

Intraobserver and Interobserver Agreement of Structural and Functional Software Programs for Measuring Glaucoma Progression

Javier Moreno-Montañés, MD, PhD; Vanesa Antón, OD; Alfonso Antón, MD, PhD;
 José M. Larrosa, MD, PhD; José María Martínez-de-la-Casa, MD, PhD; Gema Rebolleda, MD, PhD;
 Fernando Ussa, MD, PhD; Marta García-Granero, PhD

IMPORTANCE It is important to evaluate intraobserver and interobserver agreement using visual field (VF) testing and optical coherence tomography (OCT) software in order to understand whether the use of this software is sufficient to detect glaucoma progression and to make decisions regarding its treatment.

OBJECTIVE To evaluate agreement in VF and OCT software among 5 glaucoma specialists.

DESIGN, SETTING AND PARTICIPANTS The printout pages from VF progression software and OCT progression software from 100 patients were randomized, and the 5 glaucoma specialists subjectively and independently evaluated them for glaucoma. Each image was classified as having no progression, questionable progression, or progression. The principal investigator classified the patients previously as without variability (normal) or with high variability among tests (difficult). Using both software, the specialists also evaluated whether the glaucoma damage had progressed and if treatment change was needed. One month later, the same observers reevaluated the patients in a different order to determine intraobserver reproducibility.

MAIN OUTCOMES AND MEASURES Intraobserver and interobserver agreement was estimated using κ statistics and Gwet second-order agreement coefficient. The agreement was compared with other factors.

RESULTS Of the 100 observed patients, half were male and all were white; the mean (SD) age was 69.7 (14.1) years. Intraobserver agreement was substantial to almost perfect for VF software (overall κ [95% CI], 0.59 [0.46-0.72] to 0.87 [0.79-0.96]) and similar for OCT software (overall κ [95% CI], 0.59 [0.46-0.71] to 0.85 [0.76-0.94]). Interobserver agreement among the 5 glaucoma specialists with the VF progression software was moderate (κ , 0.48; 95% CI, 0.41-0.55) and similar to OCT progression software (κ , 0.52; 95% CI, 0.44-0.59). Interobserver agreement was substantial in images classified as having no progression but only fair in those classified as having questionable glaucoma progression or glaucoma progression. Interobserver agreement was fair regarding questions about glaucoma progression (κ , 0.39; 95% CI, 0.32-0.48) and consideration about treatment changes (κ , 0.39; 95% CI, 0.32-0.48). The factors associated with agreement were the glaucoma stage and case difficulty.

CONCLUSIONS AND RELEVANCE There was substantial intraobserver agreement but moderate interobserver agreement among glaucoma specialists using 2 glaucoma progression software packages. These data suggest that these glaucoma progression software packages are insufficient to obtain high interobserver agreement in both devices except in patients with no progression. The low agreement regarding progression or treatment changes suggests that both software programs used in isolation are insufficient for decision making.

JAMA Ophthalmol. 2017;135(4):313-319. doi:10.1001/jamaophthalmol.2017.0017
 Published online February 23, 2017.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Javier Moreno-Montañés, MD, PhD, Department of Ophthalmology, Clínica Universidad de Navarra, Universidad de Navarra, Av. Pío XII, 36, Pamplona, Spain (jmoreno@unav.es).

Progression of glaucomatous damage is one of the most important and difficult challenges in the management of glaucoma. There are no accepted gold-standard criteria for determining worsening (progression) of glaucoma changes in the visual field (VF) or retinal nerve fiber layer thickness (RNFL). The challenges of identifying glaucoma progression include difficulty discriminating true disease-related changes from variability, artifacts, patient cooperation, and natural age-related changes in structural or functional measurements.^{1,2} Clinicians most often review serial VFs or optic disc stereophotographs to detect progression. However, these methods are time consuming and highly subjective.^{3,4}

Several computational approaches recently have become available to analyze serial VF or RNFL measurements for progression that rely on trend or event analysis. Glaucoma Progression Analysis (Carl Zeiss Meditec) is a commercially available software that compares an algorithmic method to identify glaucomatous VF progression with the Humphrey Field Analyzer II (Carl Zeiss Meditec). Guided Progression Analysis software analyzes the progression of optic disc and RNFL damage using Cirrus-Optic Coherence Tomography (OCT) (Carl Zeiss Meditec).⁵ The goals of the current study were to evaluate the interobserver and intraobserver agreement of both software programs, to determine if the interobserver and intraobserver agreement was better with the VF algorithm or the OCT algorithm used to characterize progressive change, and to understand the factors associated with interobserver agreement when glaucoma progresses.

