Comparing Glaucomatous Disc Change Using Stereo Disc Viewing and the MatchedFlicker Software Program in Ophthalmologists-in-Training

JAMIE L. SCHAEFER, ZACHARY L. LUKOWSKI, ALISSA M. MEYER, ANTHONY J. LEONCAVALLO, ANTHONY GREER, GINA M. MARTORANA, BAIMING ZOU, JONATHAN J. SHUSTER, AND MARK B. SHERWOOD

• PURPOSE: To compare the accuracy and speed of using the computerized MatchedFlicker software program (EyeIC Inc, Narberth, Pennsylvania, USA) to evaluate glaucomatous optic disc change against the traditional gold standard of manually examining stereoscopic disc photographs.

• DESIGN: A prospective evaluation of diagnostic technology.

• METHODS: Two resident ophthalmologists and 1 glaucoma fellow at the University of Florida independently evaluated 140 image pairs from 100 glaucomatous/ocular hypertensive patient eyes using a handheld stereo viewer and the MatchedFlicker program. Fifty had progression to glaucoma as determined by the Ocular Hypertension Treatment Study (OHTS) Optic Disc Reading Group and the OHTS Endpoint Committee in the OHTS, and 50 more had photographs taken a few minutes apart, which were negative controls with no progression. Twenty photograph pairs from each group were duplicated to determine reviewer variability. Photographs were examined in alternating blocks of 70 photograph pairs for each method, with the starting viewing method randomized. Reviewer accuracy and time to review for each method were measured.

• RESULTS: Using the handheld stereo viewer, the reviewers correctly identified progression or nonprogression in 76.0% of the slide pairs. Using the MatchedFlicker software, 87.6% were correctly identified ($P = .011$). Evaluator speed averaged 34.1 seconds per image pair with the stereo viewer vs 24.9 seconds with the MatchedFlicker program ($P = .044$). Overall, Flicker was significantly more specific but less sensitive than stereo slides. Trainees appeared more reluctant to identify glaucoma progression from slides than from Flicker. For the 2 less experienced trainees Flicker was significantly more accurate.

• CONCLUSION: The MatchedFlicker software had a greater accuracy and was quicker to perform than using a handheld stereoscopic viewer. (Am J Ophthalmol 2016;167:88–95. © 2016 Published by Elsevier Inc.)

AJO.com Supplemental Material available at AJO.com.
Accepted for publication Mar 23, 2016.

From the University of Florida, Department of Ophthalmology, Gainesville, Florida (J.L.S., Z.L.L., A.M.M., A.J.L., A.G., G.M.M., B.Z., J.J.S., M.B.S.) and Ross Eye Institute, University at Buffalo, Department of Ophthalmology, Buffalo, New York (J.L.S.)

Inquiries to Mark B. Sherwood, 1600 SW Archer Rd, Suite M123D, Gainesville, FL 32610; e-mail: sherwood@ufl.edu

IT IS ESTIMATED THAT IN THE UNITED STATES MORE than 2 million people currently suffer from open-angle glaucoma (OAG).[1] Monitoring for glaucoma progression includes visual field and retinal nerve fiber analysis and careful examination of the optic nerve. All 3 methods have limitations, with quality of image scan or photograph being the primary problem with retinal nerve fiber layer (RNFL) optical coherence tomography (OCT) and disc evaluation, respectively, and subjective factors related to patient alertness, their learning curve for the field testing, technician experience, and small variations in head positioning also playing a role in visual field evaluation.

Accurate disc evaluation for focal rim defects, disc hemorrhages, optic nerve size, and nerve fiber layer abnormalities is important for identification of glaucomatous progression.[2] Photo documentation of the optic nerve head using stereo photographs has been a traditional method for ophthalmologists to monitor glaucomatous progression because of its superiority to written subjective documentation and drawings.[3] Large randomized National Eye Institute (NEI) trials have employed optic disc stereo photograph analysis using a disc reading center as one of the primary methods to detect glaucoma progression.[4–6]

Evidence suggests that structural damage can often be detected before the development of detectable visual field defects.[7] In ocular hypertensive patients that were considered normal at baseline for automated visual field and disc morphology, Kass and associates[8] in the Ocular Hypertension Treatment Study (OHTS) (full study group listed in the Appendix; Supplemental Material available at AJO.com) found that of 55% of eyes with primary open-angle glaucoma (POAG) endpoints were initially identified on stereo photographs alone, 35% by visual field changes, and 10% based on concurrent visual fields and stereo photograph changes.

