Improving Glaucoma Detection Using Spatially Correspondent Clusters of Damage and by Combining Standard Automated Perimetry and Optical Coherence Tomography

Ali S. Raza,[1,2] Xian Zhang,[1] Carlos G. V. De Moraes,[3] Charles A. Reisman,[4] Jeffrey M. Liebmann,[3,5] Robert Ritch,[5,6] and Donald C. Hood[1,7]

[1]Department of Psychology, Columbia University, New York, New York
[2]Department of Neurobiology and Behavior, Columbia University, New York, New York
[3]Department of Ophthalmology, New York University School of Medicine, New York, New York
[4]Topcon Advanced Biomedical Imaging Laboratory, Topcon Medical Systems, Oakland, New Jersey
[5]New York Eye and Ear Infirmary, New York, New York
[6]Department of Ophthalmology, New York Medical College, Valhalla, New York
[7]Department of Ophthalmology, Columbia University, New York, New York

Correspondence: Donald C. Hood, Department of Psychology, 406 Schermerhorn Hall, 1190 Amsterdam Avenue, MC 5501, Columbia University, New York, NY 10027; dch3@columbia.edu.

Submitted: May 3, 2013
Accepted: December 24, 2013

Citation: Raza AS, Zhang X, De Moraes CGV, et al. Improving glaucoma detection using spatially correspondent clusters of damage and by combining standard automated perimetry and optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2014;55:612–624. DOI:10.1167/iovs.13-12351

Purpose. To improve the detection of glaucoma, techniques for assessing local patterns of damage and for combining structure and function were developed.

Methods. Standard automated perimetry (SAP) and frequency-domain optical coherence tomography (fdOCT) data, consisting of macular retinal ganglion cell plus inner plexiform layer (mRGCPL) as well as macular and optic disc retinal nerve fiber layer (mRNFL and dRNFL) thicknesses, were collected from 52 eyes of 52 healthy controls and 156 eyes of 96 glaucoma suspects and patients. In addition to generating simple global metrics, SAP and fdOCT data were searched for contiguous clusters of abnormal points and converted to a continuous metric (p_{cc}). The p_{cc} metric, along with simpler methods, was used to combine the information from the SAP and fdOCT. The performance of different methods was assessed using the area under receiver operator characteristic curves (AROC scores).

Results. The p_{cc} metric performed better than simple global measures for both the fdOCT and SAP. The best combined structure-function metric (mRGCPL&SAP p_{cc}, AROC = 0.868 ± 0.032) was better (statistically significant) than the best metrics for independent measures of structure and function. When SAP was used as part of the inclusion and exclusion criteria, AROC scores increased for all metrics, including the best combined structure-function metric (AROC = 0.975 ± 0.014).

Conclusions. A combined structure-function metric improved the detection of glaucomatous eyes. Overall, the primary sources of value-added for glaucoma detection stem from the continuous cluster search (the p_{cc}), the mRGCPL data, and the combination of structure and function.

Keywords: glaucoma, glaucomatous, detection, diagnosis, sensitivity, specificity, visual fields, standard automated perimetry, optical coherence tomography, retinal nerve fiber layer, retinal ganglion cells, macula

Because of the progressive nature of the disease, early detection of glaucomatous damage is one of the key objectives of the glaucoma specialist. However, despite the availability of tests that yield quantitative structural and functional measures relevant to glaucoma, there is currently no "gold standard" for diagnosis. In fact, even for a particular diagnostic test, there is usually no universal consensus on what constitutes an abnormal result. For example, a wide variety of criteria are used when interpreting the results from visual field data obtained from standard automated perimetry (SAP), ranging from simple global metrics, such as mean deviation (MD), to more local analyses, such as the glaucoma hemifield test.[1]

Along with other technologies, such as confocal scanning laser ophthalmoscopy and scanning laser polarimetry, the introduction of time-domain optical coherence tomography (tdOCT)[2] more than 2 decades ago has allowed clinicians access to noninvasive in vivo imagining of optic nerve tissue, yielding a quantitative measure of structure in addition to the functional data already provided by the SAP. The tdOCT measures typically used for glaucoma detection are based on the circumpapillary retinal nerve fiber layer (cpRNFL) thickness, which has performed reasonably well in separating groups of healthy controls from glaucoma suspects and patients (see Refs. 3 and 4 for reviews).

Table 1. Population Characteristics

	Controls	Glaucoma	Significance, *P**
Eyes	52	156	
Individuals	52	96	
Age, y	52.7 ± 7.6	55.7 ± 11.9	0.208
MD, dB	−0.4 ± 1.2	−3.1 ± 3.7	0.012
PSD, dB	1.5 ± 0.4	3.2 ± 3.0	0.760

* Significance based on a GEE, accounting for intereye correlations.

The subsequent development of frequency-domain OCT (fdOCT; also referred to as Fourier- or spectral-domain OCT)[5] has allowed for faster, higher-resolution imaging, but has further increased the complexity of the data available. A fundamental question is how to make better use of this data. Although fdOCT has the potential to significantly increase the performance of glaucoma detection as compared with tdOCT, current studies have been underwhelming, suggesting performance similar to, or only incrementally better than, tdOCT (see Refs. 6–8 for reviews). A contributing factor for the apparent similarity of fdOCT and tdOCT performance is that many of these studies, to establish agreement with the tdOCT, limit their analyses of fdOCT data to cpRNFL thickness measures (i.e., 1-dimensional circular scans rather than 2-dimensional optic disc RNFL [dRNFL] thickness measures), which can be derived from volumetric scans.

One approach to make better use of the more complex fdOCT information is machine learning, which has been fairly effective in separating healthy controls from glaucoma suspects and patients using both SAP and tdOCT (see Ref. 9 for a review). For example, a recent study reported an improvement in early glaucoma detection by using fdOCT dRNFL thickness combined with machine learning as compared with using thickness alone.[10] An alternative approach for dealing with the complexity of fdOCT data is to try simpler, more intuitive strategies, such as a topographic pointwise probability map, as used in SAP analysis.[11] A prior study has suggested that measures derived from such a probability map of the dRNFL may outperform the traditional cpRNFL measures.[12]

Another approach is to use the 2-dimensional topographic information from the fdOCT macular scan protocols. In fairness, reports of quantitative structural damage from macular noninvasive imaging have predated the fdOCT (see Ref. 13 for references), although most of these reports, with a few notable exceptions,[14,15] have relied on macular total retinal (mTR) thickness measures. Nonetheless, perhaps surprisingly, even using fdOCT data and increasingly sophisticated segmentation algorithms to better distinguish the macular RNFL (mRNFL) and macular retinal ganglion cell plus inner plexiform layer (mRGCPL) from the combined macular ganglion cell complex (mGCC; defined as mRNFL+mRGCPL) has yielded diagnostic performance that is, at best, equal or marginally better than cpRNFL (see Refs. 13,16, and 17 for reviews; for use of mRGCPL separate from mGCC, see in particular Refs. 18–20). A more sophisticated approach, similar to the pattern deviation (PD) of SAP, also failed to perform better than average cpRNFL thickness.[21]

Yet another technique employed is hemiretina asymmetry (HA),[22] which is similar to the glaucoma hemifield test used in SAP analysis. HA analyses using tdOCT[23] and fdOCT[24,25] did not show a marked improvement over cpRNFL measures. One fdOCT study[26] argued for an improvement over cpRNFL measures, but the glaucoma patient population included only localized RNFL defects based on red-free fundus photos, so these results may not generalize to other populations. Notably, all of these fdOCT HA analyses used mTR measurements instead of a subset of inner retinal layers.

Along with the need to better understand how to use the richer fdOCT dataset, another fundamental question is the potential advantage of combining structural and functional measures for the purpose of glaucoma detection. Previous studies have used various strategies (e.g., simple logical rules,[27–29] machine-learning classifiers,[30–39] or a priori models[40,41]) to combine measures of visual function (e.g., SAP, short-wavelength automated perimetry, and frequency-doubling technology perimetry) with measures of structure (e.g., fundus photos, confocal scanning laser ophthalmoscopy, scanning laser polarimetry, and OCT). These methods have yielded performance equal to or better than methods using structural or functional measures alone.

In this study, we use fdOCT and SAP data to assess various methods, including a novel continuous cluster criterion, in an attempt to better classify individuals as either healthy or glaucomatous. In particular, we combine the information from the SAP and fdOCT data, using both simple logical rules as well as a relatively simple a priori model that tests for spatially correspondent patterns of damage. In addition, the aspects of these analyses that yield the most value-added are discussed.

Methods

Subjects

The glaucomatous group (156 eyes of 96 patients, aged 55.7 ± 11.9 [mean ± SD] years) consisted of patients in whom at least one eye exhibited glaucomatous optic neuropathy, defined based on stereophotography evaluation by glaucoma specialists using the following criteria: focal or diffuse neuroretinal rim thinning, focal or diffuse RNFL loss, or an intereye vertical cup-to-disc ratio asymmetry greater than 0.2 not explained by differences in disc size. All eyes had open angles as viewed during gonioscopic examination. When the fellow eye had reliable test results, it was included, even if all the test results were normal. Therefore, the glaucomatous group included eyes that were glaucoma "suspects" (based on the fellow eye) without any other indication of glaucomatous damage. Moreover, to avoid bias in evaluation of the classification ability of the SAP data and to increase the number of subtle cases included in the study, abnormal SAP results were *not* part of the inclusion criteria for the glaucoma group. Consecutive patients were enrolled retrospectively based on availability of test data. Patients with cataracts, a history of ocular surgery, or a history of any other ocular or neurological diseases that could affect structural or functional measures were excluded.