Methods

Patients

One hundred patients with glaucoma or ocular hypertension were recruited to evaluate the damage resulting from glaucoma progression. The principal investigator (J.M.-M.) selected the 100 patients. The sample included patients with stable glaucoma or progression over time based on subjective interpretation of the overview printouts and retrospective evaluation of the clinical reports (intraocular pressure, glaucoma type, and stage of damage). The principal investigator defined difficult cases as those with low change or progression and high variability among tests in trend analysis in either VF or OCT software programs.

The demographic data from these 100 patients are shown in **Table 1**. The initial VF was classified previously as glaucomatous according to the Glaucoma Staging System (GSS).⁶ Patients were excluded who had myopia (>6 D), astigmatism (>3 D), or hyperopia (>6 D). The Ethics Committee of the Clínica Universidad de Navarra approved the study, which adhered to the Declaration of Helsinki. According to this committee, no written informed consent was needed, because this evaluation was performed on data that are normally obtained in clinical practice.

Software Analysis

The printout pages of the analysis from the Humphrey Field Analyzer II (1 page) and the Cirrus-OCT (2 pages) were merged

Key Points

Question What is the level of interobserver and intraobserver agreement among 5 glaucoma specialists in the use of visual field testing and optical coherence tomography software to detect glaucoma progression?

Findings In this study, while intraobserver agreement was substantial for both software programs, interobserver agreement was moderate and associated with the variability among tests. Interobserver agreement was substantial in patients with no progression but was only fair in patients with questionable glaucoma progression or glaucoma progression.

Meanings The moderate interobserver agreement using both software programs indicates that these programs used in isolation are insufficient for decision making regarding glaucoma damage progression.

into one 3-page PDF file (Adobe Acrobat Reader, Adobe Systems Software), and the images were numbered from 1 to 100. The files were uploaded to a folder on the Basic Dropbox Business (Dropbox) file-sharing site. The intraocular pressure, glaucoma type, number of treatments, and duration of glaucoma were not included in the information provided to the specialists. All included patients had reliable VF tests using the Swedish Interactive Threshold Algorithm (Carl Zeiss Meditec). Patients with an improved visual field index (VFI) in the 3 initial and consecutive tests were considered part of the learning curve and excluded from the study. One experienced operator, who was not one of the examiners who performed the VF testing and was masked to the other findings, performed all OCT examinations. All OCT scans had a signal strength of 6 or higher.

Specialist Evaluation

Five glaucoma specialists (A.A., J.M.L., J.M.-d.-I.-C., G.R., and F.U.) from 5 different university departments of ophthalmology evaluated the 100 images. These specialists have published glaucoma studies and had used both software programs. All specialists answered the following questions: (1) Do you think that the VF damage is progressing? (2) Do you think that the OCT damage is progressing? (3) According to the changes in the VF and the OCT images, do you think the glaucoma is progressing? (4) Do you think a treatment change is needed? Each question could be answered with no, "questionable," or "progression," except for question 4, for which the answers were no, "questionable," and "definitely." No other comments or instructions were provided.

The specialists received an Excel file (Microsoft Corporation) with the 100 images ordered randomly and differently to avoid a fatigue effect in the answers. Only the cited answers were included in each box of the Excel file. After answering all questions, the specialists sent the Excel file to the principal investigator, and the 100 files disappeared from the Dropbox shared folder. No expert knew the names of the other specialists or their answers.

One or more months later, the 100 images were reordered randomly, assigned a new identification number from 1 to 100, and uploaded again to a new shared Dropbox folder. A new Excel file was sent to the specialists using a new randomized order

of images to answer. The questions remained the same as in the first analysis. The specialists sent the Excel file with the answers, and the files were removed from Dropbox. The principal investigator and an assistant (V.A.) combined the answers to evaluate the intra-agreement and interagreement.

Statistical Analysis

Agreement among the different specialists was reported using nominal κ s (ignoring the sequence) and weighted κ s (using linear weights),⁷ prevalence-adjusted and bias-adjusted κ s,⁸ and Gwet second-order agreement coefficients,⁹ with PEPI-for-Windows version 11.15 (Microsoft). Multiple raters' κ s were used to assess the agreement among the specialists. Confidence intervals for these κ values were computed using bootstrap (bias corrected and accelerated method; 100 000 repetitions) with Stata SE version 12.1 (StataCorp). All κ values were interpreted according to the guidelines of Landis and Koch.¹⁰

Descriptive statistical analyses were performed using SPSS version 20.0 (IBM Corporation). Interobserver agreement was classified as high agreement (the same answer from 5 or 4 specialists) and low agreement (the same answer from 3 or 2 specialists). This level of interobserver agreement was compared with the other parameters using logistic regression analysis with odds ratios (ORs). These parameters included the initial VFI, final VFI, change between initial and final VFIs, number of VFs, VF follow-up time, Glaucoma Staging System classification, initial RNFL, final RNFL, change in the initial and final RNFLs, number of OCT images, OCT follow-up time, and case difficulty.