It is unclear why disc evaluation through examination of sequential simultaneous stereo photographs is not done routinely in clinical practice. Possible explanations include the need for a special simultaneous-stereoscopic camera

0002-9394/$36.00
http://dx.doi.org/10.1016/j.ajo.2016.03.031

and stereo viewer (although it is possible to obtain less standardized stereoscopic photographs with a nonsimultaneous camera), and the difficulty and time-consuming nature of the stereo photograph evaluation. With the increased use of electronic medical records (EMR), it is important to have a disc imaging technique that is accurate at detecting glaucoma progression and is also well adapted to the new electronic format.

This study was designed to compare the accuracy and speed of the new MatchedFlicker (EyeIC Inc, Narberth, Pennsylvania, USA), technique using a computer screen with the traditional gold-standard method of examining slides of stereoscopic disc photographs in a group of nonexpert observers who were in various stages of ophthalmology training.

METHODS

ALL RESEARCH IN THIS EXPERIMENTAL STUDY WAS performed in accordance with the University of Florida's Institutional Review Board (IRB) with protocol approval prior to initiation of the study. Approval was given for the collection of optic nerve photographs from patient records and from the Bascom Palmer Eye Institute Optic Disc Reading Center, as well as for the participation of the photograph evaluators. A full waiver of informed consent was obtained for the use of de-identified patient optic nerve photographs, because no risk was posed to any patient with their use. This study adhered to the tenets of the Declaration of Helsinki and all federal and state laws.

From February to May of 2013, 2 ophthalmology residents (postgraduate year [PGY]-2 and PGY-3) and 1 glaucoma fellow (PGY-5) at the University of Florida independently evaluated stereoscopic pairs of disc photographs of 100 eyes taken at 2 time points. Fifty eyes were identified from patients in the OHTS study that showed glaucoma progression. This progression was previously determined by the OHTS Optic Disc Reading Group of the Bascom Palmer Eye Institute and confirmed as glaucomatous change by the OHTS Endpoint Committee. The patient's comparison photographs were taken between 1 and 5 years apart, depending on when the disc progression was first noted by the study committee. Another group of 50 eyes were obtained that were known to show no progression since the photographs were taken at the University of Florida using a Topcon TRC-50DX simultaneous stereo camera system (Topcon, Oakland, New Jersey, USA) just a few minutes apart. Duplicate photographs of 20 of the 50 eyes that showed progression and 20 of the 50 eyes with no progression were randomly selected to allow for assessment of intra- and interobserver variability in detecting progression. Thus, a total of 140 image pairs (280 photographs) were examined by each observer.

FIGURE 1. Pentax Handheld Stereo Viewer. (Pentax, Tokyo, Japan) used in study of accuracy of detecting glaucomatous progression with the traditional stereo-viewing method.

Two different examination methods were used by each observer for judging the optic disc photographs for glaucomatous progression: handheld stereo viewers and the computerized MatchedFlicker program (identified further in the study as "Flicker"). Using 2 Pentax handheld stereo viewers (Pentax, Tokyo, Japan) (Figure 1), 35 mm simultaneous-stereoscopic color slides of optic disc photographs were reviewed with the slide from the first time point in one stereo viewer and the slide from the second time point in the other stereo viewer. No poor-quality photographs were used, but photographs were of naturally slightly varying quality to reflect what a practitioner would encounter in real-world clinical practice.

The evaluators reviewed the same sets of photographs using the Flicker software. For this computerized analysis, stereoscopic images are not required. The right image of the stereo pair from each of the 35 mm stereo photographs was digitized using a Nikon Super Coolscan 5000 (Nikon, Tokyo, Japan) and uploaded into the Flicker alignment and presentation software. In 4 eyes the right image was blurred and the left image of the stereo pair was selected instead. The right image at the second time point was also digitized and uploaded into the Flicker software. The Flicker program took these 2 digitized images, registered that they were photographs of the same eye taken at 2 time points, and aligned the object (the optic disc) in the 2 images so that they were superimposed (Figure 2). The program then "flickered" rapidly between the 2 images, which simulated a sensation of movement if any structural alterations had occurred. In this way, any features that underwent a potentially glaucomatous structural change between the photographs would be detected as movement. A 16 inch Samsung Galaxy laptop (Suwon, Korea) computer screen was used for the Flicker presentations.