The control subjects (52 eyes of 52 individuals, aged 52.7 ± 7.6 years) were included based on the following criteria: spherical refraction between −6.0 diopters (D) and +3.0 D, IOP ≤ 21 mmHg, axial length between 22 mm and 26 mm, and a normal clinical examination. Subjects were excluded if they had a history of ocular disease or a family history of glaucoma. Controls were part of a previous study.[42] Normal SAP results were *not* required for the control group. The characteristics of both groups are summarized in Table 1.

Written, informed consent was obtained from all of the participants. Procedures followed the tenets of the Declaration of Helsinki, and the protocol was approved by the institutional review boards of Columbia University and the New York Eye and Ear Infirmary.

Standard Automated Perimetry

All subjects were tested with SAP (24-2 SITA Standard[43] protocol, Humphrey 750i Visual Field Analyzer; Carl Zeiss Meditec, Inc., Dublin, CA). Subjects were required to have

Investigative Ophthalmology & Visual Science

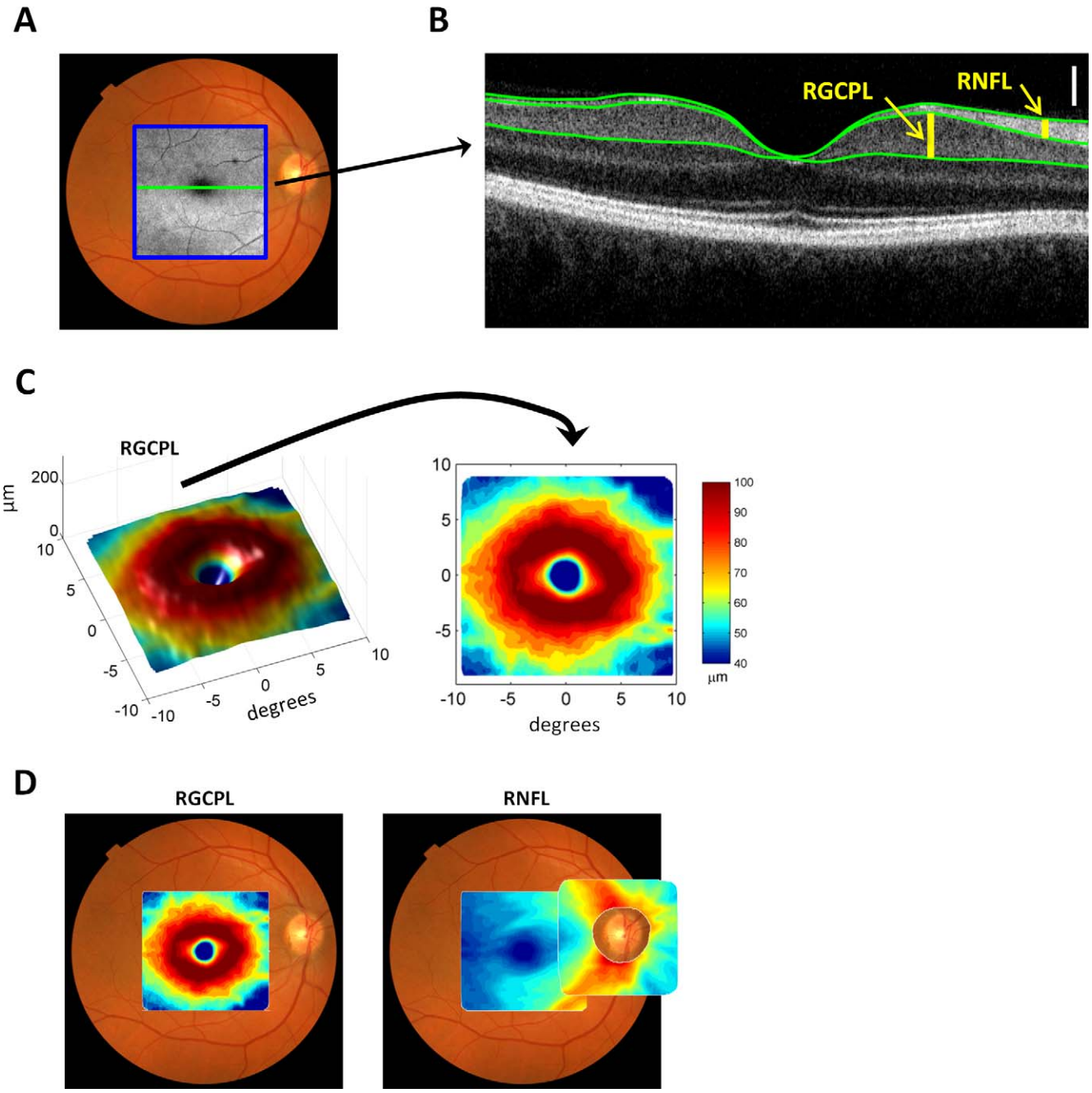

Figure 1. Measurements from fdOCT. (**A**) Fundus photo with an en face fdOCT (C-face) intensity image superimposed within a *blue square*. (**B**) The central slice through the macula of the fdOCT image (B-scan) corresponding to the *horizontal green line* in (**A**). Superimposed *green lines* mark the boundary between anatomical layers. The thicknesses of the RNFL and RGCPL are shown with *yellow vertical bars*. The *white vertical calibration bar* represents 100 µm. (**C**) The 3-dimensional macular RGCPL thickness (*left*) and a top-down view (*right*) with thickness represented in pseudo-color. (**D**) The macular RGCPL superimposed on a fundus photo (*left*) and the macular and disc RNFL superimposed on a fundus photo (*right*).

fixation losses less than or equal to 33%, false-negatives less than or equal to 33%, and false-positives less than or equal to 15%. Other than reliability of the test, the SAP was not used as an inclusion or exclusion criteria. The MD of the control group was −0.4 ± 1.2 dB and of the glaucoma group was −3.1 ± 3.7 dB. The pattern standard deviation (PSD) of the control group was 1.5 ± 0.4 dB and of the glaucoma group was 3.2 ± 3.0 dB. The Advanced Data Export module (Carl Zeiss Meditec, Inc.) was used to obtain the raw data in XML format. To classify the subjects as either normal or glaucomatous, the significance level of the MD of the 24-2 was used as a baseline metric (p_{MD}). The total deviation (TD) SAP data, converted to probability values based on the internal normative database of the machine, also were analyzed for clusters of abnormal points (see Cluster Analysis, below).

Optical Coherence Tomography

All subjects were also tested using fdOCT (3D-OCT 1000/2000; Topcon Medical Systems, Inc., Oakland, NJ) with the volume

Investigative Ophthalmology & Visual Science

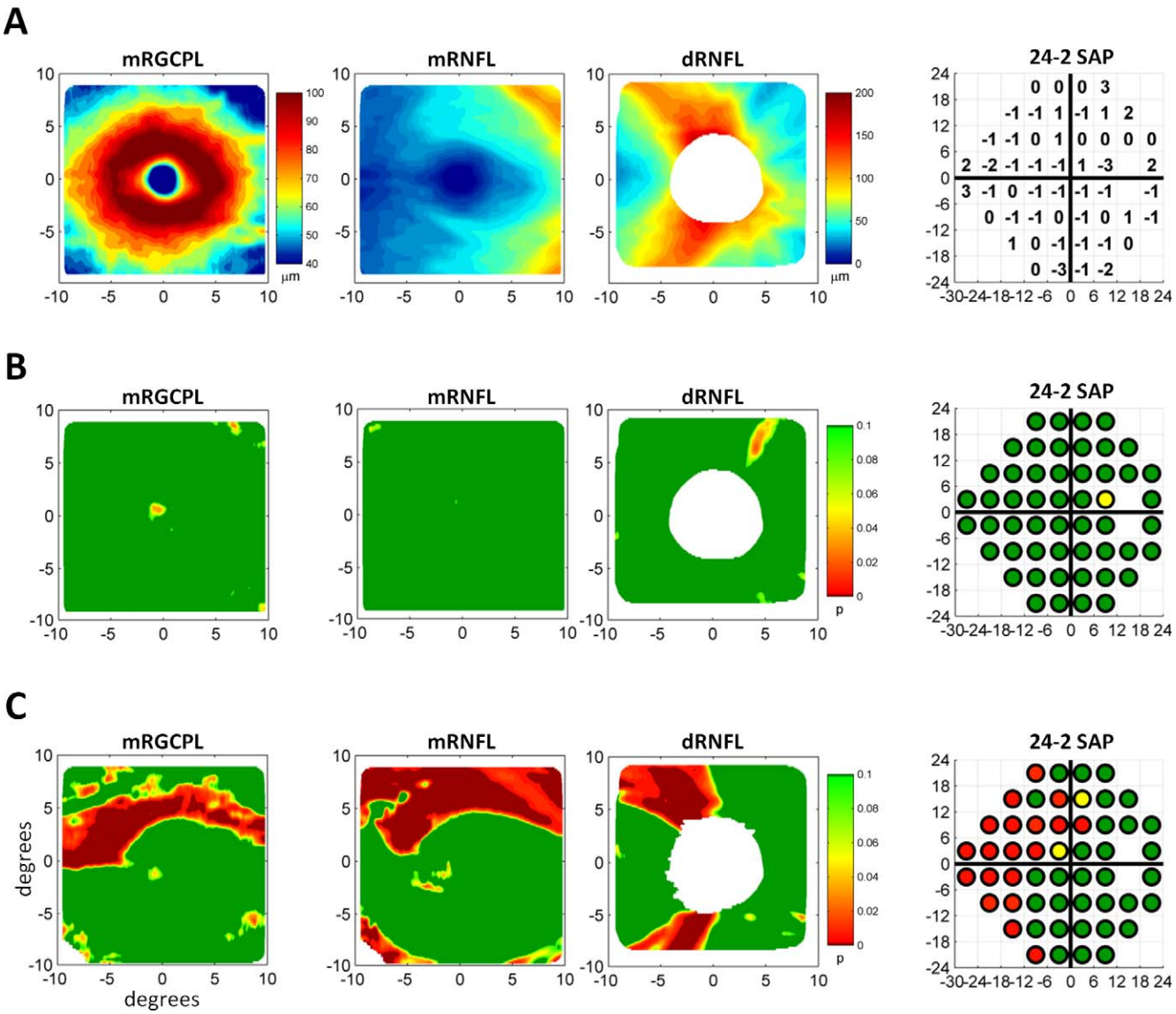

Figure 2. Measurements from fdOCT and SAP converted to probability values. (**A**) The mRGCPL as well as the mRNFL and dRNFL thicknesses for a healthy control eye (*left three panels*). The 24-2 SAP total deviation values for the same eye (*right*). (**B**) The data in (**A**) converted to probability values. (**C**) Data from a glaucomatous eye in the same format as (**B**).

