Title

Detecting Glaucoma Using Automated Pupillography

Journal

Ophthalmology, 121(6)

ISSN

0161-6420

Authors

Tatham, Andrew J

Meira-Freitas, Daniel

Weinreb, Robert N

et al.

Publication Date

2014-06-01

DOI

10.1016/j.ophtha.2013.12.015

Peer reviewed

Published in final edited form as:
Ophthalmology. 2014 June ; 121(6): 1185–1193. doi:10.1016/j.ophtha.2013.12.015.

Detecting Glaucoma Using Automated Pupillography

Andrew J. Tatham, FRCOphth, **Daniel Meira-Freitas, MD, PhD**, **Robert N. Weinreb, MD**, **Linda M. Zangwill, PhD**, and **Felipe A. Medeiros, MD, PhD**
Hamilton Glaucoma Center and Department of Ophthalmology, University of California, San Diego, La Jolla, California

Abstract

Objective—To evaluate the ability of a binocular automated pupillograph to discriminate healthy subjects from those with glaucoma.

Design—Cross-sectional observational study.

Participants—Both eyes of 116 subjects, including 66 patients with glaucoma in at least 1 eye and 50 healthy subjects from the Diagnostic Innovations in Glaucoma Study. Eyes were classified as glaucomatous by repeatable abnormal standard automated perimetry (SAP) or progressive glaucomatous changes on stereophotographs.

Methods—All subjects underwent automated pupillography using the RAPDx pupillograph (Konan Medical USA, Inc., Irvine, CA).

Main Outcome Measures—Receiver operating characteristic (ROC) curves were constructed to assess the diagnostic ability of pupil response parameters to white, red, green, yellow, and blue full-field and regional stimuli. A ROC regression model was used to investigate the influence of disease severity and asymmetry on diagnostic ability.

Results—The largest area under the ROC curve (AUC) for any single parameter was 0.75. Disease asymmetry ($P < 0.001$), but not disease severity ($P = 0.058$), had a significant effect on diagnostic ability. At the sample mean age (60.9 years), AUCs for arbitrary values of intereye difference in SAP mean deviation (MD) of 0, 5, 10, and 15 dB were 0.58, 0.71, 0.82, and 0.90, respectively. The mean intereye difference in MD was 2.2±3.1 dB. The best combination of parameters had an AUC of 0.85; however, the cross-validated bias-corrected AUC for these parameters was only 0.74.

Conclusions—Although the pupillograph had a good ability to detect glaucoma in the presence of asymmetric disease, it performed poorly in those with symmetric disease.

Glaucoma is characterized by progressive optic neuropathy, loss of retinal ganglion cells, and associated loss of visual field.[1] Glaucomatous visual field loss is irreversible, and therefore, early detection and diagnosis is important. Although standard automated

Correspondence: Felipe A. Medeiros, MD, PhD, Hamilton Glaucoma Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0946. fmedeiros@glaucoma.ucsd.edu.

Financial Disclosure(s):
The author(s) have made the following disclosure(s):

perimetry (SAP) is the most common method for detecting glaucomatous functional damage, it has several limitations. For example, SAP is influenced by patient fatigue, inattention, and learning effects, with the result that test results may be unreliable.[2,3] Furthermore, because of its variability, multiple subsequent confirmatory tests often are required to provide confidence that a defect seen on SAP is genuine.[3]

An additional problem with SAP concerns the logarithmic decibel scale used for the assessment of sensitivity thresholds and presentation of results.[4] The logarithmic scale leads to compression of data in early disease, meaning large retinal ganglion cell losses may occur before there is a significant reduction in SAP sensitivity.[5,6]

Previous studies have reported that patients with glaucoma frequently have abnormal pupillary light responses. For example, a recent systematic review found that relative afferent pupillary defect (RAPD) can be detected in one third to two thirds of glaucomatous patients using the swinging flashlight test.[7-9] An RAPD is an important marker of asymmetric impairment of the afferent visual system and, in a subject with glaucoma, is likely to indicate significant damage to the optic nerve.[10,11] Moreover, the magnitude of the pupillary response defect has been shown to correlate with the severity of disease.[12-14]

The detection of pupil response abnormalities in glaucomatous subjects raises the possibility that assessment of pupillary reactions may be a useful functional test for detecting the disease. The swinging flashlight test is a simple method of measuring the pupillary light response; however, it is subject to interobserver variation. Also, the RAPD, which is primarily a measure of asymmetry, is unlikely to be of high diagnostic value in a predominantly bilateral disease such as glaucoma.[15] A more complete assessment of the pupil response parameters is possible using computerized automated pupillography.[10] Automated pupillography is objective, and it also allows averaging of multiple responses and quantification of parameters not visible to an observer. The result is that automated pupillography is able to detect pupil abnormalities with greater sensitivity than flashlight examination.[10,15,16] A further advantage of automated pupillography is that novel testing strategies such as stimulation of specific parts of the retina within the same eye may be used.[17]

The aim of this study was to evaluate the ability of a new binocular automated pupillograph (RAPDx pupillograph; Konan Medical USA, Inc., Irvine, CA) to discriminate healthy subjects from those with glaucoma. The device measures bilateral pupil responses to monocularly presented stimuli. Parameters evaluated included the overall pupil response to a variety of white and colored full-field stimuli, the pupil response to stimuli directed at different parts of the retina, and the relative difference in pupil response between eyes. The effect of disease severity and disease asymmetry on the diagnostic ability of the pupil response parameters also was examined.

Methods

Study Participants

This was a cross-sectional study involving participants from a prospective longitudinal study designed to evaluate optic nerve structure and visual function in glaucoma, the Diagnostic Innovations in Glaucoma Study at the University of California, San Diego (UCSD). Methodologic details have been described previously.[18] The UCSD Institutional Review Board and Human Subjects Committee approved all protocols, and methods adhered to the Declaration of Helsinki.

