tvst

Glaucoma

A Metascore of Multiple Imaging Methods to Measure Long-Term Glaucoma Structural Progression

Agustina De Gainza[1], Esteban Morales[2], Alessandro Rabiolo[1], Fei Yu[3], Abdelmonem A. Afifi[4], Kouros Nouri-Mahdavi[2], and Joseph Caprioli[1]

[1] University of California, Los Angeles, Los Angeles, CA, USA
[2] Jules Stein Eye Institute, Los Angeles, CA, USA
[3] UCLA, Los Angeles, CA, USA
[4] UCLA, School of Public Health, Los Angeles, CA, USA

Correspondence: Joseph Caprioli, University of California, Los Angeles, 100 Stein Plaza, Los Angeles, CA 90095, USA.
e-mail: caprioli@jsei.ucla.edu

Received: June 15, 2021
Accepted: August 12, 2022
Published: September 21, 2022

Keywords: long-term structural progression; score for glaucoma progression; combination of structural imaging devices for glaucoma progression

Citation: De Gainza A, Morales E, Rabiolo A, Yu F, Afifi AA, Nouri-Mahdavi K, Caprioli J. A metascore of multiple imaging methods to measure long-term glaucoma structural progression. Transl Vis Sci Technol. 2022;11(9):15, https://doi.org/10.1167/tvst.11.9.15

Purpose: To develop a structural metascore (SMS) that combines measurements from different devices and expresses them on a single scale to facilitate their long-term analysis.

Methods: Three structural measurements (Heidelberg Retina Tomograph II [HRT] rim area, HD-Cirrus optical coherence tomography [OCT] average retinal nerve fiber layer [RNFL] thickness, Spectralis OCT RNFL global thickness) were normalized on a scale of 0 to 100 and converted to a reference value. The resultant metascores were plotted against time. SMS performance was evaluated to predict future values (internal validation), and correlations between the average grades assigned by three clinicians were compared with the SMS slopes (external validation).

Results: The linear regression fit with the variance approach, and adjustment to a Spectralis equivalent was the best-performing approach; this was denominated metascore. Plots were created for 3416 eyes of 1824 patients. The average baseline age (± standard deviation) was 69.8 (±13.9), mean follow-up was 11.6 (±4.7) years, and mean number of structural scans per eye was 10.0 (±4.7). The mean numbers of scans per device were 3.8 (±2.5), 5.0 (±2.9), and 1.3 (±3.0) for HRT, Cirrus, and Spectralis, respectively. The metascore slopes' median was −0.3 (interquartile range 1.1). Correlations between the average grades assigned by the three clinicians and the metascore slopes were −0.51, −0.49, and −0.69 for the first (structural measurement printouts alone), second (metascore plots alone), and third (printouts + metascore plots) series of gradings, respectively. The average absolute predictive ability was 7.63/100 (whereas 100 = entire normalized scale).

Conclusions: We report a method that converts Cirrus global RNFL and HRT global rim area normalized measurements to Spectralis global RNFL equivalent values to facilitate long-term structural follow-up.

Translational Relevance: Because glaucoma changes usually occur slowly, patients are often examined with different instruments during their follow-up, a method that "unifies" structural measurements provided by different devices, which could assist patients' longitudinal structural follow-up.

Introduction

Glaucoma is a chronic, progressive optic neuropathy that often requires life-long clinical assessment and treatment to prevent visual loss. The clinical judgement of an experienced ophthalmologist is often assisted by the interpretation of ancillary tests to guide diagnosis and treatment. Standard automated perimetry has remained the mainstay of functional testing,[1] but, in recent years, imaging techniques such as confocal scanning laser ophthalmoscopy (CSLO) and optical coherence tomography (OCT), among others, have been sequentially introduced to aid the evaluation

of glaucoma-related structural changes, aiming to provide more reproducible and objective measures of the optic nerve head (ONH) and the peripapillary retinal nerve fiber layer (RNFL).[2–7] Owing to this increasing number of commercially available structural-measuring devices, and the speed at which they are being introduced into clinical practice, patients are often examined with different instruments during their follow-up.

The present study was designed to build and validate a structural "metascore" that provides comparable measurements derived from different imaging methods and expresses them on a single normalized scale. This would facilitate long-term sequential analysis and interpretation, with the goal of improving the detection of long-term structural progression to inform appropriate treatments.

Methods

The data used to develop the structural metascore was exported from the three structural devices (Heidelberg Retina Tomograph [HRT], Cirrus OCT, and Spectralis OCT) used in the Glaucoma Division of the Stein Eye Institute, University of California, Los Angeles (UCLA). They included structural scans acquired from 1993 to 2020. This study adhered to the tenets of the Declaration of Helsinki, was approved by the UCLA Human Research Protection Program, and conformed to Health Insurance Portability and Accountability Act policies.

Inclusion criteria were clinical diagnosis of chronic glaucoma (primary open-angle glaucoma, chronic angle-closure glaucoma, uveitic, pseudoexfoliative, pigmentary, steroid-induced, traumatic), and age ≥18 years. Exclusion criteria were any other causes for optic nerve or retinal abnormalities potentially affecting structural or functional status, such as proliferative diabetic retinopathy, central retinal vein occlusion, retinal detachment, and exudative age-related macular degeneration. Visual fields were performed with Humphrey Field Analyzer's Swedish Interactive Thresholding Algorithm Standard 24-2 and 30-2 strategies and a size III white stimulus (Carl Zeiss Meditec, Inc., Dublin, CA, USA). Structural devices included in this study were Heidelberg Retina Tomograph II (HRT; Heidelberg Engineering, Heidelberg, Germany), Cirrus HD-OCT (Carl Zeiss Meditec, Inc.), and Spectralis OCT (Heidelberg Engineering). Only good-quality scans were included, defined as HRT standard deviation <50 µm, Cirrus HD-OCT signal strength ≥6, and Spectralis OCT quality ≥18. Normal subjects were recruited from the research database in the Glaucoma Division, Stein Eye Institute. The enrolled normal subjects were required to have open angles, corrected visual acuity of 20/25 or better, and normal eye examination results including normal visual fields and normal ophthalmoscopic appearance of the optic nerve head.

The process of generating a metascore for each eye is described briefly as follows (Fig. 1) and is detailed below:

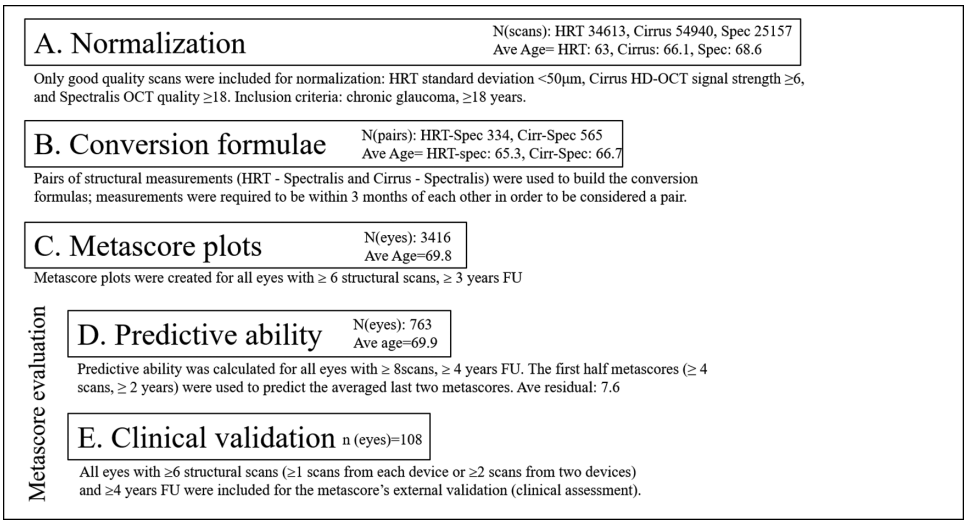

Figure 1. Summary of how subset groups were created for each step of metascore development. (**A**) Data normalization: transforming the units of measurement of each device to the same scale. (**B**) Conversion formulae: measurements provided by HRT and Cirrus were converted to a Spectralis equivalent. The resulting measurements are called metascores. (**C**) Metascore plots were created for all eyes with at least 6 structural scans and at least 3 years of follow-up. (**D**, **E**) Metascore performance was evaluated in its accuracy in predicting itself over time (predictive ability, or internal validation), and how it compared to clinical assessment (clinical, or external validation).