(cube) scan protocol (6 × 6 mm, 128 horizontal B-scans with 512 A-scans each, see Fig. 1A) with both macular and optic disc fixation targets. Scans with poor fixation (as indicated by poor B-scan alignment or by grossly off-center scans) and blink artifacts (as indicated by missing B-scans) were rejected. The thickness of retinal layers was determined using a previously validated segmentation algorithm,[44] which was manually corrected as necessary[45] (see Fig. 1B). In particular, the thickness of the retinal nerve fiber layer (RNFL) and the combined retinal ganglion cell plus inner plexiform layer (RGCPL) was determined for the macular and disc scans (see Figs. 1B–D). Because the RGCPL is quite thin near the disc, only the mRGCPL, mRNFL, and dRNFL were analyzed (Fig. 1D).

The mean thickness of the cpRNFL, converted to probability (p_{MT}), was used as a baseline metric for classifying the subjects as either normal or glaucomatous. (The cpRNFL, extracted from the volume scan, was based on a circle, centered at the optic nerve head, with a diameter of 3.4 mm and was therefore analogous to the 1-dimensional RNFL circle scan obtained from a tdOCT machine.) The mean thickness of a layer over the entire 2-dimensional cube scan, converted to p_{MT}, was also used as a baseline metric. The thickness data were also converted to point-by-point probability values as previously described[11] (see Figs. 2A, 2B for a control and Fig. 2C for a patient) and clusters of abnormal data points were determined (see Cluster Analysis, below). For ease of comparison, the fdOCT data are presented in field view from Figure 2 onward. In addition, the layer thicknesses were normalized in a manner similar to the PD analysis of SAP (by dividing the thickness values based on the 85th percentile of thickness within a scan), and the HA across the horizontal midline was assessed by taking the difference between the superior and inferior retinal thicknesses. Both PD and HA values were also converted to probability values based on controls and were analyzed for clusters of abnormal points. Note that, unlike the thickness probability values, the probability values for the PD and HA were two-tailed because both low and high values suggested abnormalities.

Investigative Ophthalmology & Visual Science

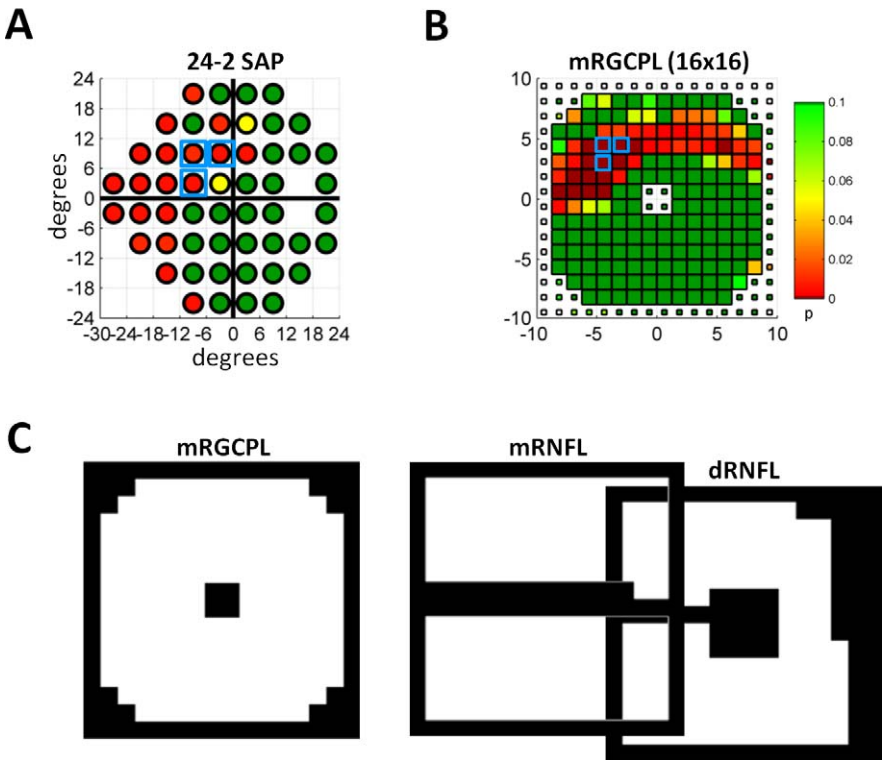

FIGURE 3. Cluster analysis and ROIs. (**A**) The 24-2 SAP data as in Figure 2C with a set of points meeting the 5-5-1 cluster criterion outlined by *light blue boxes*. (**B**) Probability values from mRGCPL thickness data downsampled into a 16 × 16 grid. Data are from the same eye shown in Figure 2C. The *small squares* indicate values outside the ROI. *White squares* indicate missing data (which often occurs at scan edges after centering the scan). A set of points meeting the 5-5-1 cluster criterion is outlined by *light blue boxes*. (**C**) The ROIs shown in *white* for the mRGCPL, mRNFL, and dRNFL data. The dRNFL is offset from the mRNFL data based on anatomy, as shown in Figure 1D.[17,42]

Cluster Analysis

Both SAP and fdOCT data were analyzed for the presence of clusters of abnormal points that respected the horizontal midline. As a baseline metric, the common 5-5-1 cluster criterion was used, requiring a contiguous set of three or more points, all of which were significant at $P \leq 0.05$, with at least one significant at $P \leq 0.01$.[46] For instance, Figure 3A shows a 24-2 SAP with an example of a 5-5-1 cluster (blue squares). (Note that there are many possible 5-5-1 clusters, but only one is indicated.)

For the fdOCT, the data was first downsampled from 128 × 512 pixels to a 16 × 16 grid and then this reduced dataset was converted from thickness to probability values (see Fig. 3B). Next, a region of interest (ROI) was determined a priori for each fixation and thickness combination (see Fig. 3C) based on exclusion of regions that were relatively thin (e.g., fovea or raphe) or regions that were near the edge (1 pixel border) of the scan.[17,42] Data outside the ROI were ignored for cluster evaluation (e.g., smaller squares in Fig. 3B, at the center and edges, corresponding to the black regions in Fig. 3C, indicate data ignored when evaluating clusters). Figure 4 shows the complete set of downsampled fdOCT data for the patient in Figure 2C, including thickness probabilities (Fig. 4A), the PD analysis (Fig. 4B), and the HA analysis (Fig. 4C).

In addition to the 5-5-1 cluster criterion, a novel continuous cluster metric was used. Unlike the 5-5-1, which returned a binary (yes/no) value for each hemifield or hemiretina based on whether or not a cluster was found, the continuous cluster metric returned a continuous value (p_{cc}). This value is related to the probability of observing a particular cluster of points. Similar to the 5-5-1, the p_{cc} required a contiguous set of points, all of which were significant at $P \leq 0.05$ (the threshold probability). However, the p_{cc} metric also considered the number of contiguous points as well as the actual probabilities of these points.

For the fdOCT, the continuous cluster metric generated the p_{cc} value by first assessing, through computer simulation, the probability of observing a set of n contiguous points in a 16 × 16 grid of points (equivalent to the downsampled fdOCT), where the set of points respected the previously defined ROIs and was contained entirely within one hemiretina. For computational simplicity, the maximum tested value for n was 6. Note that the simulation was based on a randomly generated grid of 16 × 16 points and ignored any correlation between the points, which does exist in fdOCT from healthy controls. Thus, this value should not be thought of as a true probability of observation, but rather as a quantitative metric of convenience that ranks the relative probability. To further distinguish between the probabilities of different clusters of points, particularly when the number of points n is the same, the probability of the lowest observed P value in the set was also taken into account when determining p_{cc} by using a separate set of computer simulations. Thus, for each hemiretina, the value of p_{cc} was assigned by multiplying the simulation-determined probability value based on the largest number of n contiguous points found by the simulation-determined probability value based on the lowest observed P value in that set. (If no contiguous set of abnormal points was found [i.e., $n = 1$], then p_{cc} was defined based on the point within the ROI with the lowest probability value, also adjusted based on computer simulations.) For the SAP, the p_{cc} metric was determined in the same manner, although the values of p_{cc} were based on the set of 24-2 SAP points rather than a 16 × 16 grid, requiring a separate set of computer simulations.

