

Between-Subject Variability in Healthy Eyes as a Primary Source of Structural–Functional Discordance in Patients With Glaucoma

Bright S. Ashimatey and William H. Swanson

Indiana University School of Optometry, Bloomington, Indiana, United States

Correspondence: Bright S. Ashimatey, Indiana University School of Optometry, 800 E. Atwater Avenue, Bloomington, IN 47405, USA; sbashima@indiana.edu.

Submitted: November 11, 2015
Accepted: January 10, 2016

Citation: Ashimatey BS, Swanson WH. Between-subject variability in healthy eyes as a primary source of structural–functional discordance in patients with glaucoma. *Invest Ophthalmol Vis Sci.* 2016;57:502–507.
DOI:10.1167/iovs.15-18633

PURPOSE. To test with an independent data set the finding that between-subject variability in healthy eyes is the primary source of structural–functional discordance in patients with glaucoma.

METHODS. Neuroretinal rim area, retinal nerve fiber layer thickness, and perimetric data were analyzed for one eye in each of 55 control subjects and for 245 right eyes of patients in the United Kingdom Glaucoma Treatment Study. Data were gathered with the Heidelberg Retina Tomograph (HRT), Stratus Optical Coherence Tomograph (OCT), and Humphrey Field Analyzer (HFA). Discordance was quantified as width of the limits of agreement from a Bland–Altman analysis of depth of defect. The ratio of variances (*F* test) for the patient and control groups was computed for comparisons of HFA–OCT, HFA–HRT, and OCT–HRT. Bonferroni adjustment required *P* less than 0.017 for statistical significance. The discordance in the patients was also quantified as the 95% prediction interval computed from the discordance in controls using the Hood–Kardon model for the HFA–OCT comparison.

RESULTS. The *F* ratio comparing discordance in patients and controls was 0.77, 1.43, and 1.32 for the HFA–OCT, HFA–HRT, and OCT–HRT comparisons with *P* values 0.88, 0.06, and 0.11, respectively. For the Hood–Kardon model, 4.7% of the patients had discordance outside the 95% prediction interval computed from the discordance in controls. Similar results were obtained when all comparisons were repeated for left eyes of patients.

CONCLUSIONS. These results confirm previous findings that between-subject variability in healthy eyes is the primary source of structural–functional discordance in patients with glaucoma, and extends this finding to a structural–structural comparison.

Keywords: structural–functional discordance, RNFL, depth of defect

Glaucoma is a group of chronic progressive neurodegenerative conditions that lead to characteristic patterns of visual field loss. Various mechanisms have been proposed to explain its pathophysiology; ultimately, the condition leads to loss of ganglion cell function and cell death.^{1–3} Clinical diagnostic techniques rely on perimetric tests measuring the functionality of the ganglion cells and imaging tests, which assess the structural presence of the ganglion cells (directly or indirectly).

Due to the differences in the methodology employed by structural and functional techniques, different artifacts influence them individually. Perimetric testing using Goldmann size III stimuli can be influenced by pupil size,⁴ peripheral refractive defocus,⁵ and the subject's internal criterion. Structural tests such as circumpapillary retinal nerve fiber layer (RNFL) thickness measured using optical coherence tomography (OCT) are influenced by disc area,⁶ axial length,^{6,7} and segmentation errors. Some factors such as ocular media opacity (introducing light scatter) and pupil size may impact both structural and functional techniques.

On the other hand, each technique does also have its distinctive advantage. For example, functional tests (like perimetry) present clinicians with a unique opportunity to have an idea about the visual function of the patient. Relative to

the perimetry, structural techniques have the advantage of being objectively acquired (although they require some operator dependent choices); they have also been shown to be fairly reproducible.^{8,9} The differences in the potential sources of artifact independently influencing structural and functional techniques of testing, as well as the relative advantage of one technique over the other provides a complementary effect when both techniques are used. Clinicians have therefore been encouraged to combine structural and functional tests to improve the certainties of diagnosis and to monitor progression.¹⁰

However, discordance in the structural–functional relationship has made the clinical usefulness of structural–functional comparison challenging. For example, some clinical trials found that functional tests show defects consistent with glaucomatous damage preceding structural defects while in other subjects structural damage precedes functional damage.^{10–12}

Hood and Kardon¹³ explored the sources of discordance between results of static automated perimetry (SAP) and circumpapillary RNFL thickness. They found that a simple linear model could relate the two types of measurements and that the characteristics of the structural–functional discordance in patients could be predicted from the characteristics of the structural–functional relation in controls.¹⁴ Swanson et al.¹⁵

found that the limits of agreement between structural and functional measures in patients were very similar to the between-subject variability in healthy controls, as did a study comparing temporal disc rim area to macular perimetric sensitivity.¹⁶