Table 1. Floor and Ceiling Values for Each of the Three Devices Used in This Study (Variance Approach)

Variance Approach	HRT Rim Area (mm^2)	Cirrus (RNFL Average Thickness (µm)	Spectralis Global RNFL Thickness (µm)
Floor (0) = Mean (structural measurement) − 3*SD	0.07	32.47	11.37
Ceiling value (100) = mean (structural measurement) + 3*SD	2.52	122.09	127.94

1. ***Data normalization:*** Different structural measurements (HRT rim area, HD-Cirrus OCT average RNFL thickness, and Spectralis OCT RNFL global thickness) were normalized to the same scale of 0 (worse) to 100 (better). We tested two different techniques for normalization: the variance and the dynamic range approach.

2. ***Conversion formulae:*** Measurements provided by HRT, Cirrus, and Spectralis were converted to a reference device equivalent (either Cirrus or Spectralis). The resulting measurements are called metascores. Goodness of fit for the metascores calculated with three different statistical methods (univariable linear regression, calibration equation and Bland-Altman plots) was calculated with the root mean squared error for the different approaches (Table 1).

3. ***Metascore plots:*** Metascores (vertical axis) were plotted for each included eye against time in years (horizontal axis) to provide a graphical tool. A prediction interval (range in which a future individual observation will fall, based on the model estimates) was calculated for all the metascore slopes.

4. ***Metascore evaluation:*** Metascore performance was evaluated in two ways: its accuracy in predicting itself over time (predictive ability, or internal validation), and how it compares to clinical assessment (clinical, or external validation). The methodology and number of images used is summarized in Figure 4.

Data Normalization

Data normalization was performed to transform the units of measurement of each device to the same scale. We chose the units of measurement which we considered optimal for each instrument. Both clinical practice and literature support using RNFL thickness as the main global parameter to evaluate structural integrity and progression with OCT, mainly because of its reliability and reproducibility.[8] HRT, on the other hand, can only measure height values of the retinal surface with respect to a reference plane, but it cannot distinguish between different retinal layers[9]; hence, we chose rim area as its most robust global structural measurement. Two normalization methods, variance and dynamic range approaches, were evaluated.

Variance Approach

We rescaled the structural measurements provided by HRT, Cirrus OCT, and Spectralis OCT (rim area, average RNFL thickness and global RNFL thickness, respectively) to fit values between 0 (worst) and 100 (best). We did so with the mean ± 3 standard deviation (SD) of the included data for each device, as follows:

$$\text{Min}\,(0) = \text{mean}\,(\text{structuralmeasurement}) - 3 * \text{SD}$$
$$\text{Max}\,(100) = \text{mean}\,(\text{structuralmeasurement}) + 3 * \text{SD}$$

Original values that were ≤(mean − 3*SD) or ≥(mean + 3*SD) were set to the minimum (0) or maximum (100) of the normalized scale, respectively.

Dynamic Range Approach

The dynamic range of global rim area for HRT and RNFL thickness for OCT was determined by subtracting the average residual layer thickness (floor) of glaucoma patients from the average thickness +1 SD of normal subjects. Normal subjects were recruited from the research database in the Glaucoma Division, Stein Eye Institute. The enrolled normal subjects were required to have open angles, corrected visual acuity of 20/25 or better, and normal eye examination results including normal visual fields and normal ophthalmoscopic appearance of the optic nerve head. The floor value (0), was determined by the regressed y-intercept (using the broken stick model)[10] of structural measurements against VFI (Fig. 2).The ceiling value (100) corresponds to the average thickness of normal subjects +1 SD (Table 2).

Conversion Formulas

We selected pairs of structural measurements (HRT-Spectralis and Cirrus-Spectralis) to build the conversion formulas; measurements were required to be within three months of each other to be considered a pair. We compared three different

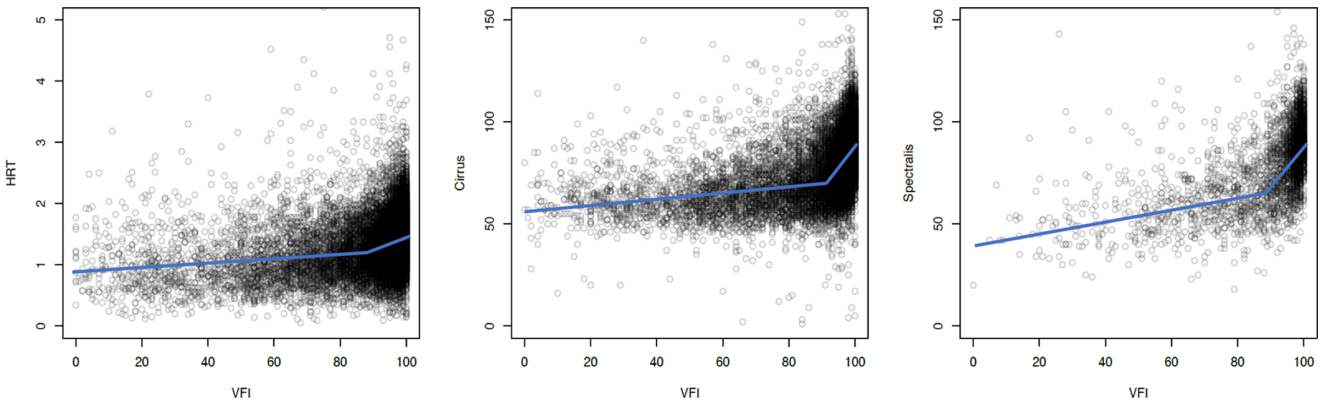

Figure 2. Broken-Stick model to determine floor values (y-intercept) for each structural device's dynamic range.

Table 2. Results of Three Different Conversion Formulas (Linear Regression, Calibration Equation, and Bland-Altman Correction Equation) Used To Convert All Measurements to a Reference Device (Spectralis)

	Conversion Formulas		
	Linear Regression	Calibration Equation	Bland-Altman
HRT Spectralis (n = 334 pairs)	Y = 0.44X + 34.3, $R^2 = 0.22$*	Spec_RNFL = 8.65 + 1*HRT_RA	Spec_RNFL = (21.4 + HRT_RA* 2.0)/2.0, $R^2 = 0.02$, $P = 0.12$†
Cirrus Spectralis (n = 565 pairs)	Y = 0.89X + 0.59, $R^2 = 0.77$*	Spec_RNFL = 6.66 + 1*Cirrus_RNFL	Spec_RNFL = (10.3 + Cirrus_RNFL* 2.0)/2.0, $R^2 = 0.05$, $P = 0$†

*R^2 = squared correlation between the two measurements.
†R^2 = squared correlation between the difference and the mean of the two measurements.

approaches to conversion (linear regression, calibration equations and Bland-Altman analysis), as detailed below.