Investigative Ophthalmology & Visual Science

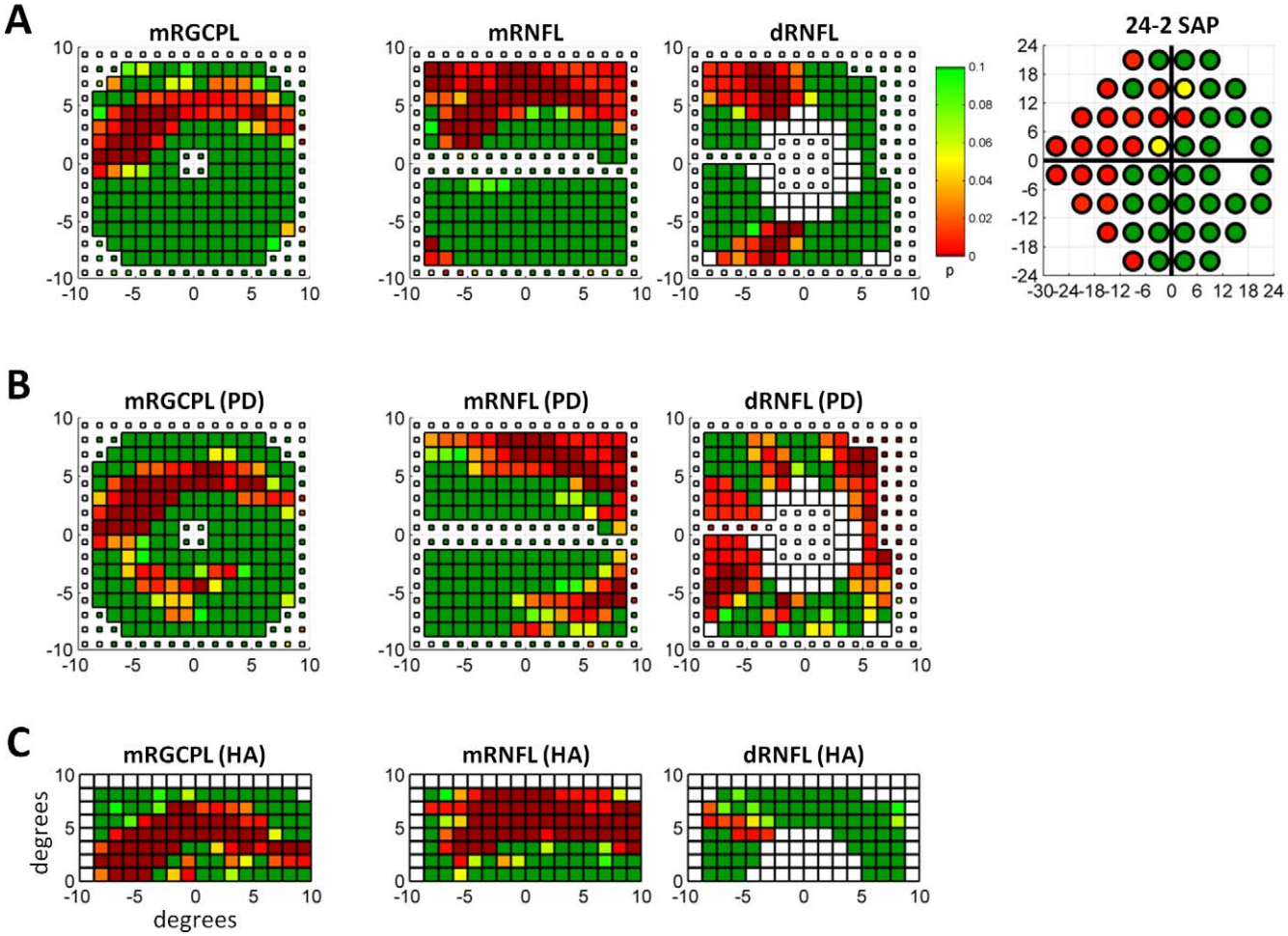

FIGURE 4. Sample data from a glaucomatous eye. (**A**) The downsampled mRGCPL, mRNFL, and dRNFL fdOCT data, as well as the 24-2 SAP data from the same glaucomatous eye shown in Figure 2C. The *white squares* near the center of the dRNFL data represent data missing due to the optic disc. (**B**) The fdOCT data after the PD analysis in a similar format as (**A**). (**C**) The fdOCT data after the HA analysis in a similar format as (**A**).

Combining Structure and Function

Although there are several reasonable ways to combine structural and functional measures, the analysis here was confined to using the p_{cc} metric in a very simple manner. First, structure and function were "combined" by simply multiplying the p_{cc} of the inferior or superior retinal mRGCPL with the p_{cc} of the correspondent hemifield of the SAP. Then, a metric combining the mRGCPL and SAP (mRGCPL&SAP p_{cc}) was defined as the minimum value of the two hemifields. Finally, a more generic combination of fdOCT and SAP data (fdOCT&SAP p_{cc}), using both the mRGCPL and dRNFL, was defined as the minimum of the mRGCPL&SAP p_{cc} and dRNFL&SAP p_{cc}. To separate the effect of using the p_{cc} metric from the combination of structure and function, a "baseline" metric for combining structure and function (fdOCT&SAP $p_{MT/MD}$), based on the mean thickness of the OCT and the MD of the visual field, was defined as the minimum of the mRGCPL p_{MT}, the dRNFL p_{MT}, and the SAP p_{MD}.

Data Analysis and Evaluation

Data were analyzed using custom code written in MATLAB (version 2012a; MathWorks, Inc., Natick, MA). For cluster simulations, cloud-based computing resources were used (EC2 Elastic Compute Cloud; Amazon.com, Inc., Seattle, WA). A generalized estimating equation (GEE) approach, which accounts for intereye correlations, was used to compare means among groups (as in Table 1). When calculating thickness probability in any of the fdOCT metrics, the control group was also used as the normative database. However, to prevent a bias in the reported specificity, a "leave-one-out" approach was used when analyzing the control group. To evaluate the classification of subjects as either normal or glaucomatous, areas under receiver operator characteristic curves (AROC scores) were determined. An AROC score of 1 represents perfect discrimination of the two groups, whereas an AROC of 0.5 indicates performance no better than chance. Standard error (SE) of AROC scores and statistical tests between AROC scores of different classification metrics were performed (as in Tables 2 and 3) according to the method of Obuchowski.[47] This method is based on the nonparametric method of DeLong et al.,[48] but also adjusts for intereye correlations. (This approach has been used in previous studies using similar data [e.g., Ref. 21]) Sensitivity at an arbitrary fixed specificity value of greater than or equal to 85% was also determined. An α level of 0.05 was used to determine statistical significance.

Table 2. Performance of Classification Metrics

Metric	AROC ± SE*	% SN at ≥85% SP
SAP alone		
SAP p_{MD}	0.768 ± 0.029	62
SAP $p_{5\text{-}5\text{-}1}$	0.760 ± 0.030	62
SAP p_{cc}	0.797 ± 0.034	72
fdOCT alone		
cpRNFL p_{MT}	0.718 ± 0.041	56
mRGCPL p_{MT}	0.775 ± 0.038	56
mRGCPL $p_{5\text{-}5\text{-}1}$	0.758 ± 0.034	59
mRGCPL p_{cc}	0.818 ± 0.035	62
mRGCPL (PD) p_{cc}	0.644 ± 0.046	34
mRGCPL (HA) p_{cc}	0.676 ± 0.043	37
dRNFL p_{MT}	0.724 ± 0.041	56
dRNFL $p_{5\text{-}5\text{-}1}$	0.681 ± 0.034	31
dRNFL p_{cc}	0.739 ± 0.037	49
dRNFL (PD) p_{cc}	0.619 ± 0.043	38
dRNFL (HA) p_{cc}	0.520 ± 0.043	24
fdOCT & SAP combined		
mRGCPL&SAP p_{cc}	0.868 ± 0.032	78
fdOCT&SAP p_{cc}	0.859 ± 0.032	78
fdOCT&SAP $p_{MT/MD}$	0.831 ± 0.034	62

SN, sensitivity; SP, specificity.

* Standard error based on the method of Obuchowski,[47] accounting for intereye correlations.

Results

Evaluation of Structure and Function Independently

The performance of the metrics in classifying subjects as either controls or glaucomatous was evaluated using AROC scores (see Table 2), although the sensitivity at an arbitrary fixed specificity of greater than or equal to 85% is also provided for context. First, consider only the SAP data. An example of a clearly abnormal 24-2 SAP, with an MD of –6.74 dB, is shown in Figure 4A (right). Using only the baseline metric of SAP p_{MD} yielded an AROC of 0.768 ± 0.029. The 5-5-1 cluster criterion was applied to the SAP data and resulted in a similar AROC score (0.760 ± 0.030). The continuous cluster criterion, p_{cc}, yielded a slightly higher AROC of 0.797 ± 0.034 and a greater sensitivity of 72% (vs. 62% for SAP p_{MD}). The difference in the AROC scores between the SAP p_{cc} cluster criterion and the SAP $p_{5\text{-}5\text{-}1}$ cluster criterion was statistically significant ($P = 0.018$), although the difference between the SAP p_{cc} and SAP p_{MD} was not ($P = 0.063$).