At each visit subjects underwent a comprehensive ophthalmologic examination including visual acuity, slit-lamp biomicroscopy, intraocular pressure measurement, gonioscopy, dilated funduscopic examination, simultaneous stereoscopic optic disc photography (Kowa WX3D; Kowa OptiMed, Inc., Torrance, CA), and SAP (Humphrey Field Analyzer II-I; Carl Zeiss Meditec, Inc., Dublin, CA) using the Swedish interactive threshold algorithm (standard 24-2). Only subjects with open angles on gonioscopy were included. Subjects were excluded if they had a best-corrected visual acuity of less than 20/40; spherical refraction outside ±5.0 diopters, cylinder correction outside 3.0 diopters, or both; or any other ocular or systemic disease that could affect the optic nerve or the visual field. Details of the methodology used to grade optic disc photographs at the UCSD Optic Disc Reading Center have been provided elsewhere.[18-20]

The study included both eyes of 66 subjects with glaucoma in at least 1 eye and both eyes of 50 healthy controls. Eyes were classified as glaucomatous if they had repeatable (≥2 consecutive) abnormal SAP test results or progressive glaucomatous changes on masked grading of stereophotographs, with or without abnormal SAP results. Abnormal SAP results were defined by a pattern standard deviation outside the 95% confidence limits or glaucoma hemifield test results outside the reference range. Healthy subjects were recruited from the general population through advertisements and from the staff and employees of UCSD. Healthy eyes had intraocular pressure of less than 22 mmHg with no history of increased intraocular pressure and normal SAP results.

Pupil responses may be influenced by medications affecting the autonomic nervous system; therefore, patients using systemic or topical cholinergic or anticholinergic medications, including pilocarpine and atropine, were not included in the study. We also excluded subjects who had undergone previous intraocular surgery, with the exception of uncomplicated cataract extraction. Each patient underwent SAP, Cirrus high-definition optical coherence tomography (OCT; software version 6.0; Carl Zeiss Meditec, Inc.), and automated pupillography, with each subject undergoing all tests within a 6-month period.

Standard Automated Perimetry and Imaging

All visual fields were evaluated by the UCSD Visual Field Assessment Center.[21] Visual fields with more than 33% fixation losses or false-negative errors or more than 15% false-positive errors were excluded. The only exception was the inclusion of fields with false-negative errors of more than 33% when the field showed advanced disease. Visual fields exhibiting a learning effect (i.e., initial tests showing consistent improvement on visual field

indices) also were excluded. Visual fields were reviewed further for the following artifacts: eyelid and rim artifacts, fatigue effects, inappropriate fixation, inattention, and evidence that the visual field results were caused by a disease other than glaucoma.

Retinal nerve fiber layer (RNFL) thickness measurements were acquired using the Cirrus high-definition OCT optic disc cube 200×200 protocol, the details of which have been described previously.[22] The parapapillary RNFL thickness measurements were calculated from a 3.46-mm-diameter circular scan (10.87-mm length) automatically placed around the optic disc. The Cirrus high-definition OCT images were included if the signal strength was 7 or more, movement artifacts were absent, and there was good centering on the optic disc.

Automated Pupillography

Pupil responses were tested using the RAPDx pupillograph, a new binocular infrared computerized pupillograph. The device measures bilateral pupil responses to a sequence of monocularly presented visual stimuli. Stimuli are generated using a single liquid crystal display (LCD) screen that has a central physical barrier to create 2 optical channels. The screen displays a target (green cross) for patient fixation, and during testing, each portion can be enabled selectively to achieve separate stimulation of each eye. The screen is viewed at infinity through a pair of 50-mm objective lenses for a field of view of approximately 25° in each eye. Eyes also are illuminated by a pair of infrared-emitting diodes, with peak emission at 880 nm, mounted at a 35° angle to each other.

Under infrared conditions information regarding the so-called dark pupil diameter is captured as the number of camera pixels, and this measurement is converted to millimeters using a scaling factor. Stimuli then are presented to each eye, with the size, color, intensity, and length of time of each stimulus varied automatically using proprietary software. The device incorporates pupil tracking and blink detection systems using 60 full-frames/second digital cameras, each with a resolution of 240×240 pixels/frame, for approximately 25 pixels/mm. If a blink obscures the pupil during recording, the test is repeated automatically.

Pupillograph Stimuli—The pupillograph LCD screen produces white, red (peak 605 nm), green (peak 555 nm), yellow (peak 576 nm), and blue (peak 440 nm) stimuli, which are varied to include full-field, central, peripheral, and quadrant testing of each eye. The LCD color chromaticity coordinates are x = 0.313, y = 0.329 (white); x = 0.645, y = 0.341 (red); x = 0.312, y = 0.625 (green); and x = 0.153, y = 0.053 (blue). The background consists of a very low white setting. Each examination is broken into subsections referred to as *tests*, which are composed of a set of *trials*. Each trial consists of a period of stimulation followed by a period of darkness (nominal background luminance of 0.01 cd/m^2), during which the cameras record continuously. The total time of each trial is 2000 ms, plus a 100-ms posttrial period during which no images are acquired (Table 1). The right eye is stimulated first, followed by the left, then the right, with continued stimulation alternating between eyes. The full-field white stimulus test consists of 18 trials, and this is followed by full-field trials using the red, green, blue, and then yellow stimuli, with each full-field stimulus extending to approximately 18° from fixation.

The central test consists of a small, 3° centrally aimed stimulus presented at low (dim) and high (bright) intensities as 2 separate tests. The central test is repeated for each colored stimulus. The peripheral test consists of a peripheral macular-sparing annulus stimulus, ranging from approximately 10° to 25°, which is also presented at dim and bright intensities for each colored stimulus. Finally, a pair of superior nasal quadrant and inferior nasal quadrant tests, each with approximately a 10° macular sparing, is presented at high intensity.

Testing for this study was conducted under dark room conditions with an illuminance of less than 0.5 lux. The full-field stimuli had luminances of 384 cd/m^2 (white stimulus), 88 cd/m^2 (red stimulus), 27 cd/m^2 (green stimulus), 23 cd/m^2 (blue stimulus), and 380 cd/m^2 (yellow stimulus), and during all tests, there was a nominal background luminance of 0.01 cd/m^2. The white dim peripheral stimulus had a luminance of 24 cd/m^2, compared with 118 cd/m^2 for the white bright peripheral stimulus. The white dim central stimulus had a luminance of 0.17 cd/m^2, compared with 0.28 cd/m^2 for the white bright central stimulus. Subjects did not wear corrective lenses during testing.