In this paper, we integrated the function proposed by Hood et al.¹³ (relating RNFL thickness to perimetric sensitivity) with the Bland-Altman plots used by Swanson et al.¹⁵ to analyze an independent data set (gathered by another laboratory) to test the finding that between-subject variability in healthy eyes is the primary source of structural-functional discordance in patients with glaucoma. The motivation to maintain the analysis approach used by the previous studies (in which the data were collected and analysis techniques selected by the respective investigators) on an independent data set was to attenuate any unintentional bias in the design and choice of analysis technique that may have influenced the results of the previous studies.^{17,18} We further extended the hypothesis to include structural-structural discordance.

METHODS

Subjects and Devices

The patient data analyzed in this study were obtained from the United Kingdom Glaucoma Treatment Study (UKGTS).¹⁹ The UKGTS was a randomized, double-masked, placebo-controlled, multicenter treatment trial for open-angle glaucoma. The study recruited newly diagnosed (untreated) glaucoma cases with glaucomatous visual field defects consistent with optic nerve head changes and with open angles on gonioscopy. The study required subjects to have a visual acuity of at least 20/40. Subjects with moderately advanced visual field loss (mean deviation worse than -10 dB in the better eye or worse than -16 dB in the other eye) were excluded from the study and subjects with IOP greater than 35 mm Hg on two consecutive visits were also excluded from the study. Subjects with lens opacity greater than P1 (on the Lens Opacity Classification System III grading) and those with other ocular comorbidities such as diabetic retinopathy were also excluded from the study. Further details on the inclusion and exclusion and other types of tests included in the UKGTS study can be found elsewhere.¹⁹

We analyzed test results for perimetry, neuroretinal rim area, and circumpapillary RNFL thickness. The visual field sensitivity measures were obtained using the Goldmann size III stimulus on the Humphrey Field Analyzer (HFA) II or II-i, (Carl Zeiss Meditec, Dublin, CA, USA) with the SITA 24-2 program. There were HFA data on 490 subjects. Humphrey Field Analyzer test results with fixation losses greater than 20% or false-positive rate greater than 15% were excluded.¹⁹ The exclusion criteria did not exempt any subjects in the original UKGTS dataset from our study.

The rim area measurements were obtained using the HRT-3 (software version 3.0.60; Heidelberg Engineering, Heidelberg, Germany). When the mean pixel SD of the HRT image was greater than $40 \mu\text{m}$, the HRT data were regarded as unreliable and were excluded.¹⁹ Rim area data were collected on 482 right eyes and 481 left eyes. After applying the exclusion criteria we were left with HRT data for 482 and 480 right and left eyes, respectively.

The RNFL thicknesses were acquired using the Stratus OCT (software version 5.0; Carl Zeiss Meditec). The fast RNFL thickness scanning protocol was used (3.4-mm diameter circle centered on the disc). The OCT data were screened to exclude scans with signal strength less than 7. For our analysis we also excluded OCT scans with errors: with an error message, a pixel

lower than $10 \mu\text{m}$, or the range across repeated scans greater than $15 \mu\text{m}$.⁸ There were OCT data on 289 right eyes and 277 left eyes in the original UKGTS data set. After applying our inclusion and exclusion criteria, we were left with RNFL thickness measures for 284 and 261 right and left eyes, respectively.

We matched the HRT, OCT, and HFA data by date and eye for the last visit of each patient. We required that all three tests were acquired on the same day, in order to avoid artifacts from progression. We were able to successfully match the three tests for 245 right eyes and 223 left eyes.

The 55 controls were from a pool of 62 whose OCT and HFA data had already been published in Swanson et al.¹⁵ Only subjects who also had reliable data on the HRT were included in this study. The criteria for excluding unreliable data and matching reliable data were the same for the controls as for the patients. The instruments used in acquiring the control data were comparable with those used in the UKGTS study.

In the original studies it was reported that the methods used in gathering the data were in accordance with the Declaration of Helsinki. Informed consent was obtained from subjects after explanation of the nature and goals of the study, before testing began.^{15,19}

Analysis

In our primary analysis, we analyzed data from the right eyes in an effort to replicate the finding that between-subject variability in controls is the primary source of structural-functional (and structural-structural) discordance in patients with glaucoma. We analyzed data from the left eyes in a secondary confirming analysis, in the same way as for the right eye.