Results

Intraobserver Agreement

The κ statistics for the intraobserver analysis among the glaucoma specialists and the 4 questions are shown in Table 2. According to the Landis and Koch classification,¹⁰ the agreement was almost perfect for question 1 about VF progression among the 4 specialists and substantial for 1 specialist; similarly, the κ values were almost perfect regarding question 2 on progression on the OCT images. No difference was found between the VF and OCT intraobserver agreement. The κ for question 3 on glaucoma damage progression showed slightly lower agreement, with substantial to almost perfect agreement among the examiners. Finally, question 4 on change in treatment saw moderate to almost perfect agreement, depending on the examiners. The agreement generally was higher although different among the examiners.

Interobserver Agreement

Table 3 shows the overall κ statistics for interobserver agreement. The combined VF agreement was moderate, although it was substantial in the answer reflecting no progression. The lowest κ values were associated with the questionable or progression answers in all questions. The agreement was similar in the combined OCT progression and slightly higher in answers reflecting no progression. The differences in interobserver agreement between question 1 and question 2 were small in either the first (κ difference, -0.03 ; 95% CI, -0.12 to 0.06 ;

Table 1. Demographic Data

Characteristic	No. (%)
Male	50 (50.0)
Age, mean (SD) [range], y	69.7 (14.1) [14-90]
Right eyes examined	52 (52.0)
White	100 (100)
VF, mean (SD) [range]	
No.	9.8 (3.4) [6-18]
Initial VFI	77.5 (24.5) [15-100]
Final VFI	71.7 (27.5) [0-100]
Time of follow-up, mo	52.6 (19.2) [12-101]
OCT, mean (SD) [range]	
No.	7.5 (0.9) [5-8]
Initial average RNFL thickness	72.9 (14.3) [45-106]
Final average RNFL thickness	68.5 (13.9) [42-105]
Time of follow-up, mo	53.4 (16.5) [12-96]
GSS in the first VF, patients	
Stage 0	15 (15.0)
Stage 1	30 (30.0)
Stage 2	25 (25.0)
Stage 3	18 (18.0)
Stage 4	12 (12.0)

Abbreviations: GSS, Glaucoma Staging System⁴; OCT, optical coherence tomography; RNFL, retinal nerve fiber layer thickness; VF, visual field; VFI, visual field index.

$P = .53$) or the second (κ difference, -0.06 ; 95% CI, -0.15 to 0.03 ; $P = .18$) evaluations. The κ values from question 3 were 0.39 (95% CI, 0.32 - 0.48) and 0.38 (95% CI, 0.30 - 0.47) in the first and second evaluations, respectively, but the agreement was substantial in patients with no progression. Similar low agreement was obtained for question 4.

The agreement among the specialists according to the classification of the principal investigator regarding the ease or difficulty of the images for detecting progression is shown in Table 4. The κ values were higher in the normal images than in the difficult images regarding combined agreement and agreement in patients with no progression.

Comparison of Interobserver Agreement With Other Factors

Interobserver agreement regarding the VF was associated with the Glaucoma Staging System classification (OR, 1.52 ; 95% CI, 1.07 - 2.16 ; $P = .01$), initial VFI (OR, 0.98 ; 95% CI, 0.97 - 0.99 ; $P = .03$), final VFI (OR, 0.99 ; 95% CI, 0.96 - 0.99 ; $P = .005$), difference in initial to final VFI (OR, 1.04 ; 95% CI, 1.00 - 1.09 ; $P = .04$), and easy/difficult diagnosis (OR, 8.40 ; 95% CI, 2.53 - 27.85 ; $P = .001$). Thus, the agreement was higher in early-stage glaucoma, with high initial VFI, with low final VFI, and with higher changes in the VFI. No association was found between interobserver agreement in the VF and the number of VFs (OR, 1.05 ; 95% CI, 0.94 - 1.18 ; $P = .40$) or months of follow-up (OR, 1.02 ; 95% CI, 0.99 - 1.04 ; $P = .12$). Interobserver agreement of the OCT values was associated with the difference in the initial and final RNFL thicknesses (OR, 0.93 ; 95% CI, 0.87 - 0.99 ; $P = .04$). No association was found between