Pupil Response Parameters—The pupillograph generates pairs of simultaneous biometric waveforms that describe the average pupil responses (combined right and left eye) to the monocular stimuli. Software is incorporated to parse the pupil diameter waveforms into specific metrics. The median value from the series of repetitions is calculated to minimize noise. Parameters measured by the pupillograph include response amplitude (RespAmp), which is the maximum contraction of the pupil as a percentage of the prestimulation size, that is, the prestimulus pupil diameter minus the minimum pupil diameter, divided by the prestimulus pupil diameter; response latency (RespLat), which is the time in milliseconds between stimulus onset and time when pupil velocity has reached 50% of the peak velocity of constriction; maximum velocity of constriction (ConMaxVel); latency of the maximum velocity of constriction (ConMaxVelLat); maximum velocity of dilation (DilMaxVel); and latency of the maximum velocity of dilation (DilMaxVelLat).

In this study, 13 response parameters were obtained for each stimulus; therefore, there were a total of 65 parameters using the 5 full-field stimuli (white, red, green, blue, yellow) alone. When reporting the results, each response parameter name was appended with the letter *R* or *L* to indicate the value was obtained with stimulation of the right or left eye, respectively, and with the letter *W*, *R*, *G*, *B*, or *Y* to indicate whether the stimulus was white, red, green, blue, or yellow, respectively.

An index of the direction and magnitude of pupil response asymmetry, known as the RAPD score, also is generated automatically by the pupillograph. The RAPD score is calculated as the difference in the amplitude of pupil constriction between stimulation of the 2 eyes using the following formula[23]:

$$RAPD\ score = 10 \times log_{10}\,(od/os),$$

where *od* is the mean response amplitude of both pupils in response to right eye stimulation and *os* is the mean response amplitude in both pupils in response to left eye stimulation. A positive value indicates a relative abnormality of the left afferent system and a negative

value indicates a relative abnormality of the right afferent system.[23] The absolute value of the RAPD score was provided as an overall marker of asymmetry, without regard to the laterality of the defect for each stimulus.

Statistical Analysis

Receiver operating characteristic (ROC) curves were constructed to assess the ability of each pupillograph parameter to distinguish a subject with glaucoma in at least 1 eye from a healthy subject. A ROC curve is a plot of the true-positive rate versus the false-positive rate for all possible cutpoints. The area under the ROC curve (AUC), adjusted for age differences between cases and controls, was used to summarize the diagnostic accuracy of each parameter. An AUC of 1.0 represents perfect discrimination, whereas an area of 0.5 represents chance discrimination.[24] Initially unadjusted pooled AUCs were calculated for each stimulus and pupil response parameter.

The single pooled ROC gives the average performance of the test in the population; however, it is important to know how the test performs in subgroups of patients, for example, those with early disease. For this purpose, a ROC regression model was fitted to investigate the influence of disease severity (as measured by a subject's average SAP mean deviation [MD]) on the diagnostic ability of the best pupillograph parameter. The ROC regression methodology was proposed originally by Pepe[25] and Alonzo and Pepe[26] and was applied previously to the evaluation of glaucoma diagnostic tests by Medeiros et al[27] and Leite et al.[28] The method allows evaluation of the influence of covariates on the diagnostic performance of the test, so that ROC curves for specific values of the covariate of interest can be obtained. Briefly, the model can be written as:

$$ROC_{X,XD}(q)=\Phi\left[\alpha_1+\alpha_2\Phi^{-1}(q)+\beta X+\beta_D X_D\right]$$

where $ROC_{X,XD}$ (q) is the probability that a subject with disease, under the effect of a disease-specific covariate, *X*D, and a common covariate, *X*, will have results that are greater than or equal to the *q*th of the distribution of test results from subjects without disease. That is, when the specificity of the test is 1–*q*, $ROC_{X,XD}$ (q) represents sensitivity. The intercept and the slope of the ROC curve are represented by the coefficients α1 and α2, respectively. Φ Represents the normal cumulative distribution function. If the coefficients (β) for disease-specific (*X*D) or common (*X*) variables are more than 0, then the variable positively influences the discrimination between subjects with disease and between normal and diseased, respectively. A negative coefficient indicates an inverse relationship. In this study, the ROC regression models were fitted with severity (average eye MD) and disease asymmetry (intereye difference in MD) used as the disease-specific covariates. Estimation of AUCs for arbitrary values of severity and asymmetry were obtained using the formula:

$$AUC=\int_0^1 ROC_{X,XD}(q)\,dt=\Phi\left(\frac{\alpha_1+\beta X+\beta_D X_D}{\sqrt{1+\alpha_2^2}}\right)$$

Confidence intervals for regression parameters were obtained using a bootstrap resampling procedure (n = 1000 resamples).

The pupillograph produces many potentially useful parameters, and a combination of parameters may have better diagnostic ability than the best-performing single parameters. Stepwise logistic regression was used to identify parameters with the best discriminatory ability and develop a classification function.[29] Only variables with a *P* value less than 0.05 were retained in the final model. Because of potential model overfitting, 5-fold cross-validation was performed to obtain a bias-corrected estimate of the diagnostic ability of the combination of pupillograph parameters, that is, the best combination of parameters was determined using four fifths of the sample and then tested using the remaining one fifth of the sample, repeated 5 times. All statistical analyses were performed with commercially available software (STATA version 12; StataCorp LP, College Station, TX). The α level (type I error) was set at 0.05.

Results

The study included both eyes of 116 subjects, including 66 patients with glaucoma in at least 1 eye, and 50 healthy subjects. Sixty-one of 116 subjects (53%) were female and 28 subjects (24%) were of African ancestry. Of those with glaucoma, 26 had glaucoma in both eyes, 29 had glaucoma in 1 eye and suspected glaucoma in the other, and 11 had glaucoma in 1 eye and a healthy fellow eye. Healthy subjects were significantly younger than patients with glaucoma ($P < 0.001$). The demographic and clinical characteristics of participants are summarized in Table 2.

Subjects with glaucoma had significantly lower average MD and RNFL thickness than healthy subjects ($P < 0.001$ for both comparisons) and had significantly greater intereye differences in MD and RNFL thickness ($P < 0.001$ for both comparisons). The distribution of disease severity, measured using MD, is shown for the glaucomatous eyes in Figure 1. There was a large variation in MD in the glaucomatous eyes, ranging from −25.4 to 1.2 dB. The average of right and left eye MD for subjects with glaucoma in at least 1 eye ranged from −24.0 to 1.1 dB.