Linear Regression

We used the slope of the linear regressions calculated for each pair of structural devices to convert normalized HRT rim area and Cirrus RNFL thickness into a Spectralis normalized RNFL thickness measurement as follows:

$$\text{Spectralis RNFL} = b_1 * \text{Cirrus RNFL} + a_1,$$
$$\text{Spectralis RNFL} = b_2 * \text{HRT} - \text{RA} + a_2,$$

where a is the intercept and b is the slope of the regression equations (Fig. 3).

Calibration Equation

We calculated a calibration equation for each pair of structural devices (HRT-Spectralis, and Cirrus-Spectralis) and used these equations to convert HRT and Cirrus normalized measurements to the equivalent Spectralis normalized measurements. This method has been explained in detail elsewhere.[11,12] In our plots, the black dashed lines represent the no-bias line (zero intercept), whereas the white circles describe the true corresponding measurements among pairs of devices, and the solid black line represents the calibration curve (Fig. 4). The closer the calibration line (full) is to the no-bias line (dashed), the smaller the systematic error between two instruments is.[12]

Bland-Altman Analysis

Bland and Altman introduced a plot that can be used to compare two measurements of the same variable to illustrate the agreement between them.[13] The X-axis is the mean of the two measurements, and the Y-axis is the difference between the two measurements. If the points on the Bland–Altman plot are scattered randomly above and below zero, then it suggests that there is no consistent bias of one method versus the other. We used the slope of the Bland-Altman plots (similar to what was done for the linear fit approach)[12] to convert HRT and Cirrus normalized measurements to the equivalent Spectralis normalized measurements (Fig. 5).

Metascore Plots

All eyes from SEI glaucoma division's database that had at least six structural scans (regardless of the devices used) and at least 3 years of follow-up were

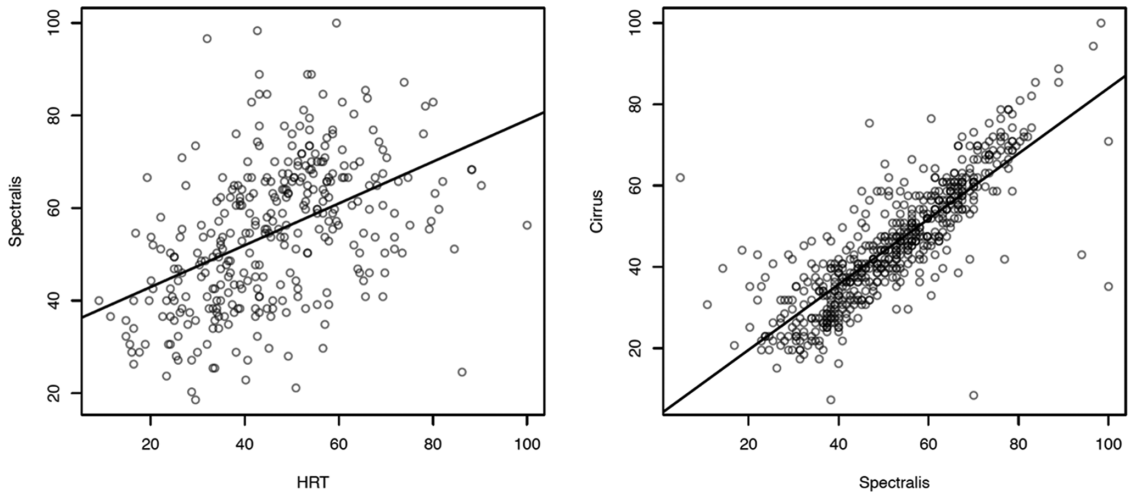

Figure 3. Linear regressions were used to estimate predicted Spectralis equivalent thickness measurement for Cirrus and HRT devices. The slopes' formulae were used to convert normalized HRT and Cirrus measurements into Spectralis normalized measurements: Spec_RNFL = 0.44*HRT_RA + 34.3, R^2 = 0.22* Spec_RNFL = 0.89*Cirrus_RNFL + 0.59, R^2 = 0.77*, *R^2 = squared correlation between the two measurements.

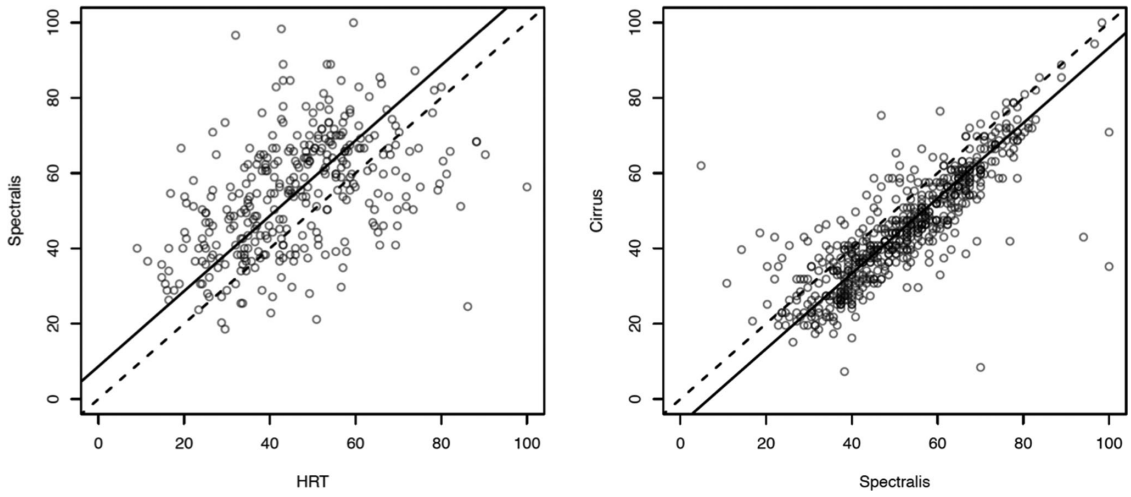

Figure 4. Calibration correction plots. We calculated a calibration correction equation for each pair of structural devices (Left: HRT-Spectralis, Right: Cirrus-Spectralis) and used these equations to convert HRT and Cirrus normalized measurements to the equivalent Spectralis normalized measurements. Spec_RNFL = 8.65 + 1*HRT_RA, Spec_RNFL = 6.66 + 1*Cirrus_RNFL.

selected to display the metascore plots (Figs. 6 and 7). Plots were created for each eye with the structural metascore on the vertical axis and follow-up (years) on the horizontal axis. Displays were coded for easy identification of measurements originating from different structural devices (HRT: orange squares, Cirrus: blue circles, Spectralis: red triangles). Metascores were calculated for all the normalization and conversion approaches mentioned above, and the goodness of fit for each approach was calculated as the root mean squared error of the resulting regression equations (Table 1). HRT and Cirrus normalized measurements were adjusted to Spectralis normalized measurements based on the goodness of fit results. Each metascore value represents a normalized (according to variance approach) and adjusted (to Spectralis) measurement. A prediction interval was calculated for each metascore slope, aiming to provide an estimated interval within which future measurements would fall based on the standard error of the model's prior measurements. This also provides an insight about the variability of the model.