Global measures from HFA, OCT, and HRT were used to reduce the impact of between-subject variations in nerve fiber projection to optic disc sectors.^{20,21} For the HFA, global visual field sensitivity was calculated by first converting the sensitivity at the various test locations to a linear scale before averaging.^{15,16,22} We then converted the average linear sensitivities back to a log scale. Two visual field points, at and just above the blind spot, were excluded from this index.

In order to overcome the impact of the differences in the metric of measurement of the various clinical measuring techniques on our analysis, we computed depth of defect (difference from mean normal) for data from each device. We computed mean normal for each testing device from controls. Depth of defect was used as a measure of the severity of the glaucomatous damage. We calculated depth of defect as log difference from mean of the control group.¹⁵

Our analysis was centered around the limits of agreement between two measures as described by Bland and Altman²³; the difference in depth of defect for two tests was plotted against their average depth of defect. The difference in defect depth may vary with the severity of damage, so we calculated the residuals from linear regression of difference versus mean. The absolute values of the residuals were plotted against the average of the two measures and then fitted with a regression to estimate whether the limits of agreement varied with severity of damage. In our analysis, the limits of agreement did not vary with the average depth of defect and so we did not proceed with the other methods described in Section 3.3 of the Bland and Altman article²³ (that describe computing limits of agreement that vary with the average depth of defect). We therefore computed the limits of agreement as $1.96 \times \text{SD}$ of the residuals.²³

We made structural-functional comparisons (perimetry versus RNFL thickness, perimetry versus rim area) and structural-structural comparisons (RNFL thickness versus rim

TABLE. The SD in Log Units of the Residuals Around the Regression Line on a Bland-Altman Plot Comparing Two Clinical Device Outputs in Patients

Global Comparison	Patient OD	Patient OS	Control	F Ratio (OD)	
				Patient/Control	P Value
SAP OCT	0.117	0.118	0.133	0.774	0.88
SAP HRT	0.190	0.186	0.159	1.428	0.06
OCT HRT	0.108	0.113	0.094	1.320	0.11

The *F* ratio for patients versus controls was considered significant if $F(244, 54) > 1.63$, $P < 0.017$.

area). It has been found, that relative to the HFA, depth of defect estimated from structural measures is limited by a floor (due to the nonneural component contained in the structural measures).^{15,24} This floor introduces a potential artifact to the Bland-Altman analysis when two measures (containing different amounts of nonneural components) are being compared. A floor of -0.5 was used in Swanson et al.¹⁵ In the Bland-Altman plot comparing HFA and OCT (and OCT and HRT) we also used a floor of -0.5 . For lack of a study that established a floor for HRT measures and for the sake of simplicity we also used a floor of -0.5 in the HFA-HRT comparison.

In the primary analysis, *F* tests were used to compare variability in patients and controls. Statistical significance was set to a *P* value of 0.017 (a Bonferroni adjustment) because three *F* values (two structural-functional comparisons and one structural-structural comparison) were assessed. For a *P* value of 0.017 and the sample sizes in the patient and control groups, this study had a power of 80% to detect a 2-fold difference in variance. For these sample sizes, the critical value will be $F(244, 54) = 1.63$, *P* less than 0.017.

To compare our findings with those of Hood et al.,¹⁴ we applied the Hood-Kardon model (which predicts the relationship between RNFL thickness and perimetry) to our data. The Hood-Kardon model assumes that the nonneural component of the RNFL thickness contributes approximately 33% of the measured thickness before the onset of the glaucomatous damage. It also assumes that RNFL thickness does not increase beyond mean normal even when visual field sensitivity is above mean normal. To allow for a direct comparison with our analyses we made similar assumptions as Hood et al.¹⁴ (although our analysis was focused on global indexes and that for Hood et al. was for superior temporal and inferior temporal disc sectors). The model predicts that approximately 95% of the patient data points should be within the prediction interval based on variability in the control data.

RESULTS

For 245 right eyes and 223 left eyes, the patients had mean (\pm SD) age of 65.5 (\pm 11.1) and 65.6 (\pm 11.2) years, respectively. The 55 controls had mean (\pm SD) age of 63.0 (\pm 9.9) years.

For the controls, mean (\pm SD) perimetric sensitivity was 0.49 (\pm 0.14) log contrast sensitivity, mean RNFL thickness was 1.99 (\pm 0.04) log μ m and mean rim area 0.18 (\pm 0.04) log mm². The mean control HRT disc area was 1.93 (\pm 0.41) mm² and that for the patients was 2.02 (\pm 0.40) mm² and 2.00 (\pm 0.47) mm² for OD and OS, respectively.

The Table shows the SDs of the residuals when depths of defect measured by two devices were compared on a Bland-Altman plot. None of the planned comparisons reached the criterion for significance.