Table 2. Intraobserver Agreement Among 5 Glaucoma Specialists

Characteristic	Specialist 1	Specialist 2	Specialist 3	Specialist 4	Specialist 5
Progression seen on VF software					
Overall κ (95% CI)	0.87 (0.79-0.96)	0.73 (0.62-0.84)	0.75 (0.63-0.87)	0.59 (0.46-0.72)	0.84 (0.75-0.93)
Weighted κ (95% CI)	0.90 (0.83-0.97)	0.82 (0.73-0.90)	0.78 (0.67-0.88)	0.68 (0.58-0.79)	0.88 (0.81-0.95)
PABAK	0.92	0.82	0.86	0.75	0.89
AC2	0.94	0.80	0.91	0.83	0.89
Progression seen on Cirrus-OCT software					
Overall κ (95% CI)	0.85 (0.76-0.94)	0.73 (0.64-0.83)	0.69 (0.56-0.82)	0.59 (0.46-0.71)	0.74 (0.63-0.85)
Weighted κ (95% CI)	0.90 (0.83-0.98)	0.81 (0.72-0.90)	0.72 (0.59-0.84)	0.70 (0.60-0.80)	0.81 (0.72-0.89)
PABAK	0.97	0.97	0.83	0.70	0.83
AC2	0.91	0.85	0.89	0.80	0.87
Glaucoma damage progression					
Overall κ (95% CI)	0.85 (0.76-0.94)	0.67 (0.55-0.80)	0.71 (0.58-0.84)	0.60 (0.47-0.73)	0.69 (0.57-0.81)
Weighted κ (95% CI)	0.87 (0.79-0.95)	0.77 (0.67-0.87)	0.74 (0.63-0.86)	0.65 (0.52-0.77)	0.77 (0.67-0.86)
PABAK	0.89	0.78	0.84	0.73	0.80
AC2	0.90	0.75	0.90	0.79	0.85
Treatment change					
Overall κ (95% CI)	0.84 (0.74-0.95)	0.83 (0.72-0.94)	0.67 (0.55-0.79)	0.56 (0.43-0.70)	0.55 (0.41-0.68)
Weighted κ (95% CI)	0.83 (0.72-0.94)	0.83 (0.72-0.94) ^a	0.65 (0.51-0.79)	0.59 (0.45-0.72)	0.53 (0.38-0.67)
PABAK	0.91	0.84	0.69	0.63	0.59
AC2	0.93	0.85	0.59	0.59	0.53

Abbreviations: AC2, Gwet second-order agreement coefficient⁹; PABAK, prevalence-adjusted and bias-adjusted κ .

^a The answers from this expert to this question were only "none" and "definitive."

Table 3. Interobserver Agreement Among 5 Glaucoma Specialists

Characteristic	No Progression, κ	Questionable Progression, κ	Progression, κ	Combined, κ (95% CI) ^a
Progression seen on VF software				
First evaluation	0.72	0.25	0.43	0.48 (0.41-0.55)
Second evaluation	0.69	0.26	0.44	0.48 (0.41-0.56)
Progression seen on OCT software				
First evaluation	0.68	0.30	0.54	0.52 (0.44-0.59)
Second evaluation	0.52	0.37	0.55	0.55 (0.47-0.63)
Glaucoma damage progression				
First evaluation	0.63	0.19	0.37	0.39 (0.32-0.48)
Second evaluation	0.64	0.18	0.33	0.38 (0.30-0.47)
Treatment change				
First evaluation	0.59	0.42	0.07	0.39 (0.32-0.48)
Second evaluation	0.62	0.43	0.10	0.42 (0.34-0.50)

Abbreviations: OCT, optical coherence tomography; VF, visual field.

^a Using bootstrap (bias corrected and accelerated 95% CI).

interobserver agreement in the OCT and number of OCT images (OR, 1.04; 95% CI, 0.67-1.62; $P = .85$) or follow-up time (OR, 1.00; 95% CI, 0.98-1.00; $P = .73$), initial RNFL thickness (OR, 0.98; 95% CI, 0.96-1.02; $P = .35$), final RNFL thickness (OR, 0.99; 95% CI, 0.97-1.03; $P = .97$), or easy/difficult diagnosis (OR, 0.04; 95% CI, 0.13-1.24; $P = .11$).

