknown or undisclosed for remaining 2). None of the 79 Caucasian and one of the seven non-Caucasian subjects identified as Hispanic.

The average Pearson correlation *R* (RMSE) between the interpolated, actual HFA 24-2 and OCT-predicted VF thresholds for RGC-AC optimized model was 0.74 (5.42 dB), averaged over all severity groups, and this correlation was significantly higher than the average *R* achieved by the other three approaches (Naïve, Donut, and Garway-Heath) on this dataset (Table 1). Figure 6 shows the correlation *R* for each sector, and for each approach, and RGC-AC optimized has significantly higher performance than all other models.

From Tables 2 and 3, we concluded that the RGC-AC optimized model was significantly better than the other three over the entire range. Table 4 shows that repeat HVF in a highly select sample of good test-takers still performs better than the RGC-AC optimized model.

For qualitative evaluation, the grayscale maps that simulate the HVF 24-2 printout were generated for actual and RGC-AC optimized predicted thresholds, grouped by glaucoma severity, as shown in Figure 7, while a detailed comparison for a specific subject with advanced glaucoma is shown in Figure 8. Figure 9 compares the predictive performance of all 86 subjects for the four models for a single HVF 24-2 sector (sector 11), which shows the higher performance of the RGC-AC optimized model in the entire range of threshold values. For comparison purposes, Figure 10 shows the predictive performance of HVF 24-2 sensitivity across all sectors for RGC-AC optimized using the same box-whisker plot as used by Zhu et al.,[27] showing superior predictive performance of the RGC-AC optimized approach compared to their approach at thresholds less than 20 dB. Average signed error from measured thresholds was −0.05 dB, and average unsigned error was 4.19 dB.[28]

If we included the 11 participants who had inadequate NFL and GCL+IPL segmentation (Fig. 11), and, thus, did not use the inclusion criteria, the RGC-AC Optimized prediction reached a lower average correlation *R* over all 97 subjects of 0.66 (95% CI, 0.63–0.69). This clearly showed the importance of accurate layer segmentation as well as the influence of correct NFL and GCL+IPL layer thickness values on the prediction outcome.

Discussion

Our results showed the high predictive performance of the RGC-AC optimized approach to predict VF thresholds from 9-field OCT image analysis, with an average correlation *R* of 0.74 to the actual HVF 24-2 thresholds. This performance was achieved on a newly recruited, independent population of glaucoma subjects with a wide distribution of glaucoma severity, using the RGC-AC optimized approach trained on a separate training set. Thus, no OCT images or HVF 24-2 thresholds of the subjects recruited for this study were ever used to train the machine learning prediction algorithms. In addition, the performance of the newly developed RGC-AC optimized approach was significantly better than three alternative approaches: Naïve, where only the NFL, GCL, and IPL thickness of the predicted sector are used for prediction; Garway-Heath, our previously published, and so far best, approach using GCL+IPL thickness of the predicted sector as well as NFL thicknesses of the sectors in the presumed nerve fiber bundle paths as determined by Garway Heath et al.[13]; and Donut, the approach that uses peripapillary NFL thickness only to mimic as close as possible this metric that is widely available on commercially available OCT devices,[20] even though Donut is a 2-D ring incorporating many more A-scans than the commercially available 1-D "peripapillary circle."

Based on our results, we made several observations: While present, there is only a limited amount of "plateauing," or leveling off of predictive performance, at increasing severity of glaucoma. Compared to the study of Zhu et al.,[27] there is less of a plateau. It is possible that the predicted thresholds are closer to the "true" sensitivity below 20 dB, since studies have questioned the reliability and accuracy of perimetry for such low sensitivities.[29] In any case, we expect to continue improving prediction performance and decreasing the plateau in future studies.