The left panels in Figure 1 show plots of discordance against average depth of defect comparing perimetry and RNFL thickness. It also features the Hood-Kardon model modified for the plot. In the right and left eyes, 4.7% and 5.4%, respectively, of the patient data fell outside the 95% limits based on the dispersion in the control population.

The right panels show the same data with a Bland-Altman analysis. The slope of the regression line on the Bland-Altman plot was 0.87 (\pm SE = 0.07) and 0.88 (\pm SE = 0.06) for right and left eyes, respectively; the intercepts of the regression lines were 0.07 (\pm SE = 0.01) and 0.06 (\pm SE = 0.01) for left and right eyes respectively.

A second structural-functional comparison is shown in Figure 2, perimetry versus rim area. The slope of the regression line was 0.44 (\pm SE = 0.10) for right eyes and 0.59 (\pm SE = 0.10) for left eyes; the intercepts were 0.13 (\pm SE = 0.02) and 0.10 (\pm SE = 0.02) for the right and left eyes, respectively.

A structural-structural comparison is shown in Figure 3, rim area versus RNFL thickness. The slope of the regression line on the Bland-Altman plot was -0.56 (\pm SE = 0.07) for right eyes and -0.47 (\pm SE = 0.08) for left eyes; the intercept was 0.00 (\pm SE = 0.01) for both right and left eyes.

DISCUSSION

In this study, we replicate the finding that normal between-subject variability is the primary source of structural-functional discordance in patients with glaucoma.^{14,15} We compared the variability of the discordance in the patients with the variability of the discordance in the controls and found them to be similar. We extended the analysis to include a structural-structural comparison and found a similar result; variability in the control group was similar to variability in the patient group.

We also analyzed the perimetry versus RNFL thickness data using an extension of the Hood-Kardon model,¹⁴ and found that approximately 95% of the patient data fell within the 95% prediction interval computed from the control population.

In the HFA-HRT comparison however, we noticed a relatively larger variance as compared with the HFA-OCT and OCT-HRT comparisons. We were therefore interested in investigating the impact of disc area on the HFA-HRT comparison using the Moorfields Regression Analysis.²⁵ Compared with our original analysis, adjusting for the rim area (using disc area and age) in the HFA-HRT comparison increased the variance in the right of patients by 1% and decreased the variability in the control group by 3%. Overall, adjusting for the rim area (using the disc size and age) did not result in a significant improvement in the HFA-HRT comparison. This is not however surprising because we did not see any consistent distribution in depth of defect based on disc size on our plots. This increased variability may be due to some image acquisition factors like the operator tracing the area of the disc and the limit imposed on the rim area by the disc size.

In the comparison of perimetry and RNFL thickness, the regression on the Bland-Altman plot yielded a positive slope, meaning that as the average defect depth increased, perimetry tended to give increasingly deeper defects than did RNFL thickness. This is consistent with a nonneural component of RNFL thickness.^{13,24} The comparison of perimetry and rim area also found positive slopes, but they were not as steep as for perimetry versus RNFL thickness. This is consistent with the proposal by Shafi et al.¹⁶ of a greater nonneural component for