Discussion

A goal in the follow-up of patients with glaucoma is to distinguish whether the observed changes are associated with test variability, artifacts, or actual disease progression. Different software to measure glaucomatous progression, such as those

used in the current study, automatically adjust each follow-up test to the baseline test to detect subtle changes in event or trend analyses. These algorithms are designed to limit the background noise and, therefore, to provide a more valid analysis of longitudinal data, theoretically increasing the ability to discern true progression of glaucomatous disease.

Previous articles on agreement among specialists using different functional or structural tests to detect progression have been published. Viswanathan et al¹¹ examined the level of agreement among 5 clinicians in assessing progressive deterioration in a VF series using standard Humphrey printouts and linear regression analysis. The authors reported a κ of 0.59 using Progressor analysis and a κ of only 0.32 using Humphrey printouts ($P = .006$). Other authors have reported

Table 4. Interobserver Agreement Among 5 Glaucoma Specialists Depending on the Ease or Difficulty of the Patients

Characteristic	No Progression, κ	Questionable Progression, κ	Progression, κ	Combined, κ (95% CI) ^a
Progression seen on VF software				
Normal	0.82	0.20	0.54	0.57 (0.49-0.65)
Difficult	0.38	0.09	0.03	0.18 (0.08-0.31)
Progression seen on OCT software				
Normal	0.74	0.28	0.60	0.56 (0.48-0.66)
Difficult	0.49	0.21	0.16	0.32 (0.18-0.48)
Glaucoma damage progression				
Normal	0.71	0.21	0.46	0.47 (0.38-0.56)
Difficult	0.35	0.04	0.07	0.14 (0.05-0.25)
Treatment change				
Normal	0.70	0.53	0.13	0.51 (0.42-0.60)
Difficult	0.29	0.09	0.08	0.11 (0.02-0.25)

Abbreviations: OCT, optical coherence tomography; VF, visual fields.

^a Using bootstrap (bias corrected and accelerated 95% CI).

similar levels of agreement.¹²⁻¹⁵ However, Tanna et al,¹⁶ who studied the agreements in 5 VF tests in 100 eyes evaluated by 5 glaucoma subspecialists, reported a moderate interobserver agreement (κ , 0.45) that did not improve when each specialist had access to Humphrey progression software.¹⁶ The authors suggested that reviewers varied regarding the degree to which they agreed with the software and that their determinations regarding progression were infrequently altered by the software data. Specialist readers might have dismissed certain progression patterns classified as such by the progression software as an artifact or error. In the current study, intraobserver agreement was high, but interobserver agreement was only moderate (κ , 0.48) using this progression software, similar to previous articles.^{11,16}

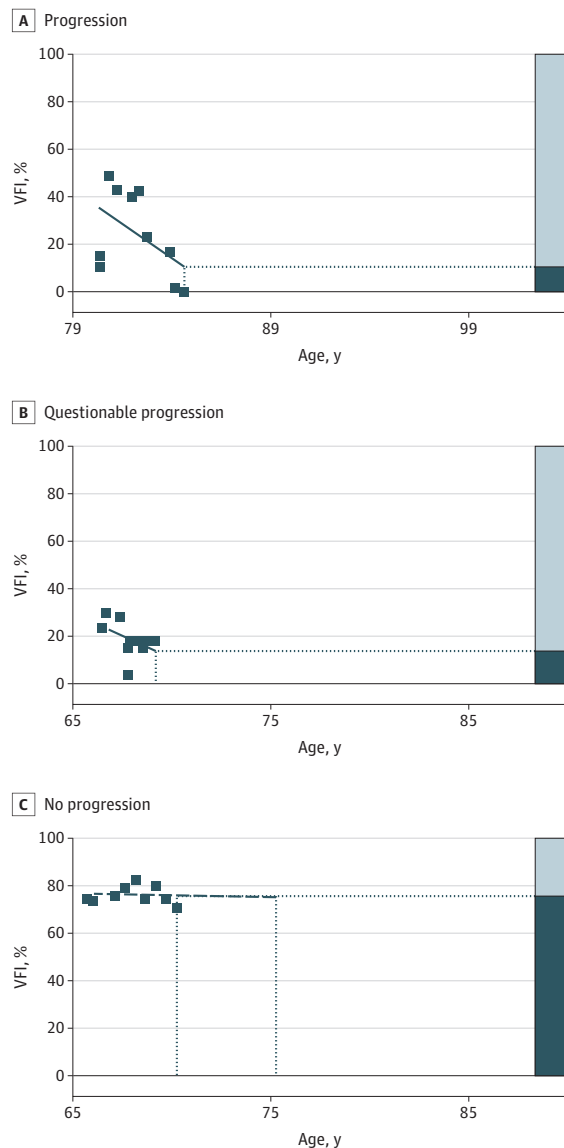
The current study also evaluated the agreement in the progression analysis from the Cirrus-OCT machine. Our results suggest only moderate interobserver agreement. To our knowledge, the current study is the first to analyze the agreement in this software from the Cirrus-OCT instrument. However, previous studies of the agreement of other structural tests have been published. Breusegem et al¹⁷ evaluated the interobserver κ coefficients among 3 specialists of a set of 2 serial optic disc color stereoslides for glaucomatous changes in 40 patients. The authors reported moderate interobserver agreement (κ , 0.51; 95% CI, 0.33-0.69).¹⁷ In another report, using Heidelberg Retina Tomograph (Heidelberg Engineering) printouts for 168 patients, 3 experienced observers found moderate to good agreement among all methods from the Heidelberg Retina Tomograph.¹⁸ In the current study, the comparison between κ values of the agreement in OCT progression was slightly higher than in VF progression, but the data did not indicate that a confident difference was identified. These results suggest that the agreement values between structural and functional software regarding progression are similar and that one is not superior to the other. However, the agreement differed in the 3 answers to each question. Thus, the agreement of the negative answers (no progression) was moderate to substantial whereas agreement was fair to moderate in the answers regarding findings that were questionable or progression. The different levels of agreement in the 3 evaluated