Contrary to our expectations, the performance improvement of the RGC-AC optimized approach over the peripapillary Donut approach is seen across the entire range of glaucoma severity (Tables 2, 3), and as illustrated for a single sector in the scatter plot in Figure 9. This is somewhat surprising as several

Investigative Ophthalmology & Visual Science

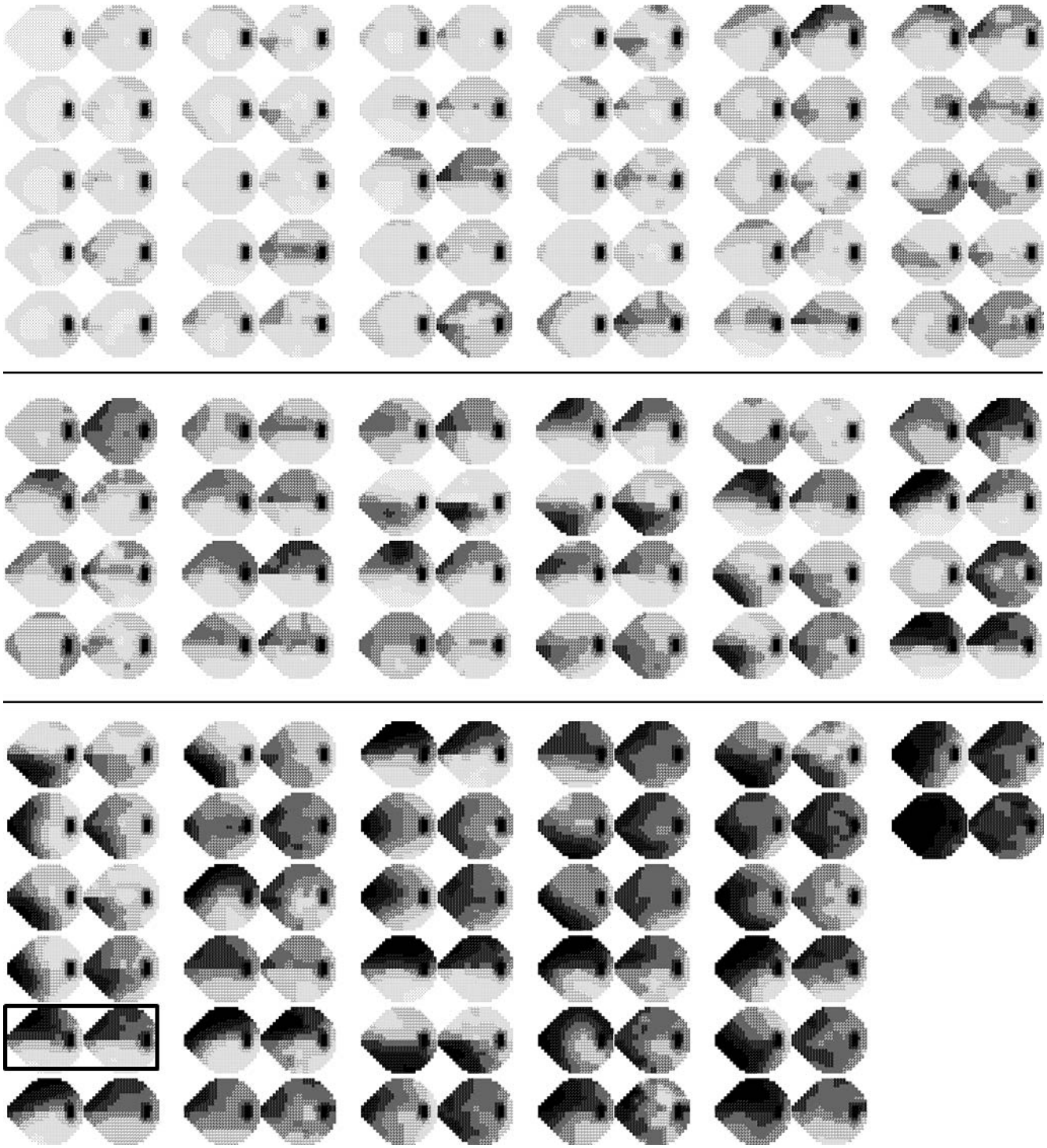

Figure 7. HFA 24-2 simulated printout of actual (*left*) and RGC-AC optimized predicted (*right*) thresholds for each subject grouped by glaucoma severity (early, *top*; moderate, *center*; and advanced, *bottom*, respectively). The *box* is placed around the subject that is shown in Figures 1, 2, and 8.