Consider next the fdOCT data alone. Figure 4A shows an example of mRGCPL, mRNFL, and dRNFL data from a patient with clear glaucomatous damage. In general, the mRNFL data performed slightly worse than either the mRGCPL or the dRNFL, so for the sake of simplicity, the mRNFL data have been omitted from the quantitative results (e.g., Table 2); however, when considering an individual subject, as in Figure 4, the data are still shown for context. The performance of the baseline metrics of mean circumpapillary RNFL thickness and mean optic disc volume RNFL thickness yielded AROC scores lower than those for the MD of the SAP (cpRNFL p_{MT} AROC = 0.718 ± 0.041 and dRNFL p_{MT} = 0.724 ± 0.041, as compared with SAP p_{MD} 0.768 ± 0.029), whereas the AROC score of the mRGCPL p_{MT} (0.775 ± 0.038) was higher than the SAP p_{MD}. However, none of these differences were statistically significant.

Perhaps surprisingly, the PD and HA analyses (see Figs. 4B, 4C for an example), which are similar in concept to the pattern deviation and glaucoma hemifield test analyses commonly used with the SAP, performed worse than the mRGCPL p_{MT} and the dRNFL p_{MT}, which were baseline global metrics based only on mean thicknesses (see Table 2 for AROC scores). Furthermore, although the dRNFL p_{cc} (AROC = 0.739 ± 0.037) performed better than the dRNFL p_{MT}, the sensitivity at 85% specificity was worse and the difference in AROC scores was not statistically significant ($P = 0.312$). However, the mRGCPL p_{cc} (AROC = 0.818 ± 0.035) did perform better than the mRGCPL p_{MT} ($P = 0.040$), with a greater sensitivity of 62% (vs. 56% for mRGCPL p_{MT}).

Evaluation of Combined Structure and Function

The performance of the best fdOCT global metric, mRGCPL p_{MT}, was similar to the SAP global metric, SAP p_{MD}. Likewise, the performance of the overall best fdOCT metric, mRGCPL p_{cc}, was similar to the best SAP metric, SAP p_{cc}. To assess the value-added from combining structure and function, we incorporated the best structural and functional metrics (the p_{cc} metric for both mRGCPL and SAP) into a single metric, mRGCPL&SAP p_{cc}. This combined structure-function technique yielded an AROC of 0.868 ± 0.032. We then compared the performance of this metric to the baseline (global) and p_{cc} (local patterns) metrics of structure and function considered

Table 3. Comparison of Classification Metrics

Metric	AROC ± SE*	Compared Metric	AROC ± SE*	Significance, P*
Entire population (n = 52 control eyes, 156 glaucoma/suspect eyes)				
mRGCPL&SAP p_{cc}	0.868 ± 0.032	SAP p_{MD}	0.768 ± 0.029	<0.001
		SAP p_{cc}	0.797 ± 0.034	0.003
		cpRNFL p_{MT}	0.718 ± 0.041	<0.001
		mRGCPL p_{MT}	0.775 ± 0.038	0.001
		mRGCPL p_{cc}	0.818 ± 0.035	0.005
		fdOCT&SAP $p_{MT/MD}$	0.831 ± 0.034	0.038
SAP (PSD) used as inclusion and exclusion criteria (n = 49 control eyes, 73 glaucoma/suspect eyes)				
mRGCPL&SAP p_{cc}	0.975 ± 0.014	SAP p_{MD}	0.897 ± 0.026	<0.001
		SAP p_{cc}	0.942 ± 0.024	0.066
		cpRNFL p_{MT}	0.809 ± 0.045	<0.001
		mRGCPL p_{MT}	0.867 ± 0.034	<0.001
		mRGCPL p_{cc}	0.922 ± 0.025	0.004
		fdOCT&SAP $p_{MT/MD}$	0.931 ± 0.025	0.003

* Significance and SE based on the method of Obuchowski,[47] accounting for intereye correlations.

Investigative Ophthalmology & Visual Science

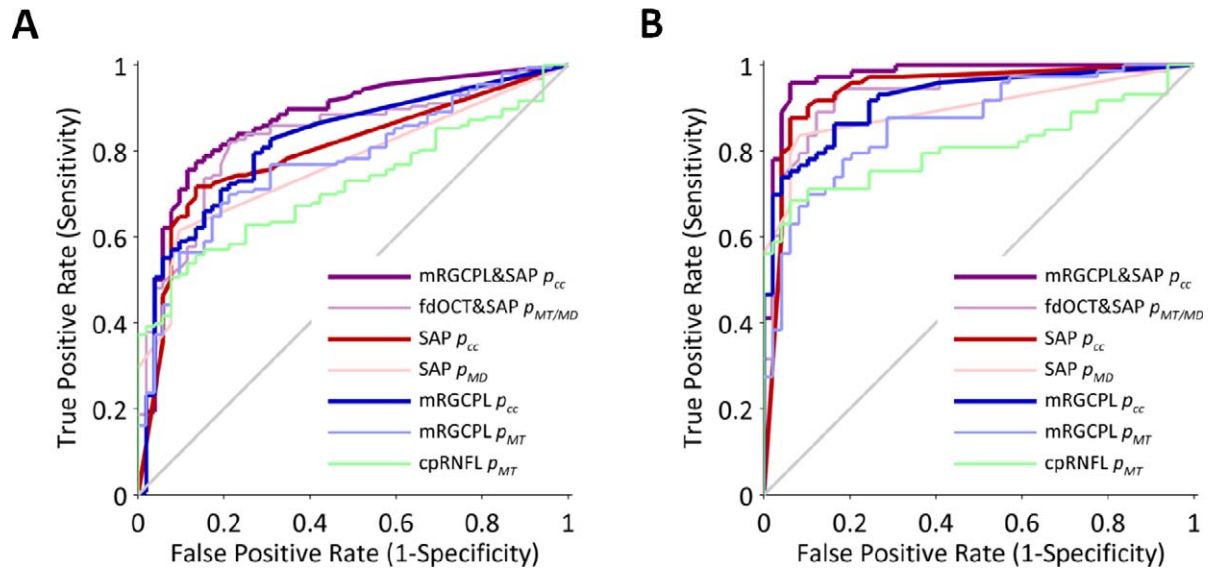

Figure 5. Receiver operator characteristic curves for various metrics. (**A**) Receiver operator characteristic curves for the entire population. (**B**) Receiver operator characteristic curves for a subpopulation in which SAP was used as inclusion and exclusion criteria.

independently (Table 3). As shown in Figure 5A, the mRGCPL&SAP p_{cc} had a significantly higher AROC than the SAP p_{MD} ($P < 0.001$) and the SAP p_{cc} metrics ($P = 0.003$) as well as the cpRNFL p_{MT} ($P < 0.001$), mRGCPL p_{MT} ($P = 0.001$), and mRGCPL p_{cc} metrics ($P = 0.005$).

However, because our combined structure-function metric relies on the continuous cluster method, we wanted to determine to what extent the higher AROC of the combined structure-function metric could be explained by merely using both fdOCT and SAP data, independent of the p_{cc} metric. Therefore, we tested a structure-function "baseline" metric based on the minimum of the SAP p_{MD}, mRGCPL p_{MT}, and dRNFL p_{MT}; this "baseline" method (fdOCT&SAP $p_{MT/MD}$) can be thought of as classifying a subject as abnormal if either structure *or* function is abnormal (simple logical rules on global measures), without testing for subtle patterns (as with the p_{cc}) and without combining the structural and functional information in a weighted and spatially correspondent manner (see Methods for details of the mRGCPL&SAP p_{cc}). When we compared the mRGCPL&SAP p_{cc} AROC (0.868 ± 0.032) to the fdOCT&SAP $p_{MT/MD}$ AROC (0.831 ± 0.034), the p_{cc} structure-function metric performed better ($P = 0.038$) than the "baseline" structure-function metric (Fig. 5A).

Evaluation of a Subpopulation Selected Based on SAP

SAP was not used as inclusion or exclusion criteria for the control or glaucomatous groups to avoid biasing the AROC scores in favor of SAP-based metrics and, more importantly, to allow for inclusion of subtle cases. However, to allow for a comparison with studies that use SAP as inclusion criteria, we analyzed a subpopulation in which a SAP criterion (PSD $P \leq 0.05$) was used as exclusion criteria for the control group and inclusion criteria for the glaucoma group. As expected, the AROC scores increased for each metric, with the mRGCPL&SAP p_{cc} metric now yielding an AROC of 0.975 ± 0.014. The mRGCPL&SAP p_{cc} metric performed significantly ($P < 0.001$ to $P = 0.004$) better than all the other methods that were previously tested on the full population (see Table 3 and Fig. 5B), with the exception of the SAP p_{cc} metric ($P = 0.066$), which is not surprising given that this subpopulation was defined based on SAP. For structural measurements considered independently, the p_{cc} metric performed better than the p_{MT} metric for the mRGCPL, mRNFL, and dRNFL, although this difference was statistically significant only for the mRGCPL ($P = 0.005$) and mRNFL ($P = 0.012$). For functional measures considered independently, the SAP p_{cc} metric performed better than the SAP p_{MD} metric ($P = 0.019$).

Discussion

Our purpose here was to combine structural and functional data, to make better use of the individual OCT and SAP measures, and to assist with glaucoma detection. We found that combining structure and function by taking into consideration spatially correspondent patterns of damage improved the proper classification of the control and glaucomatous groups. In general, the techniques used in this study improved the diagnostic capabilities of the fdOCT and SAP data in two steps. First, we considered structural and functional measures independently by using a novel technique that searches for many different types of spatially contiguous clusters of abnormal points and then converts the result into a single, continuous value (i.e., the p_{cc} metric). Next, we combined the structural and functional information in a spatially correspondent manner. Both steps offer value-added, although in different ways, as discussed below.