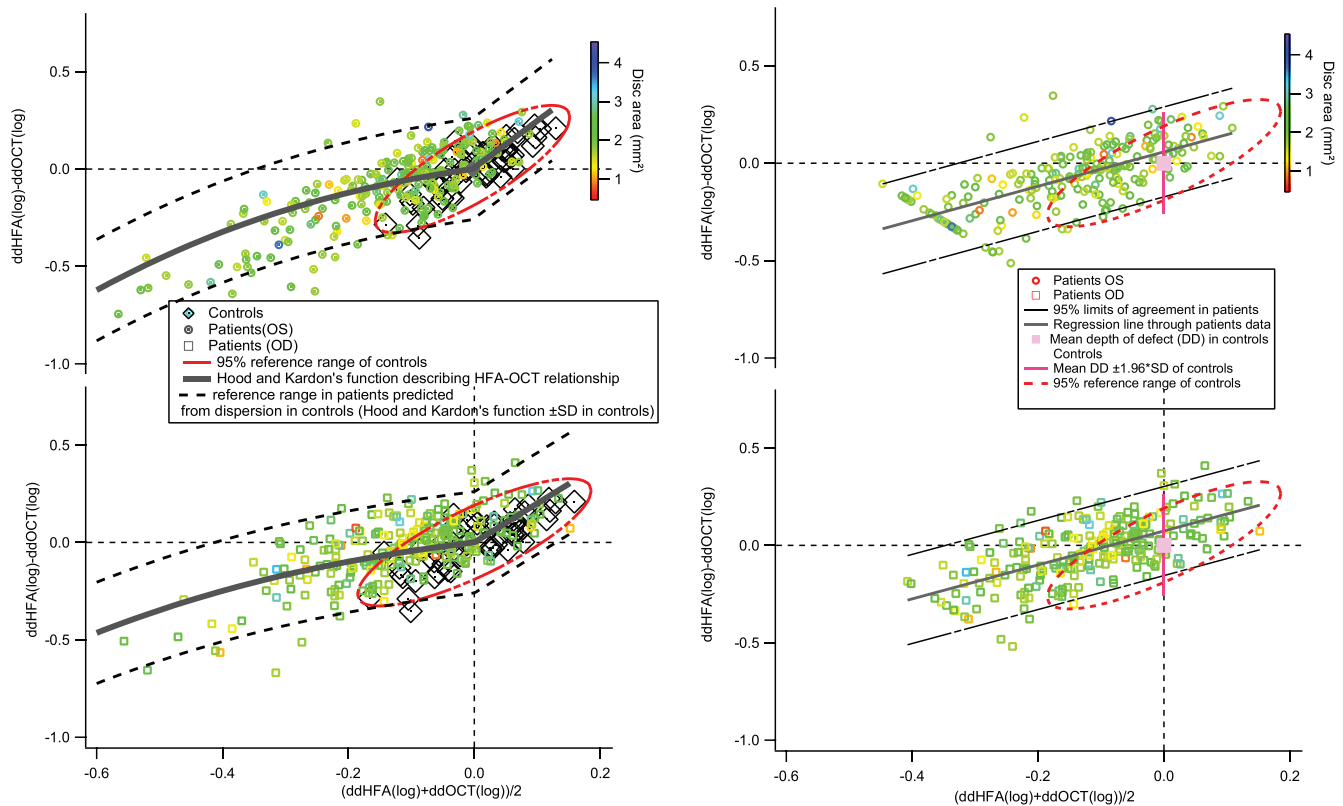


FIGURE 1. Bland-Altman plots comparing perimetric sensitivity and RNFL thickness. The *upper plots* show data for left eyes and the *lower plots* show data for right eyes. The *left panels* show limits comparable to the Hood-Kardon model,¹⁴ and the *right panels* show the 95% limits of agreement comparable to the limits of Swanson et al.¹⁵ The color of the marker indicates the disc area (in mm²). The *red ellipses* on the *left* show the 95% prediction interval for the controls; the *black broken lines* show the 95% prediction interval for patients.

RNFL thickness than for rim area. The direct comparison of rim area and RNFL thickness in Figure 3 also supports a greater nonneural component for RNFL thickness than for rim area. We recomputed the discordance between the rim area and RNFL thickness measures using a series of other floors to investigate whether the finding on the relative amount of nonneural component contained in the rim area and RNFL thickness measures was impacted by our choice of floor (-0.5 log unit). We found that although the magnitude of slope of the Bland-Altman plot changed depending on the floor, its (negative) direction and implication did not change.

The intercept of the Bland-Altman regression line was nonzero for perimetry versus RNFL thickness and for perimetry versus rim area, but not for rim area versus RNFL thickness. This is an indication that at the early onset of glaucomatous damage, structural measures estimate deeper defects than perimetry. A recent study¹⁵ from our research group found a similar result when comparing SAP with new forms of perimetry resistant to peripheral defocus: nonzero intercepts for Bland-Altman regression line indicating that SAP defects were not as deep in mild defects and became deeper in more severe defects.

Hood et al.¹⁴ pointed out that when two tests are being compared, the test with a smaller SD in the control population would require less damage to reach statistical significance and be flagged as glaucomatous than the test with a larger SD. Drawing from the statistics of identifying which loss constitutes glaucomatous damage and which loss does not, the OCT is likely to be more sensitive than the HFA because it had a smaller SD in the control group than the HFA. In an exploratory analysis, we calculated the fifth percentile for the RNFL thickness and visual field measures from the control

population. We set the value computed as threshold and identified subjects as showing glaucomatous damage when their RNFL thickness (or visual field sensitivity) fell below the threshold. The OCT identified more subjects than the HFA as showing glaucomatous damage consistent with the finding of Hood et al.¹⁴ As proposed by Hood et al.,¹⁴ this may explain why the OCT appears to be more sensitive to early damage and suggests a different reasoning (as to why the OCT may flag mild defects earlier than the HFA) than “preperimetric” glaucoma concept (which suggests that a large percentage of ganglion cells are lost before the loss is detected on perimetric testing).