Overall Pupil Response to White and Colored Full-Field Stimuli

Table 3 shows the pooled AUCs, adjusted for age, for the pupillograph parameters obtained using the full-field white stimulus. Although there were significant differences in all measured pupil parameters between healthy and glaucomatous subjects, none of the individual parameters had particularly good diagnostic ability. The white full-field stimuli parameter with the best AUC was constriction maximum velocity latency on left eye stimulation (ConMaxVelLatL, 0.65), followed by response amplitude on right-eye stimulation (RespAmpR, 0.60) and RAPD score (absolute value of the RAPD score, 0.60). The pooled AUCs for the pupil responses to full-field red, green, blue, and yellow stimuli are shown in Table 4 (available at http://aaojournal.org). The parameters with the best discriminatory ability were RespLatL with the blue stimulus (AUC, 0.73), ConMaxVelLatR with the blue stimulus (AUC, 0.74), ConMaxVelLatR with the yellow stimulus (AUC,

0.74), ConMaxVelLat with the green stimulus (AUC, 0.75 on right eye stimulation and AUC, 0.74 on left eye stimulation) and RespLatL with the green stimulus (AUC, 0.73).

Pupil Response to Stimuli Directed at Different Areas of the Retina

Table 5 shows the diagnostic ability of the best response parameters obtained using dim and bright peripheral and central white stimuli. Although glaucomatous and healthy subjects had significant differences in pupil responses obtained using these stimuli, none of the parameters had a larger AUC than the best-performing full-field stimulus. Similar results were found for pupil responses to stimuli directed at the superior and inferior nasal quadrants.

Relative Difference in Pupil Response between Eyes

The diagnostic ability of the relative difference in pupil response between eyes was assessed by generating RAPD scores for latency and amplitude of the pupil response to bright and dim full-field, central, and peripheral stimuli for each of the colored stimuli. There was moderate correlation between the full-field white stimulus RespAmp RAPD score and intereye difference in RNFL thickness ($R^2 = 0.306$; $P < 0.001$) and intereye difference in MD ($R^2 = 0.431$; $P < 0.001$). The relationship between RAPD scores using colored stimuli and in intereye difference in RNFL thickness and MD was weaker. For example, RAPD score for RespAmp bright full-field blue stimulus had an R^2 of 0.175 and 0.153 for MD and RNFL thickness, respectively. The parameters with the greatest AUCs were RAPD score for RespAmp bright full-field yellow stimulus (AUC, 0.70) and RAPD score for RespAmp bright full green stimulus (AUC, 0.69).

Effect of Disease Severity and Asymmetry

The effect of disease severity (average SAP MD) and disease asymmetry (intereye difference in MD) on the diagnostic ability of the pupillograph was assessed using ROC regression of the best performing full-field stimuli. Table 6 illustrates the results for the best performing parameter using the white stimulus (ConMaxVelLatL). There was a tendency for the pupillograph to perform better with increasing disease severity, as measured by the average eye MD ($\beta_1 = -0.056$); however, this was not statistically significant ($P = 0.058$). Figure 2A shows the ROC curves for arbitrary values of average MD of −5, −10, and −15 dB with age at the sample mean (60.9 years). The AUCs for these values of average MD were 0.68, 0.75, and 0.81, respectively. There was also no significant relationship between average eye RNFL thickness and diagnostic ability ($\beta_1 = -0.011$; $P = 0.305$). In contrast, the relationship between diagnostic ability and disease symmetry was significant, with an increase in AUC with increasing disease asymmetry ($\beta_1 = 0.098$; $P < 0.001$). Figure 2B shows the ROC curves for ConMaxVelLatL for arbitrary values of intereye difference in MD of 5, 10, and 15 dB with age at the sample mean. The AUCs for these values of intereye difference in MD were 0.71, 0.82, and 0.90, respectively. The AUC in the presence of symmetric disease, when the intereye difference in MD was 0 dB, was only 0.58. There was a similar relationship between the degree of disease asymmetry and the discriminatory ability of the other pupil response parameters.

Combination of Pupil Response Parameters

A discriminant function, consisting of a combination of pupil response parameters of good diagnostic ability, was obtained using stepwise logistic regression. Twenty parameters with *P* values less than 0.05 were retained in the final model: RespAmpR-W, RespAmpL-W, ConMaxVelLatR-W, ConMaxVelLatL-W, RespAmpR-R, RespAmpL-R, RespLatR-R, ConMaxVelL-R, ConMaxVelR-R, DilMaxVelR-R, RespAmpR-B, RespLatR-B, DilMaxL-B, ConMaxVelLatL-B, RespAmpL-Y, DilMaxVelL-G, DilMaxVelLatL-G, ConMaxVelLatR-G, RespLatL-G, and absolute value of the RAPD score G. The combination of parameters had an age-corrected AUC of 0.85 (95% confidence interval [CI], 0.77–0.92). However, 5-fold cross-validation revealed a bias-corrected AUC estimate of only 0.74 (range, 0.69–0.85). In these subjects, high-definition OCT average RNFL thickness measurements had excellent ability to distinguish healthy and glaucomatous eyes with an AUC of 0.87 (95% CI, 0.82–0.91; Fig 2C).

Discussion

This study evaluated the ability of a new automated pupillograph to distinguish healthy subjects from those with glaucoma. Subjects with glaucoma had significant differences in pupil responses, including RespAmp, RespLat, ConMaxVel, and DilMaxVel, compared with healthy controls. A large number of stimuli and response parameters were studied, with the best individual measurements achieving an AUC of 0.75 for distinguishing between glaucomatous and healthy subjects.

One of the best-performing measures was the ConMaxVelLat on right eye stimulation using the blue stimulus, which achieved an AUC of 0.75 (Table 3). Although this was a relatively well-performing parameter, it would have an expected sensitivity of only 53% for a specificity of 80%. Therefore, if it were relied on for the detection of glaucoma, we would expect only 53% of actual subjects with glaucoma to be identified correctly and a relatively large number of false-positive results (20%). At a higher specificity of 95%, the sensitivity would be even lower, only 3%. Furthermore, in practice, the specificity is likely to be lower still because pupil response abnormalities are not specific to glaucoma. The results of this study therefore indicate that single pupillograph parameters are unlikely to be sufficiently accurate to be useful in the overall detection of glaucoma.