Using the 2 different methodologies, the evaluators assessed the optic nerve head (ONH) images for evidence of progression including disc rim thinning (focal or diffuse), vessel movement related to increased cupping, or detection of new or enlarged RNFL defects.

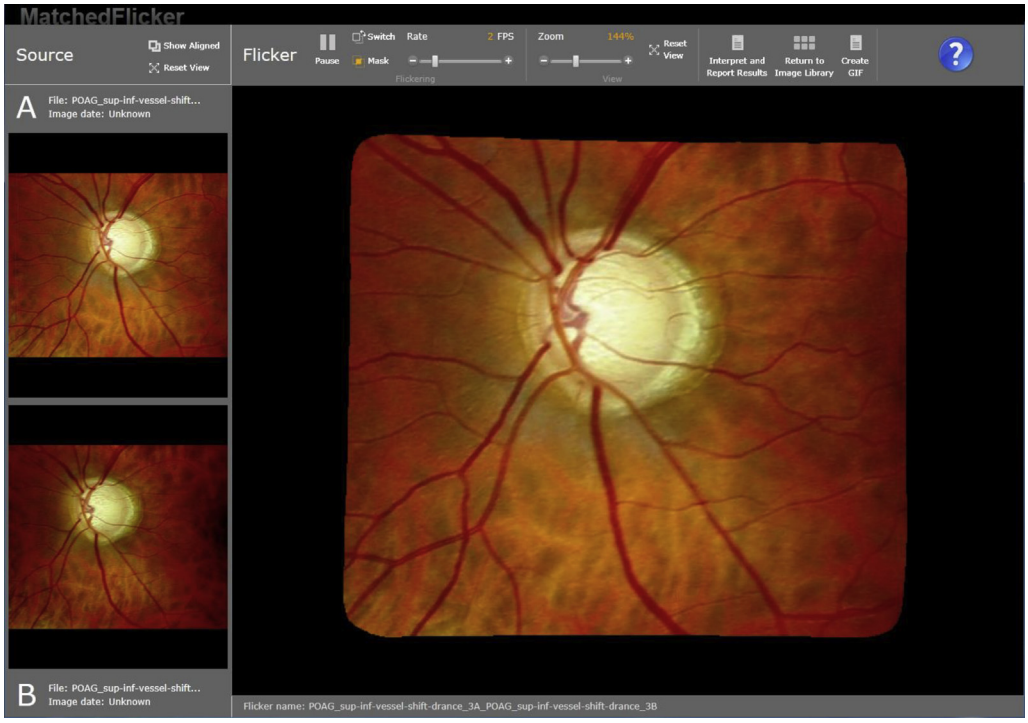

FIGURE 2. Images produced by the MatchedFlicker software. (EyeIC Inc., Narberth, Pennsylvania, USA) used in assessing accuracy of detecting glaucomatous progression with the MatchedFlicker software technique.

Prior to examining the study photographs, each evaluator was given a practice session with 10 additional photograph pairs to familiarize them with each technique. For 2 of the examiners this practice with the Flicker technique was their only exposure to the program, as they had not used it previously. All 3 had some experience reviewing disc stereo photographs with a handheld stereo viewer.

The method of examining the disc pairs was divided into 2 alternate blocks of 70 paired images, using 1 method for the entirety of that block. The method that was used first was randomized. All 280 images were reviewed separately by each of the 3 observers in a single session. A forced decision on progression or nonprogression was required for each image, even if the observer was not certain. A study coordinator recorded the evaluator's assessment of disc progression for each photograph and recorded the time taken for the overall disc determination using each method. The time taken for each evaluator to complete the set of photographs using the 2 different techniques was divided by the number of photograph pairs viewed to give a mean "time-per-assessment" measurement.

• **STATISTICAL ANALYSIS:** The difference in accuracy between the Flicker and slides method was initially analyzed with a 2-sided paired *t* test comparing each reviewer's percent accuracy for both the Flicker and slides method. A more precise analysis, in which the image pairs were scored for both Flicker and slides and the paired differences in score compared, is shown in Supplemental Table 1 (Supplemental Material at AJO.com). Each subject slide or image was scored 100% (3 of 3 reviewers correct), 66.7% (2 of 3 correct), 33.3% (1 of 3 correct), or 0% (0 of 3 correct) and the *P* value assessed by a 2-sided paired *t* test for all 100 image pairs.

The difference in the average evaluation time per assessment between the 2 methods was also analyzed with a 2-sided paired *t* test comparing the total time of each reviewer divided by all 140 slides in the block for each method.