studies have shown a saturation effect for structure–function correlation at advanced glaucoma.[5,8,9,30,31] The difference here is that we averaged the correlation values from each of the 52 sectors to calculate a final average *R* value for each eye, as opposed to correlating mean NFL thickness with global HVF MD, that is, using the average of only 2 numbers from each eye. As pointed out, Donut incorporates far more A-scans than the 1-D circle that is used commercially. It is possible that these differences account for at least part of the much more robust correlation between structure and function across all severity levels.

Substantial predictive performance can be obtained assuming that actual VF threshold sensitivity values are related directly to layer thicknesses, as a proxy of the number of axons or ganglion cells as measured with OCT. Our premise, thus, remains that HVF 24-2 threshold is related only to the number of RGC and their axons (i.e., the amount of RGC-AC remaining), and, in fact, we and others have shown age-related

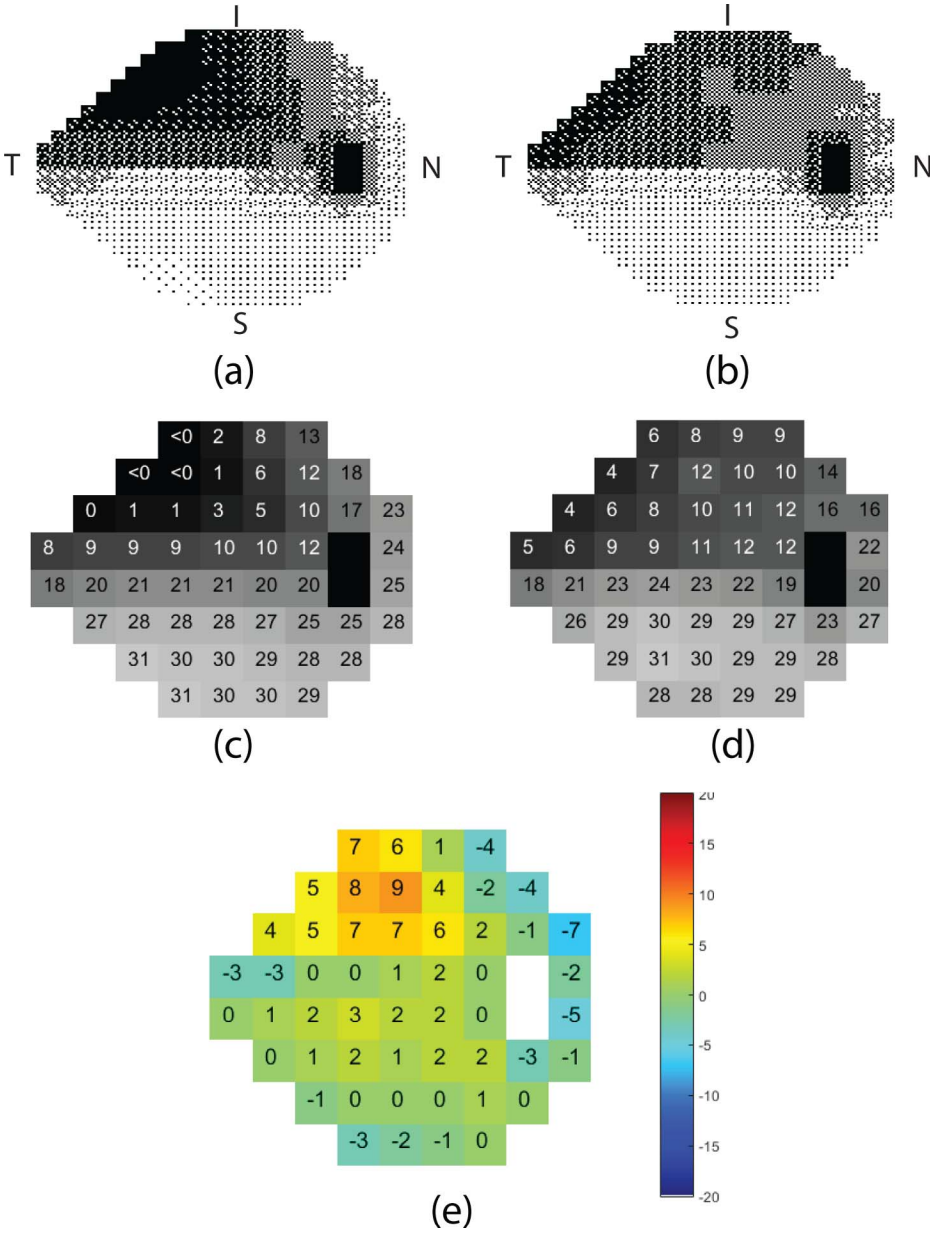

Figure 8. Detailed simulated printout of the actual HVF 24-2 (*left column*) and RGC-AC predicted HVF 24-2 (*right column*) thresholds, in a subject with advanced glaucoma, the same as in Figure 1, 2, and 7. *Top row*, dithered grayscale mapping for (**a**) actual, (**b**) predicted. *Middle row*, sensitivity thresholds in dB for (**c**) actual, (**d**) predicted. *Bottom row*, (**e**) differences between actual and predicted threshold per sector.

loss of RGC-AC in normal subjects.[32,33] However, this has not been confirmed and there also may be age-related changes in the neuroretina affecting threshold sensitivity.