A problem with comparing AUCs from different tests is that the characteristics of the patients being tested can influence test performance. One test may be more sensitive at early stages of the disease, whereas another may be more sensitive at moderate or advanced stages. Because a subject with advanced glaucoma may have a greater abnormality on pupillography than a subject with early glaucoma, it is possible that the discriminatory accuracy of the pupillograph parameters is greater in advanced disease. To test this hypothesis, ROC regression was used to model the best pupil response parameter ROC curves as a function of disease severity, measured as the average MD between left and right eyes.[25,27] Table 5 indicates that there was a trend for better diagnostic ability with increasing average MD ($\beta_1 = -0.056$), but this was not statistically significant ($P = 0.058$). There was also no significant relationship between the average RNFL thickness and diagnostic ability of the best full-field stimuli. In contrast, OCT RNFL thickness was a

significantly better diagnostic marker in more severe disease, with an AUC of 0.99 for an MD value of −10 dB in this sample ($\beta_1 = -0.055$; $P < 0.001$).

Because the pupillograph measures binocular responses to monocular stimuli, disease asymmetry also may have an important bearing on the diagnostic ability of the device. We therefore performed ROC regression to model the ROC curve as a function of intereye difference in MD. Table 5 indicates that the best response to a white full-field stimulus had significantly better diagnostic ability in subjects with asymmetric disease ($\beta_1 = 0.098$; $P < 0.001$) compared with those with symmetric disease. Figure 2B shows the ROC curves for arbitrary values of intereye difference in SAP MD of 5, 10, and 15 dB, with AUCs of 0.71, 0.82, and 0.90, respectively. There was a similar relationship between disease asymmetry and diagnostic ability for each of the best performing full-field pupillograph parameters. Although the diagnostic ability of the pupillograph was better in asymmetric disease, in practice, large degrees of asymmetry in glaucoma are likely to be rare. Despite our sample including 11 subjects with glaucoma in 1 eye, overall we found those with glaucoma to have an average intereye MD difference of only 3.2 dB, and only 5 of 66 patients (8%) had an intereye MD difference of more than 10 dB. In eyes with symmetric disease, the AUC of the best performing full-field white stimulus parameter (Con-MaxVelLatL) was only 0.58.

An advantage of the automated pupillograph is that it is possible to obtain multiple pupil response characteristics to multiple stimuli. Also, because the device is computerized, it would be possible to generate an index automatically that combines the most useful individual measures to aid glaucoma detection. We therefore constructed a classification function using a combination of the best-performing pupillograph response parameters. The AUC for the combination of parameters was significantly better than the AUC for the best-performing single parameter (0.85 vs. 0.75; $P = 0.015$). The combination of parameters had a sensitivity of 70% for a specificity of 80% and a sensitivity of 50% for a specificity of 95%. These results suggest that the combination of pupil response parameters has reasonable ability to detect glaucoma; however, because there were a high number of pupil parameters relative to the sample size, there is a risk of model overfitting. To evaluate this, we performed a 5-fold cross-validation that resulted in a bias-corrected estimate of AUC for the combination of pupil response parameters of only 0.74. These results suggest that the pupillograph is less able to detect glaucoma than other devices such as OCT.[24] In previous studies, Stratus and Cirrus OCT average RNFL thicknesses both were found to have AUCs close to 0.90.[24,28] This is similar to the OCT average RNFL thickness AUC of 0.87 found in the present study. We found OCT RNFL thickness had a sensitivity of 78% for a specificity of 80% and a sensitivity of 60% for a specificity of 95%.

The results reported should be interpreted in the context of study limitations. First, there are many factors other than glaucoma that may affect pupil response. Systemic or topical medications, previous ocular surgery, or the presence of an irregular pupil shape may limit the accuracy of the device. We tested the device in a controlled population and excluded subjects taking medications that were likely to affect the pupil response and those with significant comorbidities such as nonglaucomatous optic neuropathies and retinal disease. In clinical practice, the performance of the pupillograph therefore may be different. If the device were to be used in a setting such as community screening, it is likely that the

specificity for detecting glaucoma would be lower because many conditions can affect pupil response. Further testing also would be needed to determine the cause of pupil response abnormality.

Although the relationship between disease severity and diagnostic ability of the best-performing full-field pupil response parameters did not reach statistical significance, it is possible that other pupil response parameters may be more affected by disease severity. We did not evaluate the effect of disease severity on the performance of every pupil response parameter because many had poor AUCs.

A further challenge was identifying the best combination of the large number of parameters generated by the pupillograph. In the future, when larger sample sizes are available for study, combinations of parameters with better diagnostic ability may be identified using techniques such as machine learning. It is also important to note that some parameters performed better on left eye stimulation and others on right eye stimulation. This suggests that the relatively good performance of some parameters was the result of characteristics of study patients. Furthermore, because most patients with glaucoma have bilateral disease, approaches that rely on the detection of asymmetry may not have high diagnostic yield. The development of novel stimuli and assessment of individual eye pupil responses may have improved ability to detect glaucoma. For example, multifocal pupillographic objective perimetry, which assesses individual pupil responses to a large number of smaller size stimuli, has been reported to have diagnostic accuracy similar to that of subjective perimeters.[30] Although we tested central and peripheral stimuli, light scatter may have meant that the stimulus was less focal than intended. In future, adjustment of the stimuli and testing conditions to target specific retinal ganglion cell subtypes, such as melanopsin containing intrinsically photosensitive retinal ganglion cells, may be useful.[31] For example, Kankipati et al[31] found that subjects with glaucoma have a decrease in the intrinsically photosensitive retinal ganglion cell–mediated postillumination pupil response, which refers to the period of sustained pupil constriction after light offset. The pupillograph used in the current study uses a standard battery of stimuli, and at present the user is unable to easily modify stimulus parameters.