Metascore Evaluation

Predictive Ability (Internal Validity)

In order to evaluate the predictive ability (internal validity) of the normalized and converted

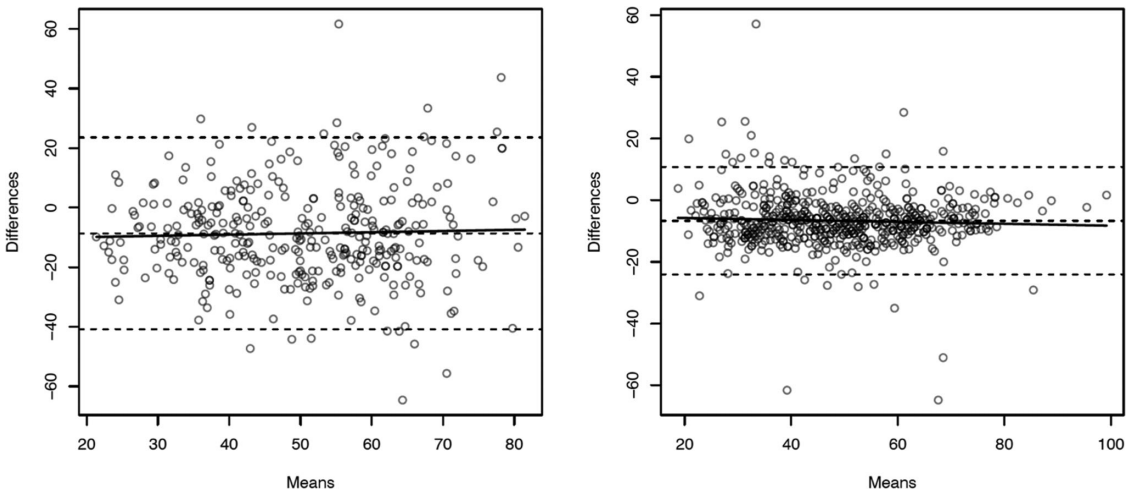

Figure 5. The Bland-Altman plot was used to assess the agreements between the HRT and Spectralis normalized measurements (left figure) and between the Cirrus and Spectralis normalized measurements (right figure). This was done by a scatter plot showing the paired differences of the two measurements on y-axis against the means of the two measurements on x-axis for each subject. The plots also showed the limits of agreement (±1.96 SD of the mean of the paired differences) as two *dashed lines*. A linear regression model was fitted in each Bland-Altman plot, and the estimates from the fitted model were used to calculate the conversion from HRT and Cirrus normalized measurements to the equivalent Spectralis normalized measurements: Spec_RNFL = (21.4 + HRT_RA * 2.0)/2.0, $R^2 = 0.02$, $P = 0.12$** Spec_RNFL = (10.3 + Cirrus_RNFL * 2.0)/2.0, $R^2 = 0.05$, $P = 0$**. ** R^2 = squared correlation between the difference and the mean of the two measurements.

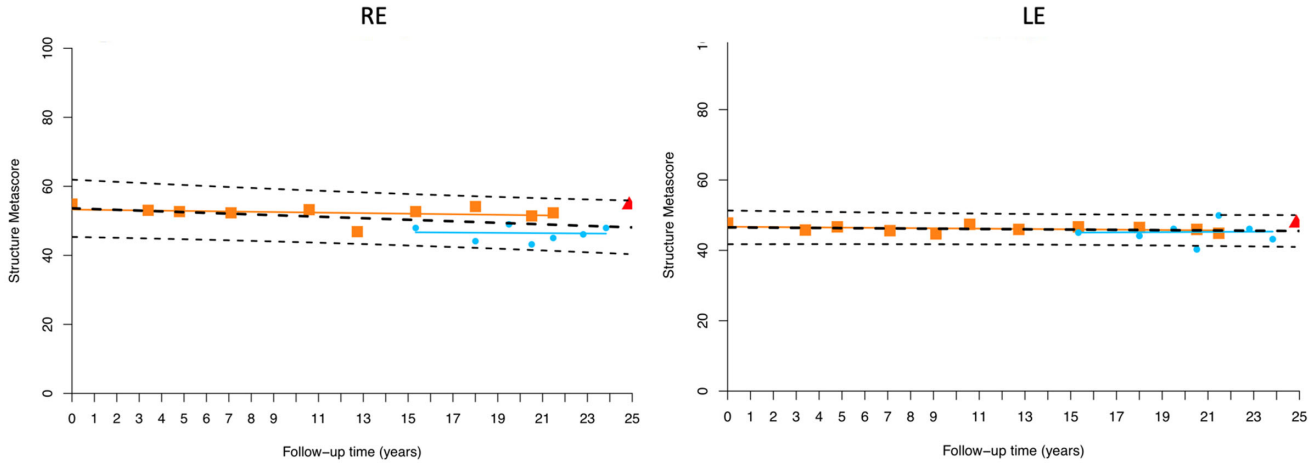

Figure 6. Metascore plots for both eyes of a patient with a 25-year follow-up period that includes measurements of all three structural devices. The slope of the metascores in these cases show structural stability on both eyes. HRT: *orange squares* and *orange regressed line*; Cirrus: *blue circles* and *blue regressed lines*; Spectralis: *red triangle*. *Dashed line*: overall regression (including all devices' measurements). The *thinner dashed lines* represent the prediction interval (range in which a future individual observation will fall) for each slope.

measurements (metascores) in predicting the structural measurements at the end of each follow-up period, we identified all eyes with ≥8 structural scans and ≥4 years of follow-up. Follow-up periods for these eyes were split so that the first and second halves included ≥4 scans and ≥2 years follow-up each. We used the linear fit on the first half to predict the average of the last two metascore measurements. We measured the predictive ability as the absolute difference between the averaged last two predicted metascores minus the averaged last two observed metascores.

Clinical Validation (External Validity)

We identified all eyes with four or more years of follow-up, six or more structural scans, and either one or more scans from each device (≥1 HRT, ≥1 Cirrus, and ≥1 Spectralis) or two or more scans from two devices (regardless of the combination of included machines) and compared the metascore slopes to

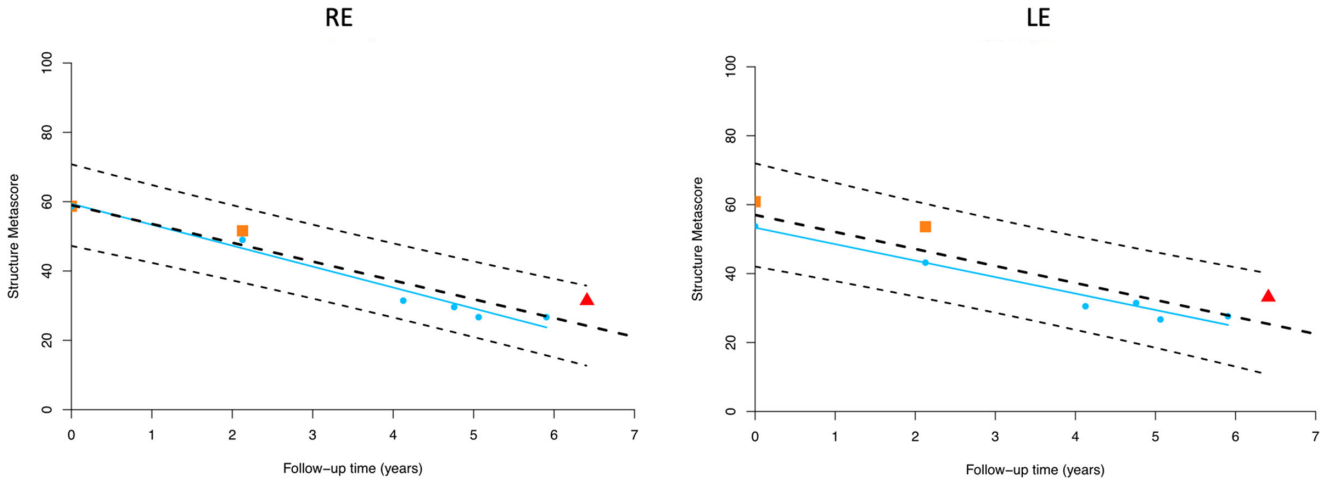

Figure 7. Metascore plots for both eyes of a patient with a seven-year follow-up period that includes measurements of all three structural devices. In this case, Cirrus OCT measurements are weighting the final slope more than the other structural devices. The metascores show structural progression of both eyes. HRT: *orange squares*; Cirrus: *blue circles* and *blue regressed lines*; Spectralis: *red triangle*. *Black dashed line*: overall regression (including all devices' measurements). The *thinner dashed lines* represent the prediction interval (range in which a future individual observation will fall) for each slope.