Significantly better performance was reached by the RGC-AC optimized predictive model. This uses RGC-AC bundle paths, which are similar, but not identical, to the functionally derived bundle paths that originally were described by Garway-Heath et al.[21] As can be seen in Figure 4, the RGC-AC optimized bundle paths, and thereby the underlying ganglion cell axons, display substantial overlap. In some sectors, RGC-AC optimized paths include NFL sectors temporal of the sector to be predicted, which is unlikely to be anatomically correct. Rather, this is caused by the loss of axons in the more temporal sector covarying substantially with the loss in the predicted sector. Refinement of the paths by making the structural sectors smaller, rather than being determined by the spacing between HVF-24-2 test locations, may further elucidate the population-derived RGC-AC bundle paths.

Accurate automated segmentation of the NFL and GCL+IPL layers is a prerequisite to achieving the reported prediction performance. Even though our NLF and GCL+IPL segmentation has matured, we excluded 11 participants (approximately 11%), because their segmentations clearly were insufficient. If we included these as subjects, predictive performance was substantially lower. Though we have developed automated segmentation quality methods,[34] segmentation performance clearly must be improved[35] before our approach can be considered for use on glaucoma patients.

We also replicated our second initial finding, that the correlation between structure and function is higher in the superior than in the inferior retina. The correlations of the RGC-AC optimized model over the entire field thus calculated for the superior hemifield and the inferior hemifield were 0.82,