The RAPDx pupillograph is a new device, and therefore there are few previous studies evaluating its ability to detect glaucoma. One previous study found a combination of 5 pupil response parameters resulting in an AUC of 0.92 (95% CI, 0.87–0.96), giving a specificity of 97% with a sensitivity of 80% (Invest Ophthalmol Vis Sci 53:ARVO E-Abstract 5621, 2012). Although this AUC was higher than the AUC for our combined parameter, the quoted AUC of 0.92 is a pooled result that does not take account of disease asymmetry or severity. Furthermore, all 5 parameters were measures of asymmetry (RAPD). We tested the same 5 combined parameters (RAPD scores for amplitude bright full-field white stimulus, amplitude dim central yellow stimulus, amplitude bright central green stimulus, latency bright peripheral red stimulus, and latency dim central green stimulus) and found a pooled AUC of only 0.67 (95% CI, 0.60–0.74). As we have demonstrated, the difference in performance likely reflects differences in the study populations. Chang et al (Invest Ophthalmol Vis Sci 2012;53:ARVO E-Abstract 5621). found that measures of asymmetry had better diagnostic ability than the measures of overall pupil response and pupil responses

to stimuli directed at different parts of the retina, which suggests that the sample had a large number of subjects with asymmetric damage. The difference in AUC obtained in the 2 populations also illustrates the value of validating a model using an external sample.

In conclusion, the pupil response parameters evaluated in this study had moderate ability to distinguish healthy subjects from those with glaucoma. The performance of the pupillograph depended on the characteristics of the population being tested and was significantly worse in those with more symmetric disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Robert N. Weinreb: Consultant - Carl Zeiss Meditec, Inc.; Financial support - Carl Zeiss Meditec, Inc., Heidelberg Engineering, Optovue, Kowa, Nidek, Konan, and Topcon.

Linda M. Zangwill: Financial support - Carl Zeiss Meditec, Inc., Heidelberg Engineering.

Felipe A. Medeiros: Financial support - Carl Zeiss Meditec, Inc., Heidel-berg Engineering.

Supported in part by the National Eye Institute, National Institutes of Health, Bethesda, Maryland (grant nos.: EY021818 [F.A.M.], EY11008 [L.M.Z.], EY14267 [L.M.Z.], EY019869 [L.M.Z.], and P30EY022589 [L.M.Z.]); Brazilian National Research Council-CNPq (grant no.: 200178/2012-1 [D.M-F.]; and an unrestricted grant from Research to Prevent Blindness, Inc., New York, New York. Grants for participants' glaucoma medications were received from Alcon, Allergan, Pfizer, Merck, and Santen.

References

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. Lancet. 2004; 363:1711–20. [PubMed: 15158634]
2. Katz J, Sommer A. Reliability indexes of automated perimetric tests. Arch Ophthalmol. 1988; 106:1252–4. [PubMed: 3046588]
3. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. Br J Ophthalmol. 2008; 92:569–73. [PubMed: 18211935]
4. Garway-Heath DF, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. Invest Ophthalmol Vis Sci. 2000; 41:1774–82. [PubMed: 10845598]
5. Medeiros FA, Zangwill LM, Bowd C, et al. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. Invest Ophthalmol Vis Sci. 2012; 5:6939–46. [PubMed: 22893677]
6. Medeiros FA, Lisboa R, Weinreb RN, et al. Retinal ganglion cell count estimates associated with early development of visual field defects in glaucoma. Ophthalmology. 2012; 120:736–44. [PubMed: 23246120]
7. Skorkovská K, Wilhelm H, Lüdtke H, Wilhelm B. Relative afferent pupillary defect in glaucoma. Klin Monbl Augenheilkd. 2011; 228:979–83. [in German]. [PubMed: 21847783]
8. Charalel RA, Lin HS, Singh K. Glaucoma screening using relative afferent pupillary defect. J Glaucoma. In press.
9. Chang DS, Xu L, Boland MV, Friedman DS. Accuracy of pupil assessment for the detection of glaucoma: a systematic review and meta-analysis. Ophthalmology. 2013; 120:2217–25. [PubMed: 23809274]
10. Wilhelm H, Wilhelm B. Clinical applications of pupillography. J Neuroophthalmol. 2003; 23:42–9. [PubMed: 12616089]