From the comparisons, we also noticed that the limits of agreement of OCT-HFA comparison (a structural-functional comparison) in the patient group were similar to the limits of agreement of the OCT-HRT comparison (a structural-structural comparison). This provides further evidence that the discordance seen in comparing structural measures to functional measures may not be necessarily due to the fact that one method assesses glaucomatous damage through structural loss while the other uses functional loss.

We investigated the characteristics of disc parameters (such as cup-to-disc volume, cup volume, cup-disc ratio, etc.) reported by the HRT in the control subjects with the most extreme discordance for each comparison to determine whether there are any specific structural patterns reported by the HRT that may explain the variability of the discordance in controls. For the HFA-OCT comparison we did not find any significant difference between subjects with the most extreme positive discordance (greater HFA depth of defect than OCT) and those with the most extreme negative discordance (greater OCT depth of defect than HFA). For example, the mean disc

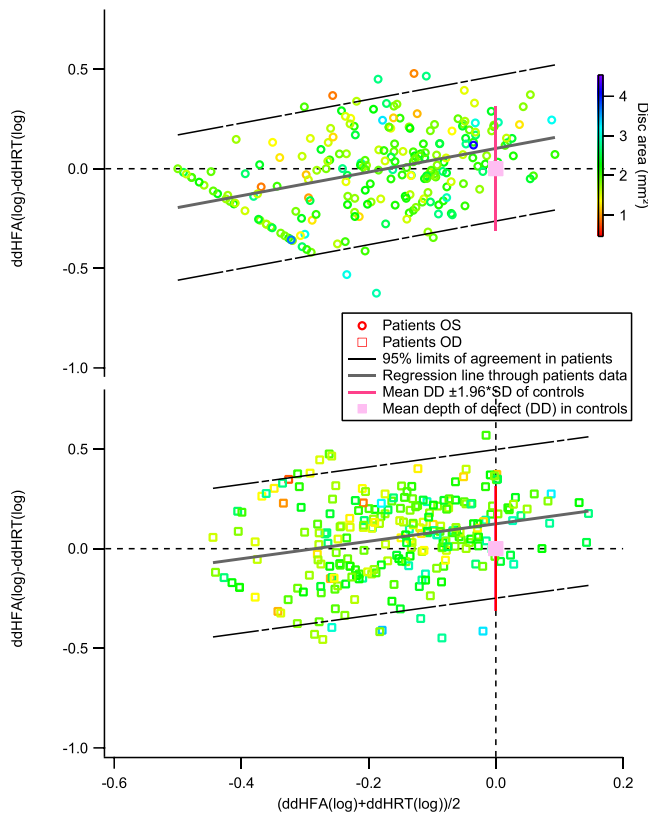


FIGURE 2. Bland-Altman plots comparing perimetric sensitivity and rim area. The black broken lines show the 95% limits of agreement for the patient data. The upper plot shows data for left eyes and the lower plot shows data for right eyes.

area of controls with the highest 10% positive discordance (greatest HFA depth of defect than OCT) and those with the highest 10% negative discordance (greatest OCT depth of defect than HFA) were 1.99 mm² (SD = 0.15) and 1.82 mm² (SD = 0.30). Similar results were found in an HFA-HRT comparison.

However, in the HRT-OCT comparison we noticed that the maximum cup depth, mean cup depth, horizontal and vertical cup-to-disc area ratio were significantly higher (at $P < 0.05$) in subjects with the most extreme positive discordance (greater HRT depth of defect than OCT) than those with the most extreme negative discordance (greater OCT depth of defect than HRT). In a consistent manner, the rim disc area ratio and rim volume were significantly smaller in subjects with the most extreme positive discordance than subjects with the most extreme negative discordance. Thus, adjusting for maximum cup depth, mean cup depth, or horizontal and vertical cup-to-disc ratio may reduce the discordance in the healthy population for OCT-HRT comparison.