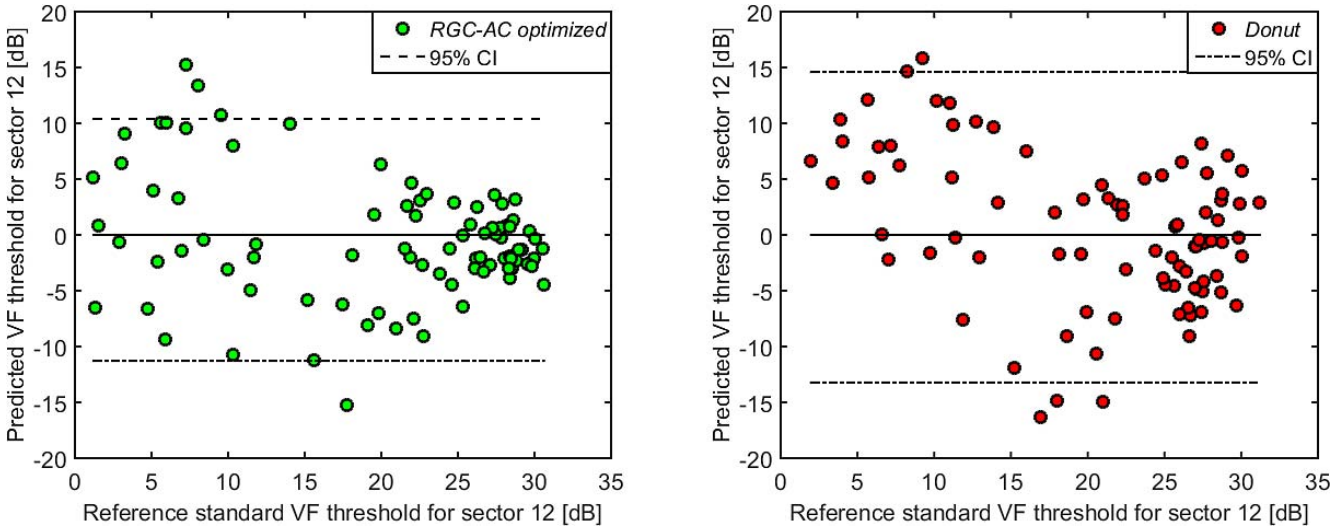

Figure 9. Bland-Altman plots of prediction error (*vertical axis*) and mean of the predicted and measured VF threshold for sector 11 of all 86 subjects for four models. The RGC-AC optimized has better predictive performance over the entire range of glaucoma severities.

0.83, and 0.85, respectively. This again raises the question whether this is the consequence of an evolutionary adaptation to the inferior hemifield being more important for survival in primates.

In addition to the requirement for better segmentation, there are some additional issues: though patients with glaucoma anecdotally prefer 9-field OCT to HVF 24-2, it still can be cumbersome. Hopefully, additional improvements in wide field swept source OCT will allow the entire 60° of the posterior pole to be imaged with fewer fields. Currently, coregistration, cosegmentation, and prediction take approximately 15 minutes, and use in a busy glaucoma clinic may require faster processing times, which are achievable by using Graphics Processing Units.

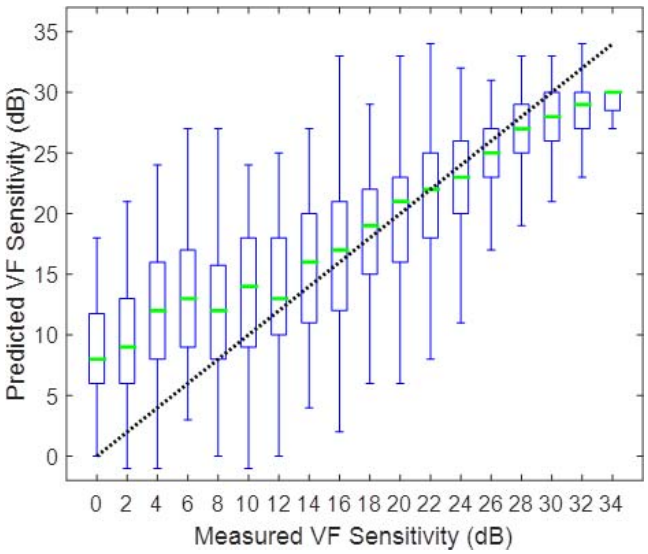

Figure 10. Relationship between RGC-AC optimized predicted and measured sensitivity for each HVF 24-2 test location in 86 eyes, stratified by sensitivity. The *bar* summarizes the predictive performance over a 2-dB range from 0 to >36 dB. *Thin vertical lines* are 90% prediction limits (5th and 95th percentile of error), each *box* indicates the interquartile range of the prediction error (25th and 75th percentile of error) with the *line* in the *box* indicating the median error. The *dotted line* of unity indicates perfect prediction. Compared to Zhu et al.,[27] the error range for RGC-AC optimized is smaller at thresholds less than 20 dB.