Evaluation of Structure and Function Independently

The continuous cluster (p_{cc}) metric has theoretical advantages when compared with similar techniques used in the past. When evaluating SAP data, the common cluster criterion of three contiguous points significantly abnormal at $P \leq 0.05$, with at least one point significant at $P \leq 0.01$ (i.e., the 5-5-1 cluster criterion), has been shown to offer classification performance that is on par with or better than other commonly used SAP metrics.[46] However, this criterion yields a binary response (whether or not a cluster is found) rather than a

Investigative Ophthalmology & Visual Science

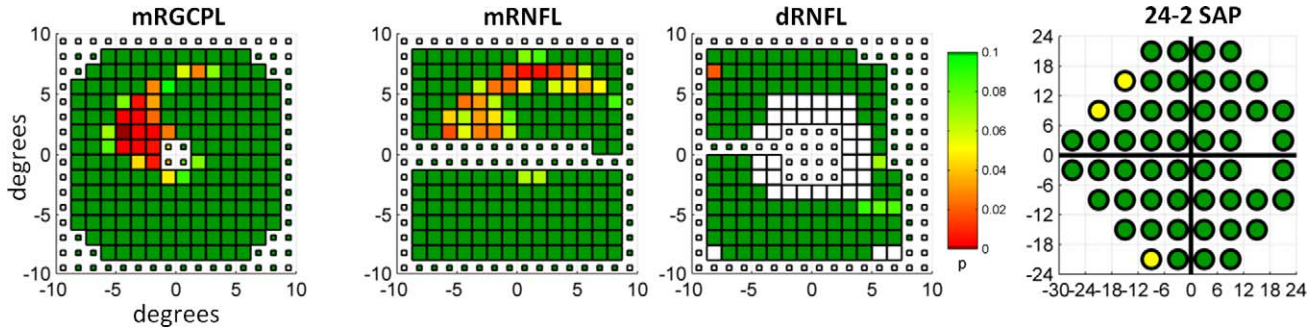

Figure 6. Example of an eye, belonging to the glaucomatous group, with a subtle arcuate defect in the macula. Data are in the same form as shown in Figure 4A.

continuous metric and, at times, this rule can seem arbitrary. For instance, the 5-5-1 criterion does not consider the number of contiguous points nor the actual probabilities. As long as there are three or more contiguous points significant at $P \leq 0.05$, the 5-5-1 does not distinguish between, for example, three contiguous points versus six contiguous points, even though the latter is much less likely to be observed due to chance alone. Thus, a set of six contiguous points at $P = 0.02$ would not meet the 5-5-1 criterion (because not a single point has $P \leq 0.01$), despite being less probable than three contiguous points with two points at $P = 0.05$ and one point at $P = 0.01$. Likewise, the 5-5-1 criterion treats two points at $P = 0.05$ and one point at $P = 0.01$ the same as three points at $P < 0.001$, even though, again, the latter is much less likely.

Our use of the p_{cc} metric does involve some assumptions. For the fdOCT data, the appearance of abnormal points may be correlated (i.e., spatially clustered) in controls due to interindividual variability. For example, if a control has an RNFL distribution at the optic disc that differs from most other controls, a pattern of contiguous abnormal points may appear. One way to address this issue is the use of "superpixels" to group correlated areas.[10] Another approach is to better understand factors leading to interindividual variability among controls and to attempt to correct or control for these underlying factors in the fdOCT data. In this study, we simply take the p_{cc} metric applied to the fdOCT data as an approximate "score" that ranks relative probabilities of observing clusters of abnormal points. Because the metric yields continuous values, an ROC curve can be generated by varying the threshold for considering a p_{cc} value as abnormal, which yields an empirical result for the optimal value in our population (without making any assumptions regarding the true probabilities of observing particular clusters of abnormal points in the fdOCT data). In the future, an independent study with a larger number of controls or a more advanced simulation might yield a better understanding of how the p_{cc} metric relates to true probabilities of observation in the fdOCT data. In any case, empirically, the p_{cc} metric performs better than the other methods tested in this study.

A specific example illustrating the advantage of the p_{cc} metric when applied to fdOCT data can be seen in Figure 6. The average thickness of the mRGCPL for this eye lies near the 32nd percentile for the eyes in the control group. Although the mRGCPL p_{MT} classifies this eye as a control when using the optimal point of the ROC curve, the mRGCPL p_{cc} metric classifies this eye as glaucomatous. Surprisingly, the PD and HA analyses did not perform very well overall. (It is important to realize that the focal loss volume metric,[21] although it makes use of a pattern deviation analysis for the mGCC, is implemented differently from the PD method tested here. For details, see Ref. 15.) Certainly, there are some specific cases in which these strategies would be helpful. For instance, Figure 7A shows a control mRGCPL that had considerably fewer abnormal points once the PD analysis was applied. Similarly, Figure 7B shows the mRNFL of a patient (same eye as Fig. 6) in whom the abnormal points within an arcuate became more statistically significant when using the HA analysis. However, overall, these strategies did not offer much value-added beyond mean thickness. It is possible that the PD analysis removed some diffuse thinning that assisted with classifying the glaucomatous eyes when using mean thickness. Similarly, it is possible that symmetric damage was removed when using the HA analysis. It is also possible that in most cases the defects enhanced by the HA technique would already be detected using analyses based on mean thickness. In other words, perhaps early glaucomatous damage in the macula is always either diffuse (symmetric across the horizontal midline) or focal damage in the form of a deep, thin arcuate (as in Fig. 6). If true, then the HA technique would not offer much of an advantage in either of these cases. Notably, recent studies[24–26] using an HA strategy in fdOCT do not offer comparisons to the performance of macular measurements based simply on mean thickness. In any case, it is clear that there are some advantages to the PD and HA strategies on a case-by-case basis (as seen in Fig. 7).

Whereas our use of the mRGCPL thickness distinguishes our study from previous reports combining structure and function, recent studies[18–20] have also considered the diagnostic value of the mRGCPL (separate from the mGCC) when considering structure alone. In particular, Mwanza et al.[18] and Takayama et al.[19] have used a "minimum spoke" approach to analyzing the mRGCPL. Although minimum probability values performed worse than the p_{cc} metric in our study (data not shown), the "minimum spoke" approach is more complex than the analysis we used, so it is difficult to make a meaningful comparison. Further work would be needed to compare these methods in a fair manner. In any case, the p_{cc} metric is a more general approach that can be applied to the thickness of any layer at various regions of the retina, whereas the "minimum spoke" approach is probably most applicable to mRGCPL thickness.

Prior studies considering the diagnostic power of the macular region in fdOCT have almost always found performance similar to cpRNFL or dRNFL (see Refs. 13 and 16 for reviews). However, in our study, the best mRGCPL parameter performed considerably better than cpRNFL or the best dRNFL parameter. The reason for this discrepancy is not entirely clear; it may be due to population differences or other factors. Regardless, although our population may include more glaucomatous individuals with central defects than in other studies, the lack of a large source of bias suggests that a fair

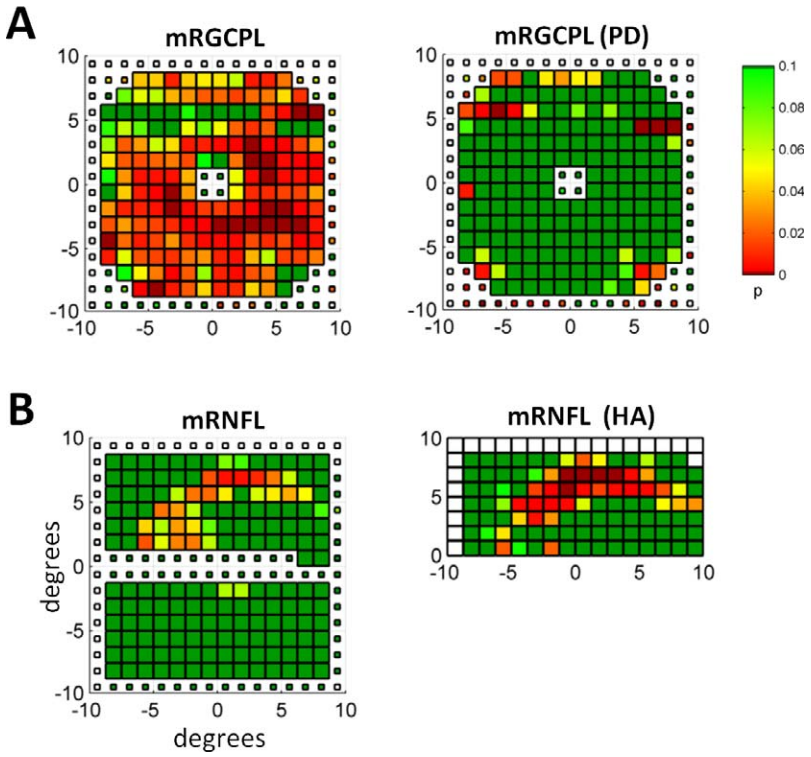

Figure 7. Examples of PD and HA analyses. (**A**) The mRGCPL probability values based on thickness (*left*) and the PD analysis (*right*) for an eye belonging to the healthy control group. (**B**) The mRNFL probability values based on thickness (*left*) and the HA analysis (*right*) for an eye belonging to the glaucomatous group. Same eye as shown in Figure 6.

number of these individuals do exist and would therefore benefit from better analysis of the macular fdOCT data.