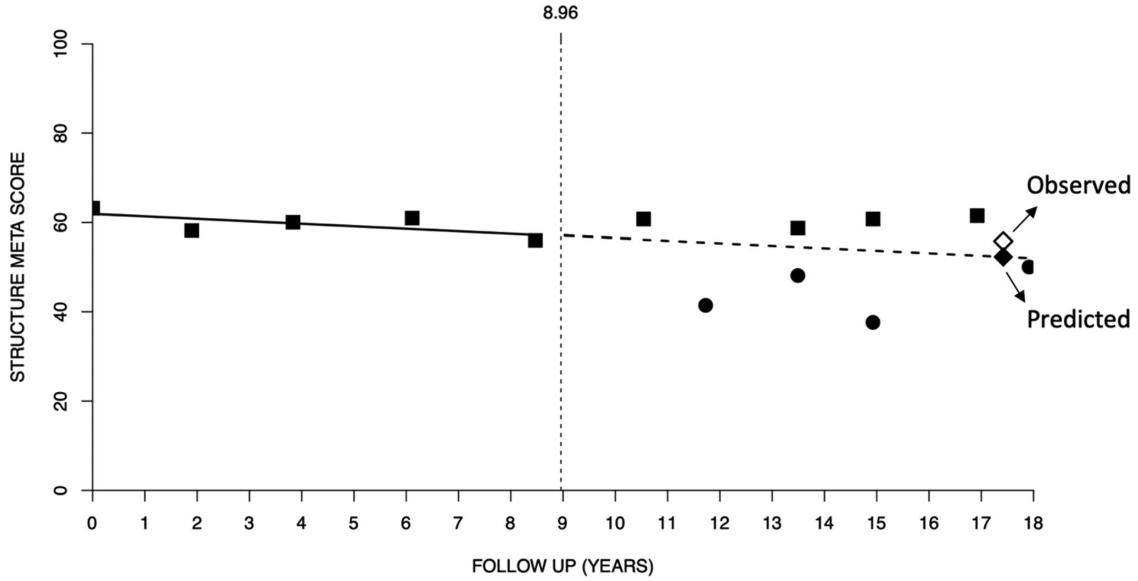

Figure 8. Metascore evaluation: Predictive ability. Linear regression model of the first half data used to predict the result of the averaged last two metascore measurements for an eye with an 18-year follow-up period. *Black squares* = HRT measurements. *Black circles* = Cirrus measurements. *White diamond* = average of last two measurements. *Black diamond* = metascore prediction of averaged last two measurements.

the clinical interpretation of the corresponding scan reports.

Subsequently, three glaucoma specialists (J.C., A.D.G., and A.R.) graded all the included structural series for likelihood of progression on a scale from 1 to 10 (1 = definitely no progression, 10 = definite progression). The gradings were performed for all the included eyes by the three specialists, with three separate times as follows: the first grading was based on the analysis of all the structural measurement printouts collected per eye; the second grading was performed with the analysis of the metascore plots alone for each eye, and the third grading was based on the analysis of the printouts plus the metascore plots for each eye (Supplemental Fig. 1).

The graders' evaluations were averaged, and correlations were calculated between their gradings and the corresponding metascore slopes (Fig. 8).

Results

A summary of the data used for normalization and adjustment, together with the results of each individual approach can be found in Tables 2 and 3. Table 4 shows the fits for all the included approaches based on their root mean squared error. The linear regression fit with the variance approach and adjusting to Spectralis was the best performing approach (RMSE 4.1), hence, conversion formulae for HRT-Spectralis and Cirrus-Spectralis measurements' adjustment are shown below whereas HRT-Cirrus conversion formulae can be found in Supplemental Material.

Normalization

Variance Approach

The data used for normalization included 34,613, 54,940 and 25,157 HRT, Cirrus, and Spectralis scans from 14,414, 8509, and 2907 eyes of 14,414, 8509, and 2907 patients, respectively. Mean ages of the included patients for each structural device were 62.5 (±15.0), 66.1 (±14.4), and 68.5 (±13.9) years for the HRT, Cirrus, and Spectralis groups, respectively. The resulting floor and ceiling values for each of the three devices for the variance approach are reported in Table 1.

Dynamic Range Approach

The data used to calculate the floor values of the dynamic range for each structural device included 19,546, 15,537, and 3072 HRT, Cirrus, and Spectralis scans from 13,440, 12,601, and 2605 eyes of 10,322, 8191, and 1619 patients, respectively. Mean (±SD) ages of the included patients for each structural device were 64.7 (±14.3), 66.5 (±14.0), and 65.2 (±14.1) years, respectively. The normal data used to estimate the ceiling of the dynamic range for each device included 126, 287, and 106 eyes of 73, 155, and 54 healthy subjects, respectively. The mean age (±SD) of patients in the normal group were 59.6 (±9.5), 63.5 (±9.0), and 59.6 (±11.0), respectively. The resulting floor and ceiling values for structural measurements of the three devices are reported in Table 5.

Table 4. Fits for All the Included Approaches (Normalization: Statistical And Dynamic Range; Conversion Formulas: Linear Fit, Calibration Correction And Bland-Altman Equation)

	Variance Approach	Dynamic Range Approach
No Adjustment	5.7	9.8
Linear Fit		
Cirrus	4.6	5.8
Spec	**4.1**	5.1
Calibration Correction		
Cirrus	5.6	9.0
Spec	5.5	9.2
Bland-Altman		
Cirrus	5.4	9.0
Spec	5.4	7.5

Each entry is the corresponding mean squared error (RMSE). The linear regression fit using the variance approach and adjusting to Spectralis was the best performing approach (RMSE 4.1), therefore, it was selected to be used for all further calculations (metascore plots).

Device Adjustment

The number of paired overlapped scans used to build the conversion formulas between paired devices were 334 and 565 for HRT-spectralis, and Cirrus-Spectralis pairs, respectively.