The average correlation R of predicted to actual sensitivity thresholds currently is close to that obtained by repeat HVF 24-2. This is caused by the substantial intrasubject variability of 24-2 perimetry [31,36–38] when estimating the "true" threshold, the implication is that potential improvements in prediction become harder and harder to measure—the likelihood that the predicted and actual threshold differ because the prediction was incorrect is more and more similar to the difference being caused by noise in the actual measured threshold. In other words, a more noisy reference standard necessarily negatively affects the actual performance that can be measured.[39] This is a challenge because one alternative, repeat VFs, is patient unfriendly, and in our experience makes it hard to recruit a sufficient number of subjects. The repeat HVF 24-2s were obtained in a number of extremely good and motivated test takers, and, thus, this sample may represent the maximum obtainable repeatability of HVF 24-2. Unfortunately, no studies of the repeatability of individual test points in HVF 24-2 currently are available to us, so until we have finalized a formal study this is the only data point we have – as mentioned, we believe $r = 0.83$ does not adequately represent the HVF 24-2 variability in glaucoma patients. Potentially, adding pattern electroretinogram (PERG) or frequency doubling technology perimetry may allow the reference standard to be improved.[38,40]

Much effort has been devoted to studying the structural–functional (S-F) correlation in glaucoma. Studies have shown a curvilinear relationship between peripapillary retinal NFL (PP-NFL) thickness and global VF index, such as mean deviation.[10] Below approximately −10 dB loss, the PP-NFL thickness

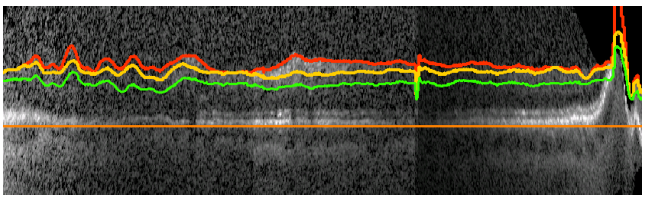

Figure 11. An example wide field OCT B-scan of a participant that was excluded from analysis because there is inadequate NFL and GCL+IPL segmentation.

reaches a "floor," thus, limiting its dynamic range. A similar curvilinear relationship was found between macular GCL thickness and VF with a dynamic range limited to VF better than −10 dB.[9] Given the nonlinear correlation between OCT and VF correlation thus established, some have advocated a combined structure–function index (CSFI) to estimate the number of retinal ganglion cells in the retina to improve early detection of glaucoma.[33] Others have advocated careful examination of 2 joint OCT scans, that is, ONH NFL and macular GCL thicknesses, for localized correlation with VF thresholds, paying particular attention to the so-called "macular vulnerability zone.[41,42] More recently, Hood et al.[43] has shown how single wide-field (9 × 12 mm) swept source OCT, thus encompassing ONH and the macula, can be used to detect early glaucoma. Our results confirmed that wide-field OCT spanning the optic nerve and macula provides a more robust and accurate picture of the retinal structure and should improve correlation with visual function. In fact, our coregistered 9-field OCT covers approximately a 60° view of the posterior pole, matching the same area as the HVF 24-2 grid. This allows us to estimate the structure–function correlations, and thereby sensitivity thresholds, for all HVF 24-2 test locations, at substantial higher correlation than the circumpapillary circle scan and global VF indices. Finally, our results demonstrated how more sophisticated models for integrating OCT structural information, what we call RGC-AC optimized, further improved the correlation to an average 0.74, and even higher in some individual, clinically relevant test locations.

In summary, our results showed a high predictive performance of individual VF thresholds predicted from OCT image analysis using the RGC-AC concept and the RGC-AC optimized approach, with an average correlation *R* of 0.74 to actual HFA 24-2 perimetry. This performance was obtained on a newly recruited, independent population of glaucoma subjects with a wide distribution of glaucoma severity. We believe we have laid the foundation to predict visual function accurately based on OCT structural information, using more of available information and smart algorithms. Potentially, predicted function derived objectively from OCT structure in patients with glaucoma may complement subjective VF testing in clinical management.

Acknowledgments

The authors thank Teresa Kopel for the help in providing the research subjects.