answers for each question suggested that the progression software programs used alone are more useful in patients with no progression than in those with questionable glaucoma progression or glaucoma progression.

The main cause of low agreement regarding the answers of questionable or progression might be the variability of the different tests. There are many causes of variability, such as patient-reported alertness, increasing age, test-retest variability, glaucoma stage, test duration, and fixational eye movements.¹⁹⁻²² Previous studies have suggested that even experienced examiners have difficulty assessing progression in reliable VF series when tests are variable.¹ A consensus has indicated that measurement variability affects the ability of any device to detect progression.²³ The goal of the current study was not to analyze the causes of variability in OCT scans or VFI from the Humphrey software. There are 4 frequently used measures of variability: range, interquartile range, variance, and standard deviation. The only available data used to analyze the variability of the printout pages from the Humphrey progression software were the 95% CIs of the VFI trend analysis, whereas the standard deviation of VFI trend analysis was used in the progression software in the Cirrus-OCT instrument. In the current 100 patients, we found that the 95% CI was higher (greater than 1%) than the mean VFI in 32 patients and a standard deviation higher than the mean in 31 patients in the average RNFL thickness from OCT. In patients with a high change value in the VFI or RNFL thickness with high variability among tests, the case will be diagnosed as progression. However, in patients with a low change value but with high variability among tests, it is difficult to distinguish between variability and true progression (Figure). A previous report²⁴ showed that the VF fluctuation is low in normal eyes and those with early glaucoma and increases from stage 0 to 4 from the Glaucoma Staging System. Similar findings were published indicating that a thin RNFL in patients with advanced glaucoma was associated with more variability in the RNFL thickness.²⁵ These results suggest that the variability among different tests is the cause of the low agreement in the current study.

Questions 3 and 4 showed only fair to moderate agreement. These questions are important because they indicate

Figure. Trend Analysis of the Visual Field Index (VFI)



Trend analysis of the VFI for 3 patients with 95% CIs higher than 1% of the mean VFI. A, A patient with a progression rate of -5.8% (95% CI, -13.9 to 2.3), with high agreement among the specialists that there was glaucoma progression. The first VFI was 16% and the second was 11% (this patient was not considered a learning curve). B, A patient with a progression rate of -3.9% (95% CI, -10.2 to 2.4) with low agreement because the specialists consider this patient to have questionable progression or progression. C, A patient with a progression rate of -0.2% (95% CI, -4.1 to 0.1) with high agreement considered by the specialists as having no progression. The blue squares indicate VFI values; the solid line, trend analysis; the dashed line, age and final VFI in the trend analysis; box, final VFI in the trend analysis.

whether the software used to measure progression is useful alone to make clinical decisions in daily practice. The patients in this study had early and advanced glaucoma. Previous studies have suggested that glaucoma progression could be better detected using OCT in patients with early glaucoma, while VF testing could better detect advanced glaucoma.^{26,27} Our results suggest that the examiners have

different criteria to detect progression or determine treatment changes probably because no other clinical parameters have been included to facilitate decisions. Nowadays, neither of the most widely used algorithms meant to inform ophthalmologists whether a VF or a RNFL thickness has deteriorated are sufficiently reliable for determining whether a particular patient with glaucoma has worsened or not. Therefore, to base therapeutic decisions on these algorithms in isolation is inaccurate and could increase the risk of damage in glaucomatous patients.