Table 3. Data Used for Normalization of HRT, Cirrus and Spectralis Structural Measurements (to Transform the Units of Measurement of Each Device to the Same Scale of 0 to 100)

	Normalization								
				Dynamic Range Approach					
	Variance Approach			HRT		Cirrus		Spectralis	
	HRT	Cirrus	Spectralis	Normals	Glaucoma	Normals	Glaucoma	Normals	Glaucoma
Scans	34,613	54,940	25,157	126	19,546	287	15,537	106	3072
Eyes	14,414	8509	2907	126	13,440	287	12,601	106	2605
Patients	14,414	8509	2907	73	10,322	155	8191	54	1619
Mean Age (±SD)	62.5 (±15.0)	66.1 (±14.4)	68.5 (±13.9)	59.67 (±9.54)	64.71 (±14.29)	63.52 (±9.01)	66.47 (±13.95)	59.61 (±10.95)	65.16 (±14.12)

Table 5. Floor and Ceiling Values For Each Structural Measurement Of The Three Structural Devices (Dynamic Range Approach)

Dynamic Range Approach	HRT Rim Area (mm^2)	Cirrus (RNFL Average Thickness (µm)	Spectralis Global RNFL Thickness (µm)
Floor (0) = regressed y-intercept	0.87	56.42	39.81
Ceiling value (100) = Ave normal + 1 SD	1.88	96.96	105.28

Linear Regression Fit

The linear regressions for each pair of structural devices were as follows:

HRT-Spec.

$$\text{Spec\_RNFL} = 0.44 * \text{HRT\_RA} + 34.30$$

Cirrus-Spec.

$$\text{Spec\_RNFL} = 0.89 * \text{Cirrus\_RNFL} + 0.59$$

Coefficients of determination (R^2) for each pair were 0.22 and 0.77, respectively, where R^2 is the squared correlation between the two measurements.

Calibration Equations

The calibration equations for each pair of structural devices were:

HRT-Spectralis.

$$\text{Spec\_RNFL} = 4.13 + 1 * \text{HRT\_RA}$$
$$\text{HRT\_RA} = -4.13 + 1 * \text{Spec\_RNFL}$$

Cirrus-Spectralis.

$$\text{Cirrus\_RNFL} = -4.347 + 1 * \text{Spec\_RNFL}$$
$$\text{Spec\_RNFL} = 4.347 + 1 * \text{Cirrus\_RNFL}$$

Bland-Altman Analysis

The linear regressions calculated for each pair of structural devices based on Bland-Altman analysis were:

HRT-Spectralis.

$$\text{Spec\_RNFL} = (21.4 + \text{HRT\_RA} * 2.0)/2.0$$

Cirrus-Spectralis.

$$\text{Spec\_RNFL} = (10.3 + \text{Cirrus\_RNFL} * 2.0)/2.0$$

Coefficients of determination (R^2) for each pair were 0.12 and 0.05, respectively, where R^2 is the squared correlation between the difference and the mean of the two structural measurements. Table 4 shows the fits for all of the included approaches based on their root mean squared error.

Metascore Results

The linear regression fit using the variance approach and adjusting to Spectralis was the best performing approach, and therefore it was selected to be used for all further calculations (Table 5). Metascore plots were created for 3,416 eyes of 1,824 patients. The average (±SD) age of this group was 69.8 (±13.9), mean follow-up was 11.6 (±4.7) years, and mean number of structural scans per eye was 10 (±4.7). The mean number of scans per device were 3.8 (±2.5), 5.0 (±2.9), and 1.3 (±3.0) for HRT, Cirrus, and Spectralis, respectively. The metascore slopes' median was −0.3 (interquartile range 1.1) (Supplemental Fig. 2).

Figure 6 shows examples of two metascore plots against time for both eyes of a patient with a 25-year follow-up period which includes measurements of all three structural devices. The slopes of the metascores in these cases show structural stability of both eyes. Figure 7, on the other hand, shows metascore longitudinal plots for both eyes of a patient with a seven-year follow-up period. In this case, the metascore slopes are consistent with structural progression in both eyes.

Metascore Evaluation

Predictive Ability

The group used for this analysis included 763 eyes of 423 patients. The average age of this group was 70.0 (±11.6), and the mean follow-up was 6.0 (±2.3) years for the first half and 6.8 (±2.5) for the second half of the complete follow-up period (mean 13.5 ± 5.1 years). The average absolute |Predicted metascore − Observed metascore| was 7.63/100 (with 100 the representation of the entire normalized scale). Figure 8 shows an example of the predictive performance for an eye with a 10-year follow-up period, and Figure 9 shows the frequency distribution of the prediction differences (absolute difference between the averaged last two predicted metascores minus the averaged last two observed metascores).

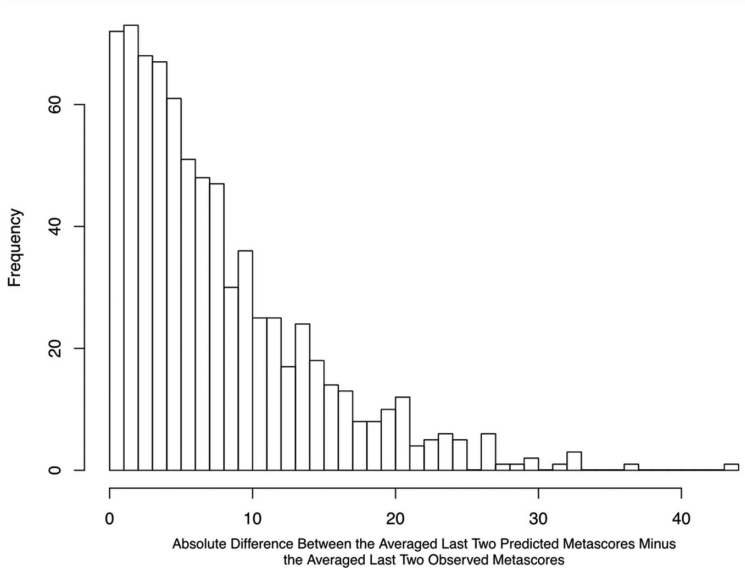

Figure 9. Frequency distribution of the absolute difference between the averaged last two predicted metascores minus the averaged last two observed metascores.

Clinical Validation

This group included 108 eyes from 108 patients. Correlations between the average grades assigned by the three clinicians and the metascore slopes were −0.51, −0.49, and −0.69 for the first (structural measurement printouts alone), second (metascore plots alone) and third (printouts + metascore plots) series of gradings, respectively; R^2 were 0.26, 0.24 and 0.48, respectively (Fig. 10). Interobserver agreement between the average clinical grading for the three gradings are shown in Table 6; overall, the metascore plots alone compared to printouts alone improved intergrader

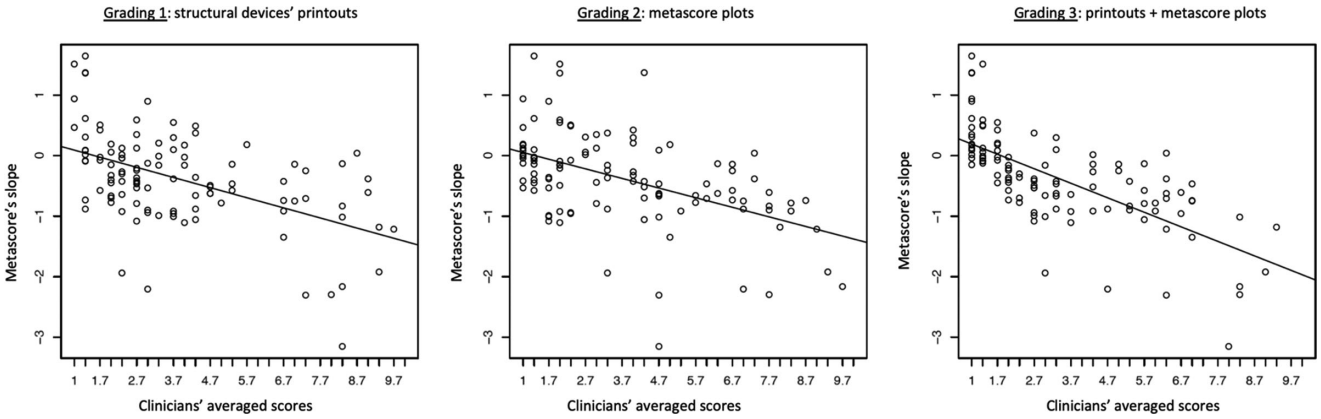

Figure 10. Clinical validation results. The scatter plots demonstrate the correlation between three glaucoma specialists' averaged gradings for structural progression (x-axis: 1 = definitely no progression, 10 = definite progression) and the corresponding metascore slopes (y-axis). First grading (left): graders' analysis of structural devices' printouts; second grading (middle): graders' analysis of the metascore plots alone; third grading (right): graders' analysis of structural printouts plus the metascore plots.