Supported by National Institutes of Health (NIH; Bethesda, MD, USA) Grants R01 EY019112, R01 EY018853, and R01 EB004640; the Department of Veterans Affairs; the Marlene S. and Leonard A. Hadley Glaucoma Research Fund; the Clifford M. & Ruth M. Altermatt Professorship (YHK); the Frederick C. Blodi Chair (WLMA); and the Robert C. Watzke MD Professorship (MDA).

Disclosure: **Z. Guo**, None; **Y.H. Kwon**, None; **K. Lee**, None; **K. Wang**, None; **A. Wahle**, None; **W.L.M. Alward**, None; **J.H. Fingert**, None; **D.I. Bettis**, None; **C.A. Johnson**, None; **M.K. Garvin**, P; **M. Sonka**, P; **M.D. Abràmoff**, IDx LLC (I, C), P

References

1. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
2. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41:417–421.
3. Wall M, Johnson CA, Kutzko KE, Nguyen R, Brito C, Keltner JL. Long- and short-term variability of automated perimetry results in patients with optic neuritis and healthy subjects. *Arch Ophthalmol*. 1998;116:53–61.
4. Werner EB, Petrig B, Krupin T, Bishop KI. Variability of automated visual fields in clinically stable glaucoma patients. *Invest Ophthalmol Vis Sci*. 1989;30:1083–1089.
5. Mwanza JC, Budenz DL, Godfrey DG, et al. Diagnostic performance of optical coherence tomography ganglion cell–inner plexiform layer thickness measurements in early glaucoma. *Ophthalmology*. 2014;121:849–854.
6. Mwanza JC, Gendy MG, Feuer WJ, Shi W, Budenz DL. Effects of changing operators and instruments on time-domain and spectral-domain OCT measurements of retinal nerve fiber layer thickness. *Ophthalmic Surg Lasers Imaging*. 2011;42:328–337.
7. Mwanza JC, Oakley JD, Budenz DL, Chang RT, Knight OJ, Feuer WJ. Macular ganglion cell-inner plexiform layer: automated detection and thickness reproducibility with spectral domain-optical coherence tomography in glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52:8323–8329.
8. Hood DC, Anderson SC, Wall M, Raza AS, Kardon RH. A test of a linear model of glaucomatous structure-function loss reveals sources of variability in retinal nerve fiber and visual field measurements. *Invest Ophthalmol Vis Sci*. 2009;50:4254–4266.
9. Raza AS, Cho J, de Moraes CG, et al. Retinal ganglion cell layer thickness and local visual field sensitivity in glaucoma. *Arch Ophthalmol*. 2011;129:1529–1536.
10. Leite MT, Zangwill LM, Weinreb RN, Rao HL, Alencar LM, Medeiros FA. Structure-function relationships using the Cirrus spectral domain optical coherence tomograph and standard automated perimetry. *J Glaucoma*. 2012;21:49–54.
11. Hood DC, Raza AS. Method for comparing visual field defects to local RNFL and RGC damage seen on frequency domain OCT in patients with glaucoma. *Biomed Opt Express*. 2011;2:1097–1105.
12. Lee K, Kwon YH, Garvin MK, Niemeijer M, Sonka M, Abramoff MD. Distribution of damage to the entire retinal ganglion cell pathway: quantified using spectral-domain optical coherence tomography analysis in patients with glaucoma. *Arch Ophthalmol*. 2012;130:1118–1126.
13. Bogunovic H, Kwon YH, Rashid A, et al. Relationships of retinal structure and Humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci*. 2015;56:259–271.
14. Wahle A, Lee K, Harding AT, et al. Extending the XNAT archive tool for image and analysis management in ophthalmology research. In: *Proc. SPIE 8674, Medical Imaging 2013: Advanced PACS-based Imaging Informatics and Therapeutic Applications*. Orlando, Florida: 2013:86740M.
15. Garvin MK, Abramoff MD, Wu X, Russell SR, Burns TL, Sonka M. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans Med Imaging*. 2009;28:1436–1447.
16. Quellec G, Lee K, Dolejsi M, Garvin MK, Abramoff MD, Sonka M. Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula. *IEEE Trans Med Imaging*. 2010;29:1321–1330.
17. Bogunovic H, Sonka M, Kwon YH, Kemp P, Abramoff MD, Wu X. Multi-surface and multi-field co-segmentation of 3-D retinal optical coherence tomography. *IEEE Trans Med Imaging*. 2014;33:2242–2253.
18. Garvin MK, Abramoff MD, Lee K, Niemeijer M, Sonka M, Kwon YH. 2-D pattern of nerve fiber bundles in glaucoma emerging from spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53:483–489.
19. Lee K, Niemeijer M, Garvin MK, Kwon YH, Sonka M, Abramoff MD. Segmentation of the optic disc in 3-D OCT