11. Chew SS, Cunnningham WJ, Gamble GD, Danesh-Meyer HV. Retinal nerve fiber layer loss in glaucoma patients with a relative afferent pupillary defect. Invest Ophthalmol Vis Sci. 2010; 51:5049–53. [PubMed: 20445112]
12. Thompson HS, Montague P, Cox TA, Corbett JJ. The relationship between visual acuity, pupillary defect, and visual field loss. Am J Ophthalmol. 1982; 93:681–8. [PubMed: 7091256]
13. Johnson LN, Hill RA, Bartholomew MJ. Correlation of afferent pupillary defect with visual field loss on automated perimetry. Ophthalmology. 1988; 95:1649–55. [PubMed: 3068603]
14. Kardon RH, Haupert CL, Thompson HS. The relationship between static perimetry and the relative afferent pupillary defect. Am J Ophthalmol. 1993; 115:351–6. [PubMed: 8442495]
15. Lankaranian D, Altangerel U, Spaeth GL, et al. The usefulness of a new method of testing for a relative afferent pupillary defect in patients with ocular hypertension and glaucoma. Trans Am Ophthalmol Soc. 2005; 103:200–7. [PubMed: 17057803]
16. Jonas JB, Zach FM, Naumann GO. Quantitative pupillometry of relative afferent defects in glaucoma. Arch Ophthalmol. 1990; 108:479–80. [letter]. [PubMed: 2322144]
17. Carle C, Maddess T, James A. Contraction anisocoria: segregation, summation, and saturation in the pupillary pathway. Invest Ophthalmol Vis Sci. 2011; 52:2365–71. [PubMed: 21212190]
18. Sample PA, Girkin CA, Zangwill LM, et al. ADAGES Study Group. The African Descent and Glaucoma Evaluation Study (ADAGES): design and baseline data. Arch Ophthalmol. 2009; 127:1136–45. [PubMed: 19752422]
19. Medeiros FA, Weinreb RN, Sample PA, et al. Validation of a predictive model to estimate the risk of conversion from ocular hypertension to glaucoma. Arch Ophthalmol. 2005; 123:1351–60. [PubMed: 16219726]
20. Medeiros FA, Vizzeri G, Zangwill LM, et al. Comparison of retinal nerve fiber layer and optic disc imaging for diagnosing glaucoma in patients suspected of having the disease. Ophthalmology. 2008; 115:1340–6. [PubMed: 18207246]
21. Racette L, Liebmann JM, Girkin CA, et al. ADAGES Group. African Descent and Glaucoma Evaluation Study (ADAGES): III. Ancestry differences in visual function in healthy eyes. Arch Ophthalmol. 2010; 128:551–9. [PubMed: 20457975]
22. Leite MT, Rao HL, Zangwill LM, et al. Comparison of the diagnostic accuracies of the Spectralis, Cirrus, and RTVue optical coherence tomography devices in glaucoma. Ophthalmology. 2011; 118:1334–9. [PubMed: 21377735]
23. Sarezky D, Krupin T, Cohen A, et al. Correlation between intereye difference in visual field mean deviation values and relative afferent pupillary response as measured by an automated pupillometer in subjects with glaucoma. J Glaucoma. 2013 Jan 18. [Epub ahead of print].
24. Medeiros FA, Zangwill LM, Bowd C, et al. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. Am J Ophthalmol. 2005; 139:44–55. [PubMed: 15652827]
25. Pepe MS. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. Biometrika. 1997; 84:595–608.
26. Alonzo TA, Pepe MS. Distribution-free ROC analysis using binary regression techniques. Biostatistics. 2002; 3:421–32. [PubMed: 12933607]
27. Medeiros FA, Sample PA, Zangwill LM, et al. A statistical approach to the evaluation of covariate effects on the receiver operating characteristic curves of diagnostic tests in glaucoma. Invest Ophthalmol Vis Sci. 2006; 47:2520–7. [PubMed: 16723465]
28. Leite MT, Zangwill LM, Weinreb RN, et al. Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis. Invest Ophthalmol Vis Sci. 2010; 51:4104–9. [PubMed: 20335619]
29. Chatterjee, S.; Hadi, AS. Regression Analysis by Example. 4. Hoboken, NJ: Wiley-Interscience; 2006. p. 281-316.
30. Carle CF, James AC, Kolic M, et al. High-resolution multifocal pupillographic objective perimetry in glaucoma. Invest Ophthalmol Vis Sci. 2011; 52:604–10. [PubMed: 20881285]
31. Kankipati L, Girkin CA, Gamlin PD. The post-illumination pupil response is reduced in glaucoma patients. Invest Ophthalmol Vis Sci. 2011; 52:2287–92. [PubMed: 21212172]

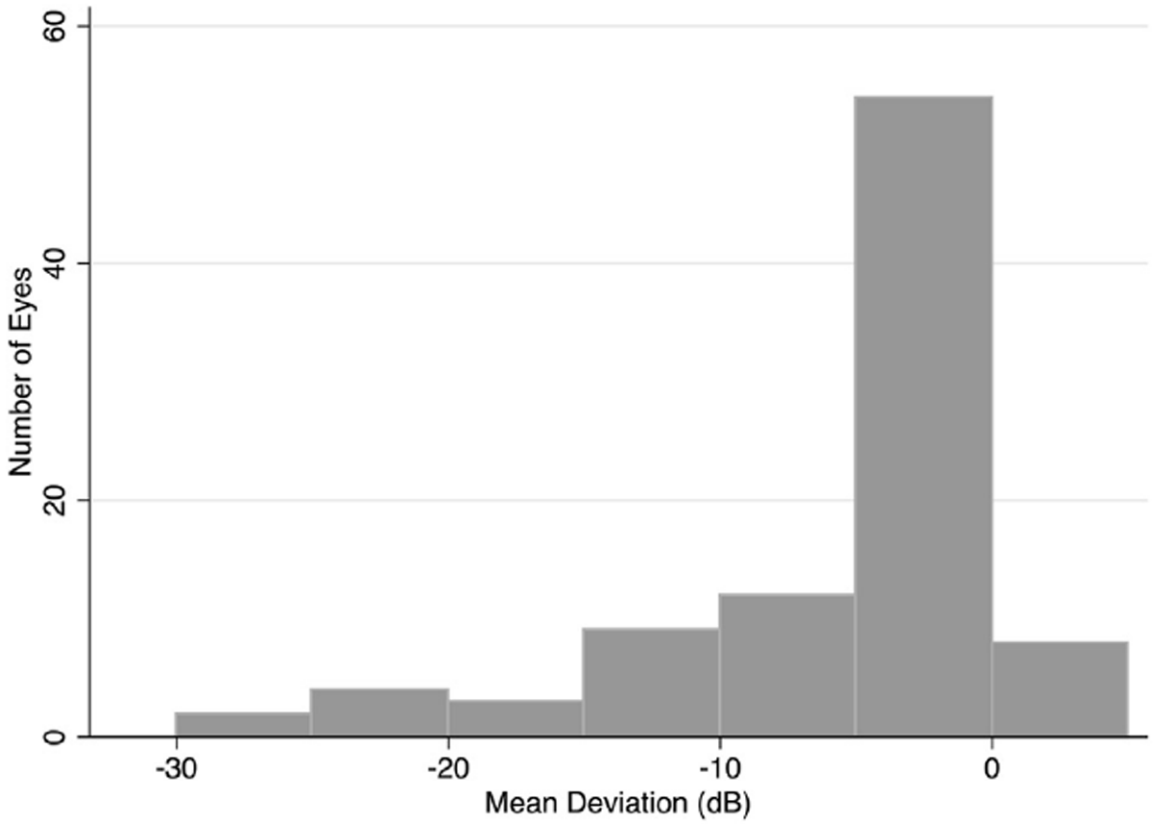

Figure 1.
Bar graph showing the distribution of disease severity, measured as mean deviation in decibels (dB) in glaucomatous eyes.