We also noticed that the majority of subjects with the most extreme discordance in the HFA-OCT comparison were not the same as the subjects with the most extreme discordance in the HFA-HRT comparison. Thus, the few subjects who have extreme discordance on both the HFA-OCT and HRT-OCT comparison may have some underlying structural pattern common to both the HRT and OCT that causes the structural depth of defect to differ from the functional depth of defect in a consistent manner. However, it is not clear whether the subjects who have higher depth of defect on the HFA-OCT comparison but not on the HFA-HRT measure or vice versa have any specific structural pattern. The trend in this group is

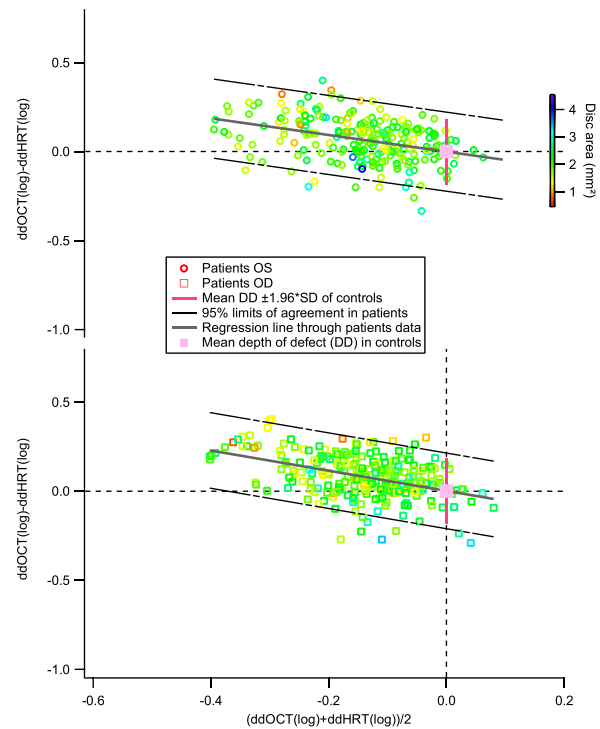


FIGURE 3. A Bland-Altman plot comparing the RNFL thickness and rim area. As in Figures 1 and 2, the black broken lines show the 95% limits of agreement for the patient data. The upper plot shows data for left eyes and the lower plot shows data for right eyes.

consistent with the HRT-OCT discordance and suggests a multidimensional source of structural-functional discordance. Thus, the discordance observed in a subject is not only a property of the structural and functional ganglion cell integrity but is also influenced by the region on the retina where the structural integrity is being assessed. In conclusion, we confirmed the finding of prior studies that between-subject variability in healthy eyes is the primary source of structural-functional discordance in patients, and extended this to structural-structural comparisons. Developing techniques with reduced between-subject variability in the healthy population may be helpful in improving the discordance in comparing two techniques.

There have been proposals on how to reduce between-subject variability in controls. Patel et al.⁷ proposed using RNFL volume instead of RNFL thickness in structural measures; elsewhere, there are proposals on how to reduce the impact of prereceptor factors on perimetric sensitivity.^{4,5} These proposals are geared toward reducing the normal between-subject variability and may consequently yield more sensitive measures with reduced structural-functional (or structural-structural) discordance when two measures are compared. Improved techniques for structural and functional measures yielding improved structural-functional discordance may not only be necessary to complement a clinician's diagnosis but will also open discussions for the use of perimetry for testing ganglion cell dysfunction.

To address the challenge of structural-functional discordance holistically, other factors that need to be addressed include the spatial sampling of perimetric measures and the difficulty in spatially mapping structural abnormality to functional abnormality using the structural-functional maps. While the structural measures (acquired in the region of the disc) assess all the ganglion cell axons present in the retina, perimetric testing with the 24-2 protocol samples 54 testing

locations on the retina. This predisposes the perimetric tests to miss localized wedge defects that may fall between perimetric testing locations while being identified by the structural measures.

Due to the fact that quite a number of the structural measures concentrated on determining glaucomatous structural abnormality at the region of the disc, most studies are faced with the challenge of choosing a structural-functional map to be able to relate structural abnormality to functional abnormality. While our study concentrated on using global measures to overcome this challenge, it is clear that the wide range of variability in the trajectory of the retinal nerve fiber bundles to the disc is a potential source of structural-functional discordance. Ballae Ganeshrao et al.²⁶ have demonstrated that accurately mapping structural measures to functional measures spatially accounting for the anatomical differences in these trajectories yield improved structural-functional discordance.²⁶

The developments in imaging techniques that allow the visualization of the nerve fiber bundles en face^{27,28} are an attractive approach to overcoming the spatial challenges implicated in the structural-functional discordance and may provide more insight into the true nature of the structural-functional relationship.

Acknowledgments

The authors thank Victor Malinovsky, OD, and Brett King, OD, for their comments on the manuscript. They also thank the UKGTS team for providing them with deidentified data from the UKGTS study.

Supported by the National Institutes of Health Grants R01EY007716 and R01EY024542 (WHS; Bethesda, MD, USA).