scans of the optic nerve head. *IEEE Trans Med Imaging*. 2010;29:159–168.

20. Mwanza JC, Durbin MK, Budenz DL; for the Cirrus OCTNDSG. Interocular symmetry in peripapillary retinal nerve fiber layer thickness measured with the Cirrus HD-OCT in healthy eyes. *Am J Ophthalmol*. 2011;151:514–521.
21. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809–1815.
22. Abramoff MD, Alward WL, Greenlee EC, et al. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Invest Ophthalmol Vis Sci*. 2007;48:1665–1673.
23. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20:273–297.
24. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
25. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25:141–151.
26. Kleinbaum D, Kupper L, Nizam A, Rosenberg EI. *Applied Regression Analysis and Other Multivariable Methods*. Scarborough, Ontario, Canada: Nelson Education; 2013.
27. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51:5657–5666.
28. Sonka M, Hlavac V, Boyle R. *Image Processing, Analysis, and Machine Vision*. 4th ed. Stamford, CT: Cengage Learning; 2014.
29. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2013;54:1345–1351.
30. Zhang X, Bregman CJ, Raza AS, De Moraes G, Hood DC. Deriving visual field loss based upon OCT of inner retinal thicknesses of the macula. *Biomed Opt Express*. 1 2011;2:1734–1742.
31. Zhang X, Raza AS, Hood DC. Detecting glaucoma with visual fields derived from frequency-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2013;54:3289–3296.
32. Demirkaya N, van Dijk HW, van Schuppen SM, et al. Effect of age on individual retinal layer thickness in normal eyes as measured with spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2013;54:4934–4940.
33. Tatham AJ, Weinreb RN, Medeiros FA. Strategies for improving early detection of glaucoma: the combined structure-function index. *Clin Ophthalmol*. 2014;8:611–621.
34. Lee K, Buitendijk GH, Bogunovic H, et al. Automated segmentability index for layer segmentation of macular SD-OCT images. *Trans Vis Sci Tech*. 2016;5(2):14.
35. Abramoff MD, Wu X, Lee K, Tang L. Subvoxel accurate graph search using non-Euclidean graph space. *PLoS One*. 2014;9:e107763.
36. Gillespie BW, Musch DC, Guire KE, et al. The collaborative initial glaucoma treatment study: baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci*. 2003;44:2613–2620.
37. Delgado MF, Nguyen NT, Cox TA, et al. Automated perimetry: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2002;109:2362–2374.
38. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci*. 2009;50:974–979.
39. Quellec G, Abramoff MD. Estimating maximal measurable performance for automated decision systems from the characteristics of the reference standard. application to diabetic retinopathy screening. *Conf Proc IEEE Eng Med Biol Soc*. 2014;2014:154–157.
40. Bayer AU, Erb C. Short wavelength automated perimetry, frequency doubling technology perimetry, and pattern electroretinography for prediction of progressive glaucomatous standard visual field defects. *Ophthalmology*. 2002;109:1009–1017.
41. Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1–21.
42. Hood DC, Raza AS. On improving the use of OCT imaging for detecting glaucomatous damage. *Br J Ophthalmol*. 2014;98(suppl 2):ii1–ii9.
43. Hood DC, De Cuir N, Blumberg DM, et al. A single wide-field OCT protocol can provide compelling information for the diagnosis of early glaucoma. *Trans Vis Sci Tech*. 2016;5(6):4.

Investigative Ophthalmology & Visual Science