Table 6. Interobserver Agreement Between Three Masked Graders for the Three Grades

	Grade 1: Printouts Agreement (%) Kappa (95% CI)	Grade 2: Plots Agreement (%) Kappa (95% CI)	Grade 3: Printouts + Plots Agreement (%) Kappa (95% CI)
Graders A & B	79.6% 0.46 (0.29–0.63)	82.4% 0.43 (0.22–0.64)	75% 0.32 (0.14–0.5)
Graders A & C	82.4% 0.49 (0.3–0.67)	66.7% 0.32 (0.18–0.46)	79.6% 0.44 (0.26–0.62)
Graders B & C	80.6% 0.55 (0.38–0.72)	73.1% 0.45 (0.31–0.59)	84.3% 0.48 (0.29–0.79)

From left to right: pairs of graders; grade 1 (structural printouts alone); grade 2 (metascore plots alone); grade 3 (structural printouts + metascore plots).

agreement in one of three combinations of graders (A and B), and the addition of metascore plots to the printouts improved agreement (as opposed to printouts alone) in one out of three combinations of graders as well (B and C).

Discussion

The diagnosis of glaucoma and detection of glaucoma progression have been traditionally based on the finding of ONH damage assessed subjectively by ophthalmoscopy or photography and by corresponding damage to the visual field assessed by automated perimetry. Clinical ONH and RNFL assessment is known to be limited by poor to fair reproducibility and by the wide variation of normal anatomy between individuals.[14] Since the advent of automated imaging devices, structural findings of the ONH have become increasingly more reproducible and objective, but there are shortcomings that need to be addressed. The time span of glaucoma follow-up period typically outlives that of rapidly evolving imaging devices. This means that oftentimes multiple structural measurements from different devices are available. Additionally, "normality" according to these devices is based on normative databases created by device manufacturers, which often do not include a wide range of ethnicities and anatomical variations. Hence, their utility is limited to patients with clinical and demographic characteristics similar to the normative databases. Also, various devices use different scanning protocols, analytical software, output scan reports, and more, all of which challenge accurate comparison of results across different scanning devices and confound detection of long-term structural changes.

The HRT uses a 670 nm diode laser to create a layered three-dimensional image. Relative topographic heights are then calculated from a reference ring (contour line) manually placed on the optic disc, after which the instrument estimates ONH stereometric parameters. The RNFL thickness measurements have been shown to have poor diagnostic accuracy in previous studies,[15] and therefore we chose HRT rim area as the structural outcome of choice for calculating our proposed metascore.

OCT is a high-resolution imaging device that uses a low coherent broadband light source from a superluminescent diode to acquire in vivo images of the retina. It applies the principle of interferometry to interpret reflectance data from a series of multiple side-by-side A-scans combined to form a cross-sectional image. The Optic Disc Cube algorithm consists of a 1024 × 200 × 200 volume scan. Parapapillary RNFL thickness is measured along a 3.46 mm diameter measurement circle automatically placed around the optic disc (256 sampled A-scans). Spectralis OCT uses a dual-beam SD-OCT (acquisition rate of 40,000 A-scans per second), a CSLO with a wavelength of 870 nm to obtain images of ocular microstructures. It incorporates a real-time eye tracking system that couples CSLO and SD-OCT scanners to adjust for eye movements and to ensure that the same location of the retina is scanned over time.

Spectralis OCT has been widely shown to have high reproducibility[8,11] and good diagnostic accuracy in detecting glaucoma and RNFL changes.[16,17] We decided to adjust measurements of the other devices to fit its normalized scale. This methodology, however, can be applied to all other devices on the market, and measurements can be theoretically adjusted to any preferred device.

With respect to the multiple machines currently available for the acquisition of automated structural ONH measurements, several studies have explored agreement,[18–20] reproducibility,[8,12,21] and diagnostic accuracy,[22] but, to our knowledge, no method has yet been introduced to unify structural measurements provided by different scanning devices on a single scale.

Tan et al.[23] compared retinal nerve fiber layer measurements between Cirrus and Spectralis and concluded that agreement of RNFL measurement between the devices was generally good; they also found that repeatability of RNFL thickness measurements in normal participants was excellent for both OCTs. Buchser et al.[11] compared RNFL thickness measurement bias and imprecision across three SD-OCT devices (RTVue-100, Cirrus HD-OCT, and 3D OCT-1000), concluding that RNFL thickness measurements showed higher imprecision (or higher measurement variability) for the RTVue-100 than the Cirrus HD-OCT and 3D OCT-1000 devices' measurements.

Leite et al.[22] assessed diagnostic accuracy and agreement[18] of RNFL thickness measurements among RTVue, Cirrus, and Spectralis OCTs and stated that, although the spectral-domain OCTs had different resolution and acquisition rates, their ability to detect glaucoma based on areas under the curve (AUCs) and sensitivities at fixed specificities of 80% and 95% was similar. With respect to agreement, they concluded that RNFL thickness measurements obtained by different SD-OCT instruments were not entirely compatible (probably attributable to differences in RNFL detection algorithms) and should therefore not be used interchangeably. Fanihagh et al.[9] explored

correlations and strength of association of RNFL thickness in glaucoma patients among OCT, scanning laser polarimetry and CSLO; they reported a high correlation in RNFL thickness between OCT and scanning laser polarimetry, while HRT's (CSLO) topographic measurements (RNFL) displayed poor correlations with the other two imaging devices. Lally et al.[25] combined structural measurements from multiple imaging devices as inputs for machine learning classifiers as to see if this would improve discriminating ability between healthy and glaucomatous eyes, concluding that combining data from multiple devices did not significantly improve discriminating ability (Lally DR, et al. *IOVS*. 2009;50:5817).

Our metascore approach aids detection of structural change. Given the large number of structural-measuring devices available on the market, the velocity at which they are being introduced to clinical practice, and the fact that glaucoma is mostly a slow progressing disease which requires life-long clinical examinations, patients are often examined with several different instruments during their lifetime. A method that puts structural measurements provided by different devices on a same scale for their sequential interpretation would be valuable to assist clinicians' interpretation of change over long follow-up periods that include diverse devices' measurements. We believe this tool would increase the relative weight of the structural components of data in decision making about treatment, since it can provide a robust long-term trend and rate. Of course, all decisions must be made in the context, and with integration, of all other relevant clinical data such as severity of the disease, patients' wishes, expected longevity, etc.