Information on each reviewer's sensitivity, specificity, and overall accuracy was compared between the 2 methods using the following logistic regression model:

$$\begin{aligned}\text{logit}[\Pr(yijk = 1)] &= \text{Log}[\text{odds}(yijk = 1{:}\ yijk = 0)] \\ &= \alpha 0 + \alpha\text{trt}\ \text{I}(i = 1) \\ &\quad + \alpha 1\ \text{I}(k = 2) + \alpha 2\ \text{I}(k = 3)\end{aligned}$$

where *yijk* is the binary response (ie, 1 for correct and 0 for incorrect) of subject *j*, reviewed by Reviewer *k* (= 1, 2, or 3) using method i (= 1 for Flicker and = 0 for slides). Reviewer 1 (PGY-5) is used as the reference. *I* (*x* = value) is the indicator function (ie, = 1 if *x* = value and 0 otherwise). This produces for each of the groups (a, b, and c) an adjusted odds ratio for Flicker to slide, with values above 1.0 favoring Flicker and values below 1.0 favoring slides.

Sensitivity was determined by the percentage of progression-positive diagnoses that were correctly attributed to truly glaucoma-positive (ie, "true-positive" images).

TABLE 1. Comparison of Individual Reviewer Percent Accuracy When Comparing Glaucomatous Disc Change Using Stereo Disc Viewing and the MatchedFlicker Software Program

Reviewer	Slides	Flicker	*P* Value[a]
Reviewer 1	80.71%	90.00%	
Reviewer 2	70.71%	84.29%	
Reviewer 3	76.43%	88.57%	
Average	75.95%	87.62%	.011

Data compare the accuracy of each individual reviewer in determining progression of the 140 slides (50 with progression and 50 without, with 20 repeated form both groups). A more precise analysis was also done comparing the individual slides, the method of which can be viewed in Supplemental Table 1 (Supplemental Material available at AJO.com).

[a]*P* value was assessed with a 2-sided paired *t* test comparing each reviewer's percent accuracy for each viewing method.

TABLE 2. Comparison of Individual Reviewer Evaluation Time per Assessment (Seconds) When Comparing Glaucomatous Disc Change Using Stereo Disc Viewing and the MatchedFlicker Software Program

Reviewer	Stereo Viewer	Flicker	Difference[a]	*P* Value[b]
Reviewer 1	33.60	20.58	13.02	
Reviewer 2	29.38	22.84	6.54	
Reviewer 3	39.15	31.29	7.85	
Average (SD)	34.1 (4.9)	24.9 (5.6)	9.2 (3.4)	.04

Data compare each individual reviewer's time, in seconds, spent determining progression on all 140 slides presented (50 with progression and 50 without, along with 20 repeated slides from each group) while using either the stereoscopic slide viewing technique or the Flicker program.

[a]Difference (in seconds) between the time spent using each method.

[b]*P* value was assessed with a 2-sided paired *t* test comparing the 3 individual reviewer times for each method.

Similarly, specificity was determined by the percentage of progression-negative diagnoses that were correct (truly negative). The logistic regression model compares the probability of a reviewer's response being correct between the 2 methods, Flicker and slides, while controlling for reviewer variability. The model was used to compare reviewer accuracy in 3 different sets of images: (1) the full image set of 100 unique image pairs, (2) the set of 50 true positives, and (3) the set of 50 true negatives.

McNemar's χ^2 was used to assess discordant pairs in reviewer responses for the same images (inter-reviewer variability), and between the 2 responses given by reviewers for the same image using the 2 different viewing methods.

TABLE 3. Analysis of Sensitivity, Specificity, and Accuracy of Individual Reviewers Using the MatchedFlicker Software Program Versus Stereo Disc Viewing to Compare Glaucomatous Disc Change

Reviewer	Sensitivity (True Positive)[a]	Specificity (True Negative)[b]	Accuracy of All Diagnoses
MatchedFlicker program			
Reviewer 1	42/47 (89%)	45/53 (85%)	87/100 (87%)
Reviewer 2	40/46 (87%)	44/54 (81%)	84/100 (84%)
Reviewer 3	44/51 (86%)	43/49 (88%)	87/100 (87%)
Stereo disc viewer			
Reviewer 1	32/32 (100%)	50/68 (74%)	82/100 (82%)
Reviewer 2	21/21 (100%)	50/79 (63%)	71/100 (71%)
Reviewer 3	29/34 (85%)	45/66 (68%)	74/100 (74%)

Data are based on evaluation of 100 slides.