Limitations

Our study had limitations. First, the sample included patients with variability among different tests, which caused more doubts when trying to identify progression or no progression. However, many patients in clinical practice show variability among tests. Also, the same examiner performed all OCT scans, which excluded the examiner as a variability factor. Second, the Cirrus-OCT scans were obtained without an eyetracker. An eyetracker has only recently been included in the Cirrus software. Third, the specialists received no recommendations for evaluating the printout pages; thus, some specialists might have placed more emphasis on the event analysis and others on the trend analysis. However, we preferred to not provide advice so that the specialists could interpret glaucoma progression as they did in daily clinical practice. The follow-up time was short (up to 101 months). However, the Cirrus is the newest-generation OCT device from Carl Zeiss Meditec, and it has only been available for clinical use for the last 8 years. Additionally, the number of scans or the follow-up time was unrelated to interobserver agreement. Finally, only a subjective evaluation of agreement was evaluated and not accuracy or precision; the goal of this study was to determine agreement regarding progression using both software packages in a group of patients to facilitate clinical decision making about progression or treatment changes. Despite these limitations, our results might contribute to the knowledge about the advantages and problems of both software packages in detecting glaucoma progression.

Conclusions

In conclusion, substantial or almost perfect intraobserver agreement was found in the evaluated questions. Nevertheless, only moderate interobserver agreement was found in the combined results. No progression had more agreement than the answers about questionable progression or progression. The factors associated with interobserver agreement depended on variability. In patients with low mean and high variability values in the trend analysis, the diagnosis of progression was questionable. The interobserver agreement levels of the global progression of glaucoma damage and treatment changes were only fair. Our results suggest that both progression software packages have similar agreement and are more clinically useful in patients with no progression than in patients with questionable progres-

sion or progression. Hence, the data suggest that these glaucoma progression software packages used alone are insufficient to obtain high interobserver agreement in patients

with glaucoma progression. More effort must be made to improve the software to reduce the variability and facilitate diagnosis of glaucoma progression.

ARTICLE INFORMATION

Accepted for Publication: January 8, 2017.

Published Online: February 23, 2017.

doi:10.1001/jamaophthalmol.2017.0017

Author Affiliations: Department of Ophthalmology, Clínica Universidad de Navarra, Universidad de Navarra, Pamplona, Spain (Moreno-Montañés, V. Antón); Institut Català de la Retina, Universidad Internacional de Catalunya, Barcelona, Spain (A. Antón); Hospital Miguel Servet, Universidad de Zaragoza, Zaragoza, Spain (Larrosa); Hospital Clínico San Carlos, Universidad Complutense, Madrid, Spain (Martínez-de-la-Casa); Hospital Ramón y Cajal, Universidad Autónoma de Madrid, Madrid, Spain (Rebolledo); Instituto de Oftalmología Aplicada, Universidad de Valladolid, Valladolid, Spain (Ussa); Unidad de Estadística, Departamento de Bioquímica y Genética, Universidad de Navarra, Pamplona, Spain (García-Granero).

Author Contributions: Dr Moreno-Montañés had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Moreno-Montañés, A. Antón, Larrosa.

Acquisition, analysis, or interpretation of data: V. Antón, A. Antón, Larrosa, Martínez-de-la-Casa, Rebolledo, García-Granero.

Drafting of the manuscript: Moreno-Montañés, Martínez-de-la-Casa, García-Granero.

Critical revision of the manuscript for important intellectual content: Moreno-Montañés, V. Antón, A. Antón, Larrosa, Martínez-de-la-Casa, Rebolledo, García-Granero.

Statistical analysis: Moreno-Montañés, Martínez-de-la-Casa, García-Granero.

Administrative, technical, or material support: Moreno-Montañés, V. Antón, Martínez-de-la-Casa.

Study supervision: Moreno-Montañés, Larrosa, Martínez-de-la-Casa, Rebolledo.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: Supported in part by Instituto de Salud Carlos III, "Red temática de Investigación Cooperativa, Proyecto RD07/O063. OftaRed: Patología ocular del envejecimiento, calidad visual y calidad de vida."