In the clinical validation of our metascore, we observed that the correlation between the specialists' gradings and the metascore slopes decreased when the metascore plots were analyzed alone, but improved when they were reviewed together with the structural devices' printouts (Fig. 10). This suggests that the metascore might be helpful as an additional tool for structural progression analysis but may not necessarily replace the analysis of structural raw data provided by the devices' printouts. Agreement between graders improved in one out of three combinations of graders (B&C) when the printouts were analyzed together with the metascore plots (as opposed to the printouts alone) and decreased for the other two pairs of graders (A&B and A&C). We attribute this to the subjectivity of interpreting a novel method, and the fact that there was no consensus training before the grading. Regarding the metascore slopes, we obtained an overall negative trend (mean and median −0.3), which is to be expected, considering glaucomatous progression. Nevertheless, we also obtained some "positive" slopes that can be attributed either to noise and variability (property of all ancillary tests), or to actual structural changes.[24]

Our study has limitations. We used data from the structural devices used at our institution, which does not include other commercially available devices. Our metascore includes global measurements (such as RNFL thickness and rim area) and does not account for different localized or regional changes only, or for stages of glaucoma. The implementation of our methods requires a significant amount of work to pull out the relevant data from the corresponding devices. Structural scans were not filtered for segmentation errors (nor other scan artifacts), which might have resulted in some unreliable scans being included in the metascore slopes. Regarding our 1-10 clinical validation scale, it's worth mentioning that by not being externally calibrated, the scores might have included unequal steps, hence, presenting a limitation in the averaged graders' scores shown in the results. The metascore has been internally and externally evaluated with its predictive ability and clinicians' validation, respectively. We did not include an objective external reference standard for a similar approach to a combined structural measure, because we believe none are currently available. It is true that the generating the metascore on a different population may yield different coefficients in the model; we plan this as additional work in large datasets. Ultimately the utility of the technique will rest on more widespread use. Finally, its retrospective design and performance at a tertiary care center may produce results that are not entirely generalizable to other populations. Future work will include optic disc photographs with the purpose of incorporating additional structural data to the structural "metascore" and would address even longer follow-up periods.

To conclude, the capability of imaging instruments to provide additional information to the traditional examination improves the detection of glaucoma and its progression. Our aim is to combine structural measurements provided by different rapidly-evolving, commercially available measuring devices in order to achieve a reliable tool with which to gauge glaucomatous structural progression in patients with long follow-up that spans the use of several, evolving imaging methods. Specifically, we report a method that converts HRT rim area and Cirrus RNFL measurements to Spectralis global RNFL equivalent, normalized values, so that they can be evaluated on a single scale to facilitate analysis and interpretation of long-term structural data in glaucomatous eyes.

Acknowledgments

Disclosure: **A. De Gainza**, None; **E. Morales**, None; **A. Rabiolo**, None; **F. Yu**, None; **A.A. Afifi**, None; **K. Nouri-Mahdavi**, None; **J. Caprioli**, None

References

1. Hu R, Wang C, Gu Y, Racette L. Comparison of standard automated perimetry, short-wavelength automated perimetry, and frequency-doubling technology perimetry to monitor glaucoma progression. *Medicine (Baltimore)*. 2016;95(7):e2618.
2. Mwanza JC, Oakley JD, Budenz DL, Anderson DR. Ability of Cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes. *Ophthalmology*. 2011;118:241–248.e1.
3. Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol*. 2014;98(Suppl 2):ii15–ii19.
4. Chen TC, Hoguet A, Junk AK, et al. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2018;125(11):1817–1827.
5. Nouri-Mahdavi K, Nikkhou K, Hoffman DC, Law SK, Caprioli J. Detection of early glaucoma with optical coherence tomography (StratusOCT). *J Glaucoma*. 2008;17:183–188.
6. Leung CKS, Chiu V, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a comparison between spectral-domain and time-domain optical coherence tomography. *Ophthalmology*. 2011;118:1558–1562.
7. shun Leung CK, lui Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology*. 2009;116:1257–1263.e2.
8. Mwanza JC, Chang RT, Budenz DL, et al. Reproducibility of peripapillary retinal nerve fiber layer thickness and optic nerve head parameters measured with Cirrus HD-OCT in glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2010;51:5724–5730.
9. Fanihagh F, Kremmer S, Anastassiou G, Schallenberg M. Optical coherence tomography, scanning laser polarimetry and confocal scanning laser ophthalmoscopy in retinal nerve fiber layer measurements of glaucoma patients. *Open Ophthalmol J*. 2015;9:41–48.
10. Amini N, Daneshvar R, Sharifipour F, et al. Structure-function relationships in perimetric glaucoma: comparison of minimum-rim width and retinal nerve fiber layer parameters. *Invest Ophthalmol Vis Sci*. 2017;58:4623–4631.
11. Buchser NM, Wollstein G, Ishikawa H, et al. Comparison of retinal nerve fiber layer thickness measurement bias and imprecision across three spectral-domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2012;53:3742–3747.
12. Pierro L, Gagliardi M, Iuliano L, Ambrosi A, Bandello F. Retinal Nerve Fiber Layer Thickness Reproducibility Using Seven Different OCT Instruments. *Invest Ophthalmol Vis Sci*. 2012;53:5912–5920.
13. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician*. 1983;32:307.
14. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc*. 1976;74:532–572.
15. Moreno-Montañés J, Antón A, García N, Olmo N, Morilla A, Fallon M. Comparison of retinal nerve fiber layer thickness values using Stratus Optical Coherence Tomography and Heidelberg Retina Tomograph-III. *J Glaucoma*. 2009;18:528–534.
16. Silverman AL, Hammel N, Khachatryan N, et al. Diagnostic accuracy of the Spectralis and cirrus reference database in differentiating between healthy and early glaucoma eyes. *Ophthalmology*. 2016;123:408–414.
17. Beltran-Agullo L, Roca-Obis M, Ayala-Fuentes E, Morilla-Grasa A, Antón-López A. Spectralis SD-OCT and Cirrus SD-OCT: RNFL thickness measurement agreement and diagnostic performance in glaucoma. *Invest Ophthalmol Vis Sci*. 2010;51:230–230.
18. Leite MT, Rao HL, Weinreb RN, et al. Agreement among spectral-domain optical coherence tomography instruments for assessing retinal nerve fiber layer thickness. *Am J Ophthalmol*. 2011;151:85–92.e1.
19. Leung CK shun, C Ye, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography a study on diagnostic agreement with Heidelberg Retinal Tomograph. *Ophthalmology*. 2010;117:267–274.
20. Patel N, Wheat J, Rodriguez A, Tran V, Harwerth R. Agreement between retinal nerve fiber layer measures from Spectralis and Cirrus Spectral Domain OCT. *Optom Vis Sci Off Publ Am Acad Optom*. 2011;89:E652–66.

21. Rohrschneider K, Burk RO, Kruse FE, Völcker HE. Reproducibility of the optic nerve head topography with a new laser tomographic scanning device. *Ophthalmology*. 1994;101:1044–1049.
22. Leite M, Rao H, Zangwill L, Weinreb R, Medeiros F. Comparison of the diagnostic accuracies of the Spectralis, Cirrus, and rtvue optical coherence tomography devices in glaucoma. *Ophthalmology*. 2011;118:1334–1339.
23. Tan BB, Natividad M, Chua KC, Yip LW. Comparison of retinal nerve fiber layer measurement between 2 spectral domain OCT instruments. *J Glaucoma*. 2012;21:266–273.
24. Vessani R, Frota T, Shigetomi G, Correa P, Mariottoni EB, Tavares I. Structural changes in the optic disc and macula detected by swept-source optical coherence tomography after surgical intraocular pressure reduction in patients with open-angle glaucoma. *Clin Ophthalmol*. 2021;15:3017–3026.