[a]Sensitivity = the percentage of positive diagnoses that were correct, or "true positive."

[b]Specificity = the percentage of negative diagnoses that were correct, or "true negative."

The logistic regression models were analyzed using the open source R statistical analysis package (GNU Software), while the McNemar test was conducted in SAS (SAS Institute), using PROC UNIVARIATE, which provides exact *P* values. All tests were 2-sided and statistical significance was qualified by whether or not $P < .05$.

RESULTS

FOR ALL 3 REVIEWERS, THE NUMBER OF CORRECT RESPONSES using Flicker was significantly higher ($P = .011$) than that achieved while using the stereo-slide viewer (Table 1 and Supplemental Table 1).

The observed time with Flicker was 9.2 seconds (27%) faster per assessment, on average, than those using conventional stereo slides ($P = .04$) (Table 2).

Specificity is the percentage of negative diagnoses that are truly negative, and sensitivity is the percentage of those diagnosed with true glaucoma progression. As a whole, reviewers were more specific and more accurate with Flicker than with slides (Table 3). Reviewers were also less sensitive with Flicker than with slides. While using Flicker, the reviewers assigned both positive and negative responses at almost an equal rate, identifying 49–54 of the 100 image pairs as negative. However, reviewers identified 66–79 of the 100 image pairs they viewed as negative for glaucomatous progression while using the slides.

According to Table 4, reviewers were significantly more accurate overall ($P = .001$) and more likely to correctly diagnose the group of 50 image pairs depicting glaucomatous progression ($P < .001$) when using the Flicker method. However, the group of 50 image pairs depicting no

TABLE 4. Comparison of Accuracy of All Reviewers Using Stereo Disc Viewing and the MatchedFlicker Software Program to Compare Glaucomatous Disc Change in the True-Positive, True-Negative, and Full Image Sets, Using a Logistic Regression Model

Subject Set[a]	Odds Ratio[b]	95% Confidence for OR	*P* Value
(a) All	1.983	1.308, 3.038	.001
(b) True-positive[c]	4.482	2.621, 7.886	<.001
(c) True-negative[d]	0.248	0.080, 0.646	.008

OR = odds ratio.

Data compare the likelihood that a slide from a certain group will be designated correctly while using either the Flicker program or the stereoscopic slide viewing method.

[a]Subject set "a" consists of all 100 unique slides, "b," the 50 true-positive slides, and "c," the 50 true-negative slides.

[b]An adjusted odds ratio above 1.0 favors Flicker for accuracy, while a ratio below 1.0 favors the slides.

[c]True-positive = the slides or image pairs whose correct diagnosis is positive.

[d]True-negative = the slides or image pairs whose correct diagnosis is negative.

TABLE 5. Comparison of Inter-reviewer Variability by Analyzing the Discordant Pairs Between Reviewers While Using Stereo Disc Viewing and the MatchedFlicker Software Program to Compare Glaucomatous Disc Change

Reviewers Compared	Ratio[a] (Flicker)	*P* Value (Flicker)	Ratio[a] (Slide)	*P* Value[b] (Slide)
1: 2	14:11	0.55	15:4	.012
1: 3	6:6	1.00	16:8	.10
2: 3	8:11	0.49	10:13	.53

Data compare each reviewer's performance against that of his peers by determining who had the correct designation when 2 reviewers had disagreeing designations for a slide.

[a]Ratio = Discordant pairs correct with first reviewer: Discordant pairs correct with second reviewer.

[b]*P* values were assessed by the McNemar χ^2 2-sided test.

TABLE 6. Comparison of Intra-reviewer Variability for Each Individual Reviewer While Using Stereo Disc Viewing and the MatchedFlicker Software Program to Compare Glaucomatous Disc Change

Method	Reviewer	Repeated Negative Controls (20)	Repeated Positive Controls (20)	Total
Flicker	Reviewer 1	17/20	15/20	32/40
	Reviewer 2	15/20	12/20	27/40
	Reviewer 3	15/20	16/20 (1)[a]	31/40 (1)[a]
Slide	Reviewer 1	19/20	15/20 (5)[a]	34/40 (5)[a]
	Reviewer 2	20/20	13/20 (8)[a]	33/40 (8)[a]
	Reviewer 3	20/20	14/20 (5)[a]	34/40 (5)[a]

Data compare each individual reviewer's designation variability for the 40 slides that were repeated randomly throughout the trials.