Evaluation of Combined Structure and Function

Combining structure and function yielded an improvement beyond using structural and functional measures independently, in agreement with many of the previous studies combining structure and function to aid in glaucoma detection.[27–41] Although some of these studies have evaluated the benefit of combining structure and function in comparison with simple independent measures, such as the average cpRNFL thickness, here we followed the logic of those studies that tested the combined structure-function method against the best independent measures of structure and function (i.e., the p_{cc} metric in this study). Additionally, previous studies that have used machine-learning classifiers or a priori models have often neglected to include a simple combination of structure and function to test for the value-added of using a more complex technique to combine structure and function. Here, we compare our best combined structure-function metric (mRGCPL&SAP p_{cc}) against a simpler model of combining structure and function (fdOCT&SAP $p_{MT/MD}$) to provide a more rigorous test for the value-added of our technique. It is important to note that even our "simple" model of combining structure and function involves more than the simple logical rules used in past studies. For instance, if we were to consider an individual glaucomatous if either the SAP p_{MD} or the mRGCPL p_{MT} or the dRNFL p_{MT} were abnormal at a significance level of $P \leq 0.05$, then the corresponding AROC score would be 0.785 ± 0.033 (compare with values in Table 2), slightly better than simple metrics using structure or function alone, but considerably worse than even our "baseline" combination of structure and function (fdOCT&SAP $p_{MT/MD}$), which merely uses the minimum value of either the SAP p_{MD} or the mRGCPL p_{MT} or the dRNFL p_{MT} to provide a continuous metric.

Many of the previous attempts to combine structure and function used machine-learning approaches,[30–39] including an early study by Caprioli.[30] (Here, we include linear discriminant analysis under the umbrella of machine learning.) Two recent studies by Bowd et al.[36] and Bizios et al.[38] used tdOCT and SAP. Bizios et al.[38] found that their artificial neural network approach tested on a moderate glaucomatous population (SAP MD = -11.0 ± 8.2 dB) yielded better discrimination than cpRNFL or SAP MD alone, although not better than using their machine-learning approach on cpRNFL alone. Bowd et al.[36] found that their Bayesian machine-learning classifiers using both structure and function, tested on a mild glaucomatous population (SAP MD = -3.1 ± 3.4 dB), performed better than cpRNFL or SAP MD alone and also performed better than the best machine-learning approach on independent structural and functional measures, although the difference between the best functional measure and the combined structure and function approach was not statistically significant. Other studies using machine-learning approaches on data other than OCT have also reported improvements when combining structure and function as opposed to the best independent structural or functional measure.[33,34,37,39] Ultimately, the philosophy behind machine-learning approaches differs fundamentally from the approach used in this study. There is no clear "right answer"; both machine-learning and a priori models involve trade-offs.

Even relatively simple logical rules to combine structure and function have been shown to be useful in previous studies.[27–29] In particular, the study of Shah et al.[27] concluded that combining structure and function offered a statistically significant advantage over the best independent structural and functional measures, even though their method of combining structure and function did not require spatial correspondence.

The study of Hirashima et al.[29] is, to the best of our knowledge, the only previous work combining structure and function to aid in glaucoma detection that uses 2-dimensional fdOCT data; however, when combining structure and function, the authors do not account for spatial correspondence and also use very simple global measures, such as average cpRNFL thickness, average mGCC thickness, and MD.

Two recent studies[40,41] have used more complex a priori models to combine structure and function in glaucoma while taking into account spatial correspondence. Boland and Quigley[40] created a simple model that relates the probability of abnormality in both confocal scanning laser ophthalmoscopy and SAP data to the probability that these abnormal data points are spatially correspondent, based on an extension of a structure-function spatial map created by Garway-Heath et al.[49] Although the Boland and Quigley[40] approach to relating structure and function is elegant, we used a coarser approach for spatial correspondence for two reasons: the Garway-Heath et al.[49] map in its current form does not fully apply to 2-dimensional fdOCT dRNFL data and the lack of 10-2 SAP data on all subjects prevented precise spatial relationships in the macula. In any case, the Boland and Quigley[40] model combining structure and function did not perform better than using the MD of SAP alone; however, their study attempted the more difficult task of separating glaucoma patients from glaucoma suspects (rather than from healthy controls).

Another important a priori model, which combines 1-dimensional cpRNFL fdOCT data with SAP, has been presented in a recent study by Medeiros et al.[41] Their work attempts the much more ambitious task of relating fdOCT to SAP by first converting both datasets to estimated ganglion cell counts based on a model by Harwerth et al.[50] The Medeiros et al.[41] model for combining structure and function performed better than simple global independent measures, such as average cpRNFL thickness and SAP MD, and this difference was statistically significant. However, the authors did not test their model against more complex independent measures of structure and function or against a simple model combining structure and function. Additionally, it is worth noting that the weighting function applied to the combination of structure and function is heavily biased toward the structural measure for mild glaucomatous defects (as defined by the SAP). The relative weighting for the fdOCT-derived estimated ganglion cell counts is 90% when the MD of the 24-2 SAP is –3 dB and 80% when the MD of the 24-2 SAP is –6 dB. Thus, it is not surprising that the performance of the Medeiros et al.[41] combined structure-function metric was equivalent to average cpRNFL thickness for their preperimetric glaucoma group (admittedly a difficult group for glaucoma detection). Although the weighting function of Medeiros et al.[41] makes their combined structure-function metric largely dependent on structural measures for mild glaucomatous damage (as defined by the SAP), in fairness, their metric appears to be aimed more at progression than early glaucoma detection. Indeed, further work by this group has framed this combined structure-function metric in the context of progression.[51,52] Although their combined structure-function metric is limited by the assumptions and validity of the Harwerth et al.[50] model (as mentioned by Medeiros et al.[41] in their discussion), the conversion of structural and functional measures to estimated ganglion cell counts allows for interesting speculations regarding disease mechanisms in glaucoma.[51]

While combining only fdOCT dRNFL and SAP data (data not shown) in this study performed better than using dRNFL or SAP data alone, the best performance was achieved by combining mRGCPL, dRNFL, and SAP data or simply mRGCPL and SAP data. Although it may seem unusual that the dRNFL seems to add little to the combination of mRGCPL and SAP, the lack of 10-2 SAP data probably makes the 24-2 SAP and dRNFL somewhat redundant, whereas the mRGCPL offers complementary information pertaining to macular damage. The dRNFL may also be subject to interindividual variability to a higher degree than the mRGCPL. Because 10-2 SAP is not routinely collected in many clinical settings and is not officially a part of the protocol for the data collected by many important multisite studies, the mRGCPL data may serve as an important indicator of macular damage that might otherwise go unnoticed in glaucoma suspects and patients.

Inclusion and Exclusion Criteria

By not using more stringent inclusion and exclusion criteria at the outset, we allowed for the inclusion of more subtle cases. For instance, we included the fellow eye of glaucoma suspects even if there was no evidence of glaucomatous optic neuropathy or abnormal visual fields in that eye. Our rationale here is similar to the arguments made by Bowd et al.[36] in their study combining SAP and tdOCT data using machine-learning classifiers. The MD of the SAP in our glaucomatous group was similar to their study and notably higher than some other studies (e.g., Refs. 33 and 38) combining structure and function. In particular, 87% of the eyes in our glaucomatous group had an MD greater than –6 dB, suggesting "mild" glaucoma according to the Hodapp-Anderson-Parrish classification system.[53] An example of an interesting subtle case that would not have been included otherwise can be seen in Figure 6. However, although our population more closely resembles a screening population, the AROC scores we report (e.g., Table 2) are probably lower because our population is not strictly defined. Thus, when assessing the AROC scores in Table 2, it is important to realize that the maximum possible AROC is probably less than 1 in this context.

To illustrate the differences in the AROC scores, as well as to demonstrate the performance of these methods on a more well-defined population, we repeated our analyses on a subpopulation using SAP as part of the inclusion and exclusion criteria (see Table 3 and Fig. 5B). Although the performance of SAP-based metrics is likely to be favorably biased in this subpopulation, it is important to note that the performance of different metrics within a particular modality (i.e., structure or function) will not be affected by such a bias. For example, regarding the mRGCPL or the SAP data, there is no compelling reason to believe that the better performance of the p_{cc} metric was affected by a bias in how the subpopulation was selected. Ultimately, the lack of a "gold standard" for glaucoma diagnosis means there will be a certain amount of ambiguity in interpreting AROC scores, particularly in studies attempting to combine structure and function, regardless of how the populations are defined.

Future Directions

The methods used here could benefit from a more refined model for spatial correspondence between structural and functional measures. Although a precise local point-to-point multiplication of probabilities between structure and function yielded performance better than structure or function alone (data not shown), the lack of 10-2 SAP data led us to use a simpler method of combining structure and function. Additionally, continuous probability values for the SAP data in this study,[54] as opposed to the discrete values derived from machine-based normative data, may further enhance the performance of the combined structure-function techniques shown here. Finally, although we used a priori assumptions in this study to derive the ROIs for the fdOCT data, optimization

Investigative Ophthalmology & Visual Science

based on our current dataset may yield better performance in future studies with independent datasets.

Conclusions

A combined structure-function metric, taking into consideration spatially correspondent patterns of damage, improved the detection of glaucomatous eyes. Overall, evaluation of the techniques used in this study suggests that the primary sources of value-added stem from the continuous cluster search (the p_{cc}), the mRGCPL data, and the combination of structure and function.

Acknowledgments

The authors thank Danilo B. Fernandes, Monica F. Chen, Diane L. Wang, and David W. Rhee for their assistance in the preparation of this manuscript.

Supported by National Institutes of Health Grant R01-EY-02115 (DCH), National Science Foundation Graduate Research Fellowship Grant DGE-11-44155 (ASR), and a grant of equipment from Topcon, Inc.