Disclosure: **B.S. Ashimatey**, None; **W.H. Swanson**, Heidelberg Engineering (C), Carl Zeiss Meditec (C, R)

References

- Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901-1911.
- Whitmore AV, Libby RT, John SW. Glaucoma: thinking in new ways—a role for autonomous axonal self-destruction and other compartmentalised processes? *Prog Retin Eye Res*. 2005;24:639-662.
- Nickells RW. The cell and molecular biology of glaucoma: mechanisms of retinal ganglion cell death. *Invest Ophthalmol Vis Sci*. 2012;53:2476-2481.
- Swanson WH, Dul MW, Horner DG, Liu T, Tran I. Assessing spatial and temporal properties of perimetric stimuli for resistance to clinical variations in retinal illumination. *Invest Ophthalmol Vis Sci*. 2014;55:353-359.
- Horner DG, Dul MW, Swanson WH, Liu T, Tran I. Blur-resistant perimetric stimuli. *Optom Vis Sci*. 2013;90:466-474.
- Budenz DL, Anderson DR, Varma R, et al. Determinants of normal retinal nerve fiber layer thickness measured by Stratus OCT. *Ophthalmology*. 2007;114:1046-1052.
- Patel NB, Luo X, Wheat JL, Harwerth RS. Retinal nerve fiber layer assessment: area versus thickness measurements from elliptical scans centered on the optic nerve. *Invest Ophthalmol Vis Sci*. 2011;52:2477-2489.
- Budenz DL, Chang RT, Huang X, Knighton RW, Tielsch JM. Reproducibility of retinal nerve fiber thickness measurements using the stratus OCT in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2005;46:2440-2443.
- Ghasia FF, El-Dairi M, Freedman SF, Rajani A, Asrani S. Reproducibility of spectral-domain optical coherence tomography measurements in adult and pediatric glaucoma. *J Glaucoma*. 2015;24:55-63.
- Malik R, Swanson WH, Garway-Heath DE 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clin Experiment Ophthalmol*. 2012;40:369-380.
- Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120:701-713, discussion 829-730.
- Miglior S, Zeyen T, Pfeiffer N, Cunha-Vaz J, Torri V, Adamsons I. Results of the European Glaucoma Prevention Study. *Ophthalmology*. 2005;112:366-375.
- Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. 2007;26:688-710.
- 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.
- Swanson WH, Malinovsky VE, Dul MW, et al. Contrast sensitivity perimetry and clinical measures of glaucomatous damage. *Optom Vis Sci*. 2014;91:1302-1311.
- Shafi A, Swanson WH, Dul MW. Structure and function in patients with glaucomatous defects near fixation. *Optom Vis Sci*. 2011;88:130-139.
- Iorns E, Chong C. New forms of checks and balances are needed to improve research integrity. *F1000Res*. 2014;3:119.
- Let's think about cognitive bias. *Nature*. 2015;526:163.
- Garway-Heath DE, Lascaratos G, Bunce C, Crabb DP, Russell RA, Shah A. The United Kingdom Glaucoma Treatment Study: a multicenter, randomized, placebo-controlled clinical trial: design and methodology. *Ophthalmology*. 2013;120:68-76.
- Lamparter J, Russell RA, Zhu H, et al. The influence of intersubject variability in ocular anatomical variables on the mapping of retinal locations to the retinal nerve fiber layer and optic nerve head. *Invest Ophthalmol Vis Sci*. 2013;54:6074-6082.
- Janssonius NM, Schiefer J, Nevalainen J, Paetzold J, Schiefer U. A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position. *Exp Eye Res*. 2012;105:70-78.
- Hot A, Dul MW, Swanson WH. Development and evaluation of a contrast sensitivity perimetry test for patients with glaucoma. *Invest Ophthalmol Vis Sci*. 2008;49:3049-3057.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135-160.
- Sihota R, Sony P, Gupta V, Dada T, Singh R. Diagnostic capability of optical coherence tomography in evaluating the degree of glaucomatous retinal nerve fiber damage. *Invest Ophthalmol Vis Sci*. 2006;47:2006-2010.
- Wollstein G, Garway-Heath DE, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology*. 1998;105:1557-1563.
- Ballae Ganeshrao S, Turpin A, Denniss J, McKendrick AM. Enhancing structure-function correlations in glaucoma with customized spatial mapping. *Ophthalmology*. 2015;122:1695-1705.
- Huang G, Gast TJ, Burns SA. In vivo adaptive optics imaging of the temporal raphe and its relationship to the optic disc and fovea in the human retina. *Invest Ophthalmol Vis Sci*. 2014;55:5952-5961.
- Chauhan BC, Sharpe GP, Hutchison DM. Imaging of the temporal raphe with optical coherence tomography. *Ophthalmology*. 2014;121:2287-2288.