[a]Data in parentheses indicate the number of times there was agreement on the incorrect diagnosis. Such agreement occurred 18 times for the slides and only once for Flicker.

glaucomatous progression were significantly more likely to be designated correctly by reviewers using the slides than by those using Flicker ($P = .008$) (Table 4). This difference from Table 3 is driven, as mentioned above, by the relative reluctance of the reviewers to call a positive progression with slide viewing, so that although they correctly called nearly all of the true negatives, there was also a group of true positives that were designated as negative by the reviewers, thus decreasing the percentage of negatives that were actually correctly called.

Reviewers 2 and 3 (PGY-2 and PGY-3, respectively) had significantly more instances where they could correctly decide whether there was glaucomatous progression between a set of image pairs while using Flicker, but were unable to decide correctly while using the slides for the same image pair ($P = .043$ and $P = .026$ for Reviewer 1 and Reviewer 2, respectively) (Supplemental Table 2, Supplemental Material at AJO.com). While results were not significant, Reviewer 1 (PGY-5) also had more instances of correctly deciding whether or not an image pair showed progression when using Flicker.

Analysis of the discordant pairs between any 2 reviewers while using Flicker revealed no cases of 1 reviewer having significantly more correct decisions on glaucomatous progress in image pairs than his cohort (Table 5). However, it should be noted that when analyzing the cases of discordant responses while using slides, Reviewer 1 and Reviewer 2 (PGY-5 and PGY-2) differed significantly in the number of correct decisions on glaucomatous progress, with Reviewer 1 making a significantly greater amount of correct decisions ($P = .012$).

Intraobserver agreement was generally high between the 2 techniques when analyzing the 40 duplicates, with a mean of 30 and 33.67 of the 40 repeated images (75% and 84%) being given the same designation with the use of Flicker and slides, respectively (Table 6). While using the slides, there were 18 cases out of 42 agreements where the wrong decision on the presence of glaucomatous progression was made twice. In all 18 cases, the reviewers determined that there was no progression when there actually was. While using Flicker, there was only 1 case of 43 agreements where the wrong determination of glaucomatous progression was chosen twice. The error made was the same as with the slides: the reviewer saw no progression between the 2 image pairs, when progression had actually taken place.

DISCUSSION

A SIGNIFICANT PROBLEM IN THE MANAGEMENT OF GLAUcoma is the difficulty in detecting progression so that adequate therapy can be instituted. Xin and associates[9] have noted the importance of following both structural and functional parameters in combination when monitoring for progression, because frequently these do not change at the same time. Disc changes commonly precede visual field loss in early glaucoma.[8,10,11]

Examination of the optic disc is an acquired skill and is underperformed. The recommended American Academy of Ophthalmology (AAO) Preferred Practice Pattern guidelines for follow-up of glaucoma and evaluation of optic nerve and visual field assessment are from 3 months to 2 years at a maximum.[12] Hertzog and associates[13] found that only 23.3% of patients had an optic nerve head drawing or photograph within 15 months of the most recent visit. In addition, the authors found that 37.8% of patient charts had neither an optic nerve head drawing nor a photograph documented after the initial visit. Friedman and associates[14] noted in a retrospective cohort study of 1712 diagnosed suspects and 3623 diagnosed glaucoma patients enrolled in a single comprehensive insurance program that after a median follow-up of 440 days, while 46% had a billed visual field testing only 13% had some sort of optic nerve head imaging.

It can be a time-consuming process to accurately and thoroughly examine each disc every year. In addition, many clinicians do not have stereoscopic disc photographs or computerized image analysis systems for examining discs for progression and have to rely on either old notes describing the nerve or hand-drawings or nonstereoscopic photographs of varying quality. The AAO Preferred Practice Pattern suggests that these are less desirable alternatives.[15]

With 243 ophthalmologist participants, Reus and associates[10] found the accuracy for diagnosing glaucomatous discs by reviewing stereoscopic optic disc photographs using stereo viewers to be at best 80.5%. A different study by Breusegem and associates[11] looked at the interobserver agreement and accuracy of glaucoma experts and nonexpert ophthalmologists. They found that the interobserver agreement of nonexpert ophthalmologists was significantly lower than that of the experts.

As with any imaging modality, correctly interpreting the image is dependent on the quality; whether it is well centered, well focused, well illuminated, without motion artifact, and without medial opacities. Poor-quality images reduce the ability to detect subtle changes, and therefore subtle glaucomatous changes may be overlooked.[16] Our study included good-quality photographs from the OHTS Disc Reading Center and from the University of Florida clinical archives, while possessing real-life characteristics such as ill-defined sloping of some disc rims or disc tilting that would be regularly encountered by physicians in a clinical setting.