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Previous Presentation: This article was presented in part at the Annual Meeting of the Association for Research in Vision and Ophthalmology; May 1, 2016; Seattle, Washington.

REFERENCES

- Nouri-Mahdavi K, Brigatti L, Weitzman M, Caprioli J. Comparison of methods to detect visual field progression in glaucoma. *Ophthalmology*. 1997;104(8):1228-1236.
- Werner EB, Petrig B, Krupin T, Bishop KI. Variability of automated visual fields in clinically stable glaucoma patients. *Invest Ophthalmol Vis Sci*. 1989;30(6):1083-1089.
- Heijl A, Bengtsson B. Diagnosis of early glaucoma with flicker comparisons of serial disc photographs. *Invest Ophthalmol Vis Sci*. 1989;30(11):2376-2384.
- Photogrammetry of the optic disc [editorial]. *Br J Ophthalmol*. 1980;64(12):879-880.
- Carl Zeiss Meditec Inc. *Details Define Your Decisions*. Jena, Germany: Carl Zeiss Meditec; 2007.
- Mills RP, Budenz DL, Lee PP, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease. *Am J Ophthalmol*. 2006;141(1):24-30.
- Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. New York, NY: John Wiley & Sons; 2003.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-429.
- Gwet KL. *Handbook of Inter-Rater Reliability: the Definitive Guide to Measuring the Extent of Agreement Among Raters*. 2nd ed. Gaithersburg, MD: Advances Analytics; 2010:76-78, 80-81.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
- Viswanathan AC, Crabb DP, McNaught AI, et al. Interobserver agreement on visual field progression in glaucoma: a comparison of methods. *Br J Ophthalmol*. 2003;87(6):726-730.
- Werner EB, Bishop KI, Koelle J, et al. A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. *Arch Ophthalmol*. 1988;106(5):619-623.
- Chauhan BC, Drance SM, LeBlanc RP, Lieberman MF, Mills RP, Werner EB. Technique for determining glaucomatous visual field progression by using animation graphics. *Am J Ophthalmol*. 1994;118(4):485-491.
- Iester M, Corallo G, Capris E, Capris P. Agreement in detecting glaucomatous visual field progression by using guided progression analysis? and Humphrey overview printout. *Eur J Ophthalmol*. 2011;21(5):573-579.
- Lin AP, Katz LJ, Spaeth GL, et al. Agreement of visual field interpretation among glaucoma specialists and comprehensive ophthalmologists: comparison of time and methods. *Br J Ophthalmol*. 2011;95(6):828-831.
- Tanna AP, Bandi JR, Budenz DL, et al. Interobserver agreement and intraobserver reproducibility of the subjective determination of glaucomatous visual field progression. *Ophthalmology*. 2011;118(1):60-65.
- Breusegem C, Fieuws S, Stalmans I, Zeyen T. Agreement and accuracy of non-expert ophthalmologists in assessing glaucomatous changes in serial stereo optic disc photographs. *Ophthalmology*. 2011;118(4):742-746.
- Vizzeri G, Weinreb RN, Martínez de la Casa JM, et al. Clinicians agreement in establishing glaucomatous progression using the Heidelberg retina tomograph. *Ophthalmology*. 2009;116(1):14-24.
- Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Res*. 2007;47(7):925-936.
- Turpin A, McKendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci*. 2011;52(6):3237-3245.
- Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci*. 2012;53(10):5985-5990.
- Gillespie BW, Musch DC, Guire KE, et al; CIGTS (Collaborative Initial Glaucoma Treatment Study) Study Group. The collaborative initial glaucoma treatment study: baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci*. 2003;44(6):2613-2620.
- Weinreb RN, Garway-Heath DF, Leung C, Crowston JG, Medeiros FA. *Progression of Glaucoma*. Amsterdam, the Netherlands: Kugler Publications; 2011:137. Consensus Series; Vol 8.
- Fogagnolo P, Sangermani C, Oddone F, et al. Long-term perimetric fluctuation in patients with different stages of glaucoma. *Br J Ophthalmol*. 2011;95(2):189-193.
- Wu Z, Vazeen M, Varma R, et al. Factors associated with variability in retinal nerve fiber layer thickness measurements obtained by optical coherence tomography. *Ophthalmology*. 2007;114(8):1505-1512.
- 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(11):6939-6946.
- Wollstein G, Kagemann L, Bilonick RA, et al. Retinal nerve fibre layer and visual function loss in glaucoma: the tipping point. *Br J Ophthalmol*. 2012;96(1):47-52.