Disclosure: **A.S. Raza**, None; **X. Zhang**, Topcon (C); **C.G.V. De Moraes**, None; **C.A. Reisman**, Topcon (E); **J.M. Liebmann**, Topcon (F, C); **R. Ritch**, None; **D.C. Hood**, Topcon (F, C)

References

1. Asman P, Heijl A. Glaucoma hemifield test: automated visual field evaluation. *Arch Ophthalmol*. 1992;110:812–819.
2. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*. 1991;254:1178–1181.
3. Sharma P, Sample PA, Zangwill LM, Schuman JS. Diagnostic tools for glaucoma detection and management. *Surv Ophthalmol*. 2008;53:17–32.
4. Chang R, Budenz DL. New developments in optical coherence tomography for glaucoma. *Curr Opin Ophthalmol*. 2008;19:127–135.
5. Wojtkowski M, Fercher AF, Leitgeb R. Phase-sensitive interferometry in optical coherence tomography. *Proc SPIE*. 2001;4515:250–255.
6. Savini G, Carbonelli M, Barboni P. Spectral-domain optical coherence tomography for the diagnosis and follow-up of glaucoma. *Curr Opin Ophthalmol*. 2011;22:115–123.
7. Sung KR, Kim JS, Wollstein G, Folio L, Kook MS, Schuman JS. Imaging of the retinal nerve fibre layer with spectral domain optical coherence tomography for glaucoma diagnosis. *Br J Ophthalmol*. 2011;95:909–914.
8. Grewal DS, Tanna AP. Diagnosis of glaucoma and detection of glaucoma progression using spectral domain optical coherence tomography. *Curr Opin Ophthalmol*. 2013;24:150–161.
9. Bowd C, Goldbaum MH. Machine learning classifiers in glaucoma. *Optom Vis Sci*. 2008;85:396–405.
10. Xu J, Ishikawa H, Wollstein G, Schuman JS. 3D optical coherence tomography super pixel with machine classifier analysis for glaucoma detection. *Conf Proc IEEE Eng Med Biol Soc*. 2011;2011:3395–3398.
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. Leung CKS, Lam S, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: analysis of the retinal nerve fiber layer map for glaucoma detection. *Ophthalmology*. 2010;117:1684–1691.
13. Sung KR, Wollstein G, Kim NR, et al. Macula assessment using optical coherence tomography for glaucoma diagnosis. *Br J Ophthalmol*. 2012;96:1452–1455.
14. Ishikawa H, Stein DM, Wollstein G, Beaton S, Fujimoto JG, Schuman JS. Macular segmentation with optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2005;46:2012–2017.
15. Tan O, Li G, Lu AT-H, Varma R, Huang D. Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis. *Ophthalmology*. 2008;115:949–956.
16. Wong JJ, Chen TC, Shen LQ, Pasquale LR. Macular imaging for glaucoma using spectral-domain optical coherence tomography: a review. *Semin Ophthalmol*. 2012;27:160–166.
17. Hood DC, Raza AS, De Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1–21.
18. Mwanza J-C, Durbin MK, Budenz DL, et al. Glaucoma diagnostic accuracy of ganglion cell-inner plexiform layer thickness: comparison with nerve fiber layer and optic nerve head. *Ophthalmology*. 2012;119:1151–1158.
19. Takayama K, Hangai M, Durbin M, et al. A novel method to detect local ganglion cell loss in early glaucoma using spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53:6904–6913.
20. Kotowski J, Folio LS, Wollstein G, et al. Glaucoma discrimination of segmented cirrus spectral domain optical coherence tomography (SD-OCT) macular scans. *Br J Ophthalmol*. 2012;96:1420–1425.
21. Tan O, Chopra V, Lu AT-H, et al. Detection of macular ganglion cell loss in glaucoma by fourier-domain optical coherence tomography. *Ophthalmology*. 2009;116:2305–2314.
22. Zeimer R, Asrani S, Zou S, Quigley H, Jampel H. Quantitative detection of glaucomatous damage at the posterior pole by retinal thickness mapping. A pilot study. *Ophthalmology*. 1998;105:224–231.
23. Bagga H, Greenfield DS, Knighton RW. Macular symmetry testing for glaucoma detection. *J Glaucoma*. 2005;14:358–363.
24. Asrani S, Rosdahl J, Allingham R. Novel software strategy for glaucoma diagnosis. *Arch Ophthalmol*. 2011;129:1205–1211.
25. Um TW, Sung KR, Wollstein G, Yun S-C, Na JH, Schuman JS. Asymmetry in hemifield macular thickness as an early indicator of glaucomatous change. *Invest Ophthalmol Vis Sci*. 2012;53:1139–1144.
26. Seo JH, Kim T-W, Weinreb RN, Park KH, Kim SH, Kim DM. Detection of localized retinal nerve fiber layer defects with posterior pole asymmetry analysis of spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53:4347–4353.
27. Shah NN, Bowd C, Medeiros FA, et al. Combining structural and functional testing for detection of glaucoma. *Ophthalmology*. 2006;113:1593–1602.
28. Tóth M, Kóthy P, Vargha P, Holló G. Accuracy of combined GDx-VCC and matrix FDT in a glaucoma screening trial. *J Glaucoma*. 2007;16:462–470.
29. Hirashima T, Hangai M, Nukada M, et al. Frequency-doubling technology and retinal measurements with spectral-domain optical coherence tomography in preperimetric glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 2013;251:129–137.
30. Caprioli J. Discrimination between normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 1992;33:153–159.
31. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol*. 1996;121:511–521.
32. Vihanninjoki K, Teesalu P, Burk RO, Läärä E, Tuulonen A, Airaksinen PJ. Search for an optimal combination of structural and functional parameters for the diagnosis of glaucoma. multivariate analysis of confocal scanning laser tomograph,

blue-on-yellow visual field and retinal nerve fiber layer data. *Graefes Arch Clin Exp Ophthalmol*. 2000;238:477–481.

33. Lauande-Pimentel R, Carvalho RA, Oliveira HC, Gonçalves DC, Silva LM, Costa VP. Discrimination between normal and glaucomatous eyes with visual field and scanning laser polarimetry measurements. *Br J Ophthalmol*. 2001;85:586–591.
34. Horn FK, Nguyen NX, Mardin CY, Junemann AG. Combined use of frequency doubling perimetry and polarimetric measurements of retinal nerve fiber layer in glaucoma detection. *Am J Ophthalmol*. 2003;135:160–168.
35. Mardin CY, Peters A, Horn F, Jünemann AG, Lausen B. Improving glaucoma diagnosis by the combination of perimetry and HRT measurements. *J Glaucoma*. 2006;15:299–305.
36. Bowd C, Hao J, Tavares IM, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2008;49:945–953.
37. Racette L, Chiou CY, Hao J, et al. Combining functional and structural tests improves the diagnostic accuracy of relevance vector machine classifiers. *J Glaucoma*. 2010;19:167–175.
38. Bizios D, Heijl A, Bengtsson B. Integration and fusion of standard automated perimetry and optical coherence tomography data for improved automated glaucoma diagnostics. *BMC Ophthalmol*. 2011;11:20.
39. Horn FK, Lämmer R, Mardin CY, et al. Combined evaluation of frequency doubling technology perimetry and scanning laser ophthalmoscopy for glaucoma detection using automated classification. *J Glaucoma*. 2012;21:27–34.
40. Boland MV, Quigley HA. Evaluation of a combined index of optic nerve structure and function for glaucoma diagnosis. *BMC Ophthalmol*. 2011;11:6.
41. Medeiros FA, Lisboa R, Weinreb RN, Girkin CA, Liebmann JM, Zangwill LM. A combined index of structure and function for staging glaucomatous damage. *Arch Ophthalmol*. 2012;130:1107.
42. Hood DC, Raza AS, De Moraes CGV, Johnson CA, Liebmann JM, Ritch R. The nature of macular damage in glaucoma as revealed by averaging optical coherence tomography data. *Transl Vis Sci Technol*. 2012;1:1–15.
43. Budenz DL, Rhee P, Feuer WJ, McSoley J, Johnson CA, Anderson DR. Sensitivity and specificity of the Swedish interactive threshold algorithm for glaucomatous visual field defects. *Ophthalmology*. 2002;109:1052–1058.
44. Yang Q, Reisman CA, Wang Z, et al. Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt Express*. 2010;18:21293–21307.
45. 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.
46. Katz J, Sommer A, Gaasterland DE, Anderson DR. Comparison of analytic algorithms glaucomatous visual field loss detecting. *Arch Ophthalmol*. 1991;109:1684–1689.
47. Obuchowski NA. Nonparametric analysis of clustered ROC curve data. *Biometrics*. 1997;53:567–578.
48. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
49. Garway-Heath DF, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43:2213–2220.
50. Harwerth RS, Wheat JL, Fredette MJ, Anderson DR. Linking structure and function in glaucoma. *Prog Retin Eye Res*. 2010;29:249–271.
51. Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci*. 2012;53:6939–6946.
52. Medeiros FA, Zangwill LM, Anderson DR, et al. Estimating the rate of retinal ganglion cell loss in glaucoma. *Am J Ophthalmol*. 2012;154:814–824.
53. Ng M, Sample PA, Pascual JP, et al. Comparison of visual field severity classification systems for glaucoma. *J Glaucoma*. 2012;21:551–561.
54. Wall M, Johnson CA, Kardon RH, Crabb DP. Use of a continuous probability scale to display visual field damage. *Arch Ophthalmol*. 2009;127:749–756.

Investigative Ophthalmology & Visual Science