To the best of our knowledge, this is the first study to compare 2 different methods for accuracy of detecting glaucomatous disc progression over time using photographs with a known status in regard to progression. The negative controls in which simultaneous stereo photographs were taken within a few minutes of each other can reliably be stated to show no progression. It is more difficult to be certain that subtle disc progression has definitely occurred, but by using the results of a well-recognized NEI study disc reading center, our study aimed to get as close as possible to being certain that there is real progression. Indeed, this system of having a disc reading center review stereoscopic disc photographs is the current gold standard for detecting disc progression for NIH-sponsored glaucoma clinical trials. We considered using slides of nonprogressing discs from the OHTS study disc reading center as our negative control in view of the concern regarding increased lighting and other photographic differences when multiple clinics take photographs over longer time periods. However, we felt that there was some risk that discs that were graded in OHTS as not showing progression by the disc reading center might in fact have a minute number of subtle signs of progression, thus introducing a type II error (false negative). Although there are pros and cons of each negative control, on balance, we felt that using slides taken a few minutes apart and that were certain to have no progression would be better.

A recent study by Radcliffe and associates[16] viewed stereo photograph image sets from patients with glaucoma using an automated alternation flicker (the Flicker software program) and found retinal blood vessel shifts in 26.4% of the longitudinally followed glaucomatous eyes. In addition, they found a statistically significant association between blood vessel movement and increased rate of visual field progression and neuroretinal rim loss.

Our study compares the new Flicker software program viewed on a standard computer screen with the traditional technique of examining a 35 mm slide in a handheld stereo viewer, for detecting glaucomatous optic disc structural progression. Our 3 observers included 2 resident ophthalmologists at different stages of their training and a glaucoma fellow at the University of Florida. The primary endpoint was the accurate identification of disc progression over time. A higher percentage of correct answers was found using the Flicker computerized program for all 3 reviewers ($P = .01$) (Table 1, Supplemental Table 1). There were no statistically significant differences in accuracy between the trainees with different levels of experience using Flicker, but when using slides the resident with the least experience examining stereo photographs was significantly less accurate than the glaucoma fellow (Table 5). This may reflect the fact that all 3 observers had a similar, limited experience using the Flicker program with just the same 10 training photograph pairs.

As mentioned previously in the statistical analysis section, specificity is the percentage of negative designations that are truly negative and sensitivity is the proportion of positive diagnoses that are truly positive (ie, have true glaucomatous progression). In our opinion, both are important, but perhaps having high specificity is of particular importance so that patients who are progressing are detected and can receive treatment. According to Table 3, the Flicker method produced more specific results (negative designations that are truly negative) than did slides. The slides method produced more sensitive results (positive designations that are truly positive) than Flicker, owing in part to the low number of positive designations that were made. The low number of image pairs that were identified as showing glaucomatous progression (Table 3) suggests that reviewers were less inclined to call the glaucomatous progression between a set of stereoscopic image pairs while using the slides method. For all 3 trainee observers, Flicker was more accurate for the true-positive image pairs ($P = .001$), less accurate for the true-negative image pairs ($P = .008$), and, regarding the full set of 100 unique image pairs, more accurate overall than stereoscopic slide review ($P = .001$) (Table 4). The test-retest reproducibility of detecting optic disc deterioration from stereo photographs by the certified graders at the Optic Disc Reading Center in the Ocular Hypertension Treatment Study was reported to show a specificity of 98%–100% and a sensitivity ranging from 64% to 81%.[17] Our 3 trainee reviewers had a similar specificity of 95%–100% and a sensitivity of 65%–75% for the 40 duplicate slides they examined (Table 6).

With the demand for the clinician to see higher volumes of patients, the time taken to perform an accurate analysis is important. All 3 observers using Flicker were faster than using 35 mm slides ($P = .04$) (Table 2). Integration with the EMR will also be increasingly important over time. Scanning of digitized stereo images into the EMR and stereo viewing of these images usually requires specialized viewing devices and a specialized camera to obtain a simultaneous stereo image. The Flicker software has the advantage of using monoscopic images and does not require the use of stereo photography.

The current study only compared physicians at different stages of training. A future study comparing Flicker vs viewing of stereo photographs among glaucoma experts (subspecialists with multiple years of experience using stereo images) to determine the accuracy of detecting glaucoma progression would be useful to fully explore the role of Flicker analysis in clinical practice.

Our study with ophthalmologists-in-training demonstrated that the use of MatchedFlicker software is comparable or even better in accuracy than the traditional methods of stereo image optic disc viewing and was faster to perform. By using a technique for evaluating optic discs that is easier, faster, and potentially more accurate than the current gold standard, and in particular does not require stereo imaging and is suitable for EMR use, it is felt that ophthalmology residents would be more likely to regularly track the optic disc structure of their glaucoma patients, as well as visual fields and RNFL OCT. This technique has the potential to assimilate more easily into a clinical practice setting with EMR, and in conjunction with other functional and structural tests would be a good complement to the current methods of glaucoma progression detection.

THIS STUDY WAS SUPPORTED IN PART BY AN UNRESTRICTED GRANT FROM RESEARCH TO PREVENT BLINDNESS (NEW YORK, NY), to the Department of Ophthalmology, University of Florida, and NIH grant 1UL1TR000064 from the National Center for Advancing Translational Sciences (Bethesda, MD). Financial disclosures: The following authors have no financial disclosures: Jamie L. Schaefer, Zachary L. Lukowski, Alissa M. Meyer, Anthony J. Leoncavallo, Anthony Greer, Gina M. Martorana, Baiming Zou, Jonathan J. Shuster, and Mark B. Sherwood. All authors attest that they meet the current ICMJE criteria for authorship.

We gratefully acknowledge the Ocular Hypertension Treatment Study group for providing the photo pairs in the positive control group, and to Mae Gordon and Michael Kass for reviewing the manuscript and their helpful comments.

REFERENCES

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020". *Br J Ophthalmol* 2006;90(3):262–267.
2. Mackenzie PJ, Mikelberg FS. Evaluating optic nerve damage: pearls and pitfalls. *Open Ophthalmol J* 2009;3:54–58.
3. Jamara RJ, Denial A, Valentini D, Thorn F. Clinical quality assessment using computer monitor photoimages of optic nerve head cupping. *Optom Vis Sci* 2000;77(8):433–436.
4. Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120(6):714–720.
5. Leske MC, Heijl A, Hyman L. Bengtsson. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999;106(11):2144–2153.
6. Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology* 1999;106(4):653–662.
7. Xu G, Weinreb RN, Leung CK. Optic nerve head deformation in glaucoma: the temporal relationship between optic nerve head surface depression and retinal nerve fiber layer thinning." *Ophthalmology* 2014;121(12):2362–2370.
8. 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(6):701–713.

9. Xin D, Greenstein VC, Ritch R, Liebmann JM, DeMoraes CG, Hood DC. A comparison of functional and structural measures for identifying progression of glaucoma. *Invest Ophthalmol Vis Sci* 2011;52(1): 519–526.
10. Reus NJ, Lemij HG, Garway-Heath DF, et al. Clinical assessment of stereoscopic optic disc photographs for glaucoma: the European Optic Disc Assessment Trial. *Ophthalmology* 2010; 117(4):717–723.
11. 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.
12. American Academy of Ophthalmology Glaucoma Panel. Preferred Practice Pattern ® Guidelines. Primary Open-Angle Glaucoma. San Francisco, CA: American Academy of Ophthalmology; 2010. Available at: www.aao.org/ppp. Accessed December 10, 2015.
13. Hertzog LH, Albrecht KG, LaBree L, Lee PP. Glaucoma care and conformance with preferred practice patterns. Examination of the private, community-based ophthalmologist. *Ophthalmology* 1996;103(7):1009–1013.
14. Friedman DS, Nordstrom B, Mozaffari E, Quigley HA. Glaucoma management among individuals enrolled in a single comprehensive insurance plan. *Ophthalmology* 2005;112(9): 1500–1504.
15. Shaffer RN, Ridgway WL, Brown R, Kramer SG. The use of diagrams to record changes in glaucomatous disks. *Am J Ophthalmol* 1975;80(3 Pt 1):460–464.
16. Radcliffe NM, Smith SD, Syed ZA, et al. Retinal blood vessel positional shifts and glaucoma progression. *Ophthalmology* 2014;121(4):842–848.
17. Parrish RK II, Schiffman JC, Anderson DR, et al. Test-retest reproducibility of optic disk deterioration detected from stereophotographs by masked graders". *Am J Ophthalmol* 2005; 140(4):762–764.