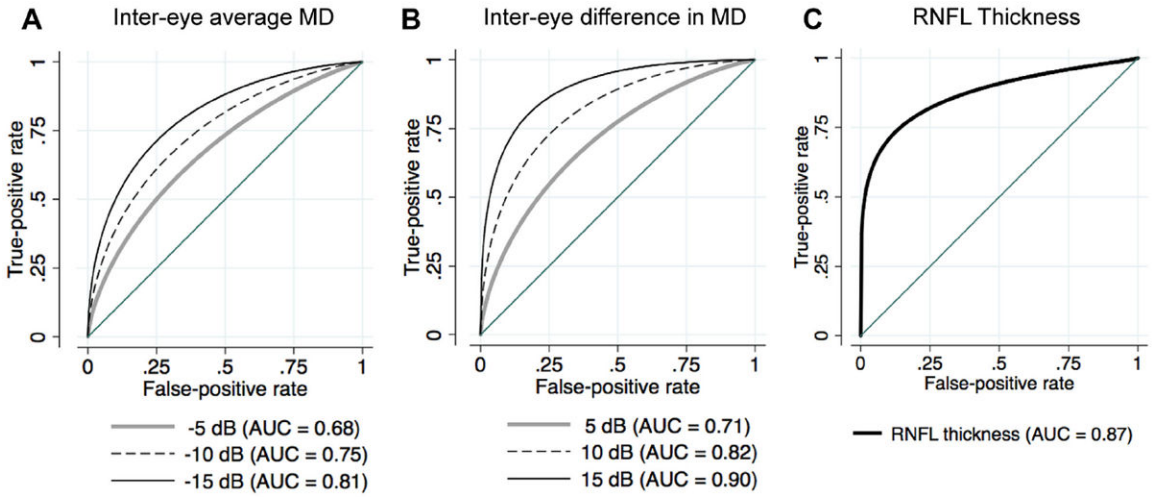

Figure 2.

Receiver operating characteristic (ROC) curves showing the ability of the best white full-field pupil response parameter (latency of the maximum velocity of constriction) to differentiate subjects with glaucoma in at least 1 eye from healthy subjects. The ROC curves are shown (**A**) for subjects with an intereye average mean deviation (MD) of −5, −10, and −15 decibels (dB) and (**B**) for those with an intereye difference in MD of 5, 10, and 15 dB for the sample mean age of 60.9 years. **C**, The ROC curve showing the ability of optical coherence tomography average retinal nerve fiber layer (RNFL) thickness to discriminate healthy and glaucomatous eyes. AUC = area under the ROC curve.

Table 1

Order and Timing of Trials for Each Stimulus Type Used during Pupillography

	Duration (ms)					
Stimulus	***Before Stimulus***	***Stimulus***	***After Stimulus***	***After Trial***	**No. of Trials**	**Total Time (Seconds)**
Full-field (white)	300	200	1500	100	18	37.8
Full-field (color*)	300	200	1500	100	10	25.2
Central (dim)	300	600	1100	100	10	21.0
Peripheral (dim)	300	600	1100	100	10	21.0
Central (bright)	300	600	1100	100	10	21.0
Peripheral (bright)	300	600	1100	100	10	21.0
INQ	300	600	1100	100	10	21.0
SNQ	300	600	1100	100	10	21.0

INQ = inferior nasal quadrant; SNQ = superior nasal quadrant.

*Red, green, yellow, and blue stimuli.

Ophthalmology. Author manuscript; available in PMC 2015 June 01.

Table 2

Demographic and Clinical Characteristics of Included Patients

	Subjects with 2 Healthy Eyes (n = 50; 100 eyes)	Subjects with Worse Eye with Glaucoma (n = 66; 132 eyes)	*P* Value
Age, yrs (range)*	51.3 (35.6–61.6)	69.1 (62.2–77.5)	<0.001
Female sex†	29 (58%)	32 (48%)	0.184
Race (no.)†	32 European ancestry	44 European ancestry	0.526
	11 African ancestry	17 African ancestry	
	7 other	5 other	
Mean deviation, db (range)*			
Worse eye	−0.2 (−1.0 to 0.4)	−3.3 (−8.8 to −1.3)	<0.001
Better eye	0.4 (−0.3 to 1.0)	−1.0 (−2.8 to −0.1)	<0.001
Absolute intereye difference	0.5 (0.2–1.0)	1.3 (0.6–4.8)	<0.001
Intereye average	0.1 (−0.6 to 0.7)	−2.5 (−4.9 to −0.8)	<0.001
RNFL thickness, μm (range)*			
Worse eye	94 (85–103)	72 (64–80)	<0.001
Better eye	95 (86–103)	81 (70–89)	<0.001
Absolute intereye difference	3 (1–4)	9 (5–13)	<0.001
Intereye average	95 (86–102)	77 (68–84)	<0.001

RNFL = retinal nerve fiber layer.

*Median (interquartile range) and Wilcoxon signed rank test.

†Fisher exact test.

Table 3

Median (Interquartile Range) of Pupil Response Parameters to the Full-Field White Stimulus with Pooled Area under the Receiver Operating Characteristic Curves for Sample Mean Age of 60.9 Years and True-Positive Rates at Fixed False-Positive Rates

Full-Field White Stimulus	Glaucoma Subjects (n = 66)	Healthy Subjects (n = 50)	*P* Value*	Area under the Receiver Operating Characteristic Curve (Standard Error)	True-Positive Rate	
					False-Positive Rate, 0.05	*False-Positive Rate, 0.2*
RespAmpR	0.28 (0.24–0.32)	0.30 (0.28–0.32)	0.001	0.60 (0.05)	0.14	0.47
RespAmpL	0.28 (0.24–0.31)	0.30 (0.27–0.32)	<0.001	0.59 (0.05)	0.10	0.38
RespLatR	183 (168–205)	168 (159–178)	<0.001	0.59 (0.06)	0.10	0.41
RespLatL	183 (169–199)	168 (161–180)	<0.001	0.57 (0.05)	0.14	0.38
ConMaxVelR	4.2 (3.6–5.0)	5.1 (4.1–6.1)	<0.001	0.54 (0.07)	0.10	0.24
ConMaxVelL	4.1 (3.6–5.0)	5.1 (4.2–5.9)	<0.001	0.51 (0.07)	0.12	0.20
ConMaxVelLatR	397 (374–421)	375 (366–393)	<0.001	0.59 (0.06)	0.11	0.39
ConMaxVelLatL	398 (385–417)	381 (369–390)	<0.001	0.65 (0.05)	0.06	0.52
DilMaxVelR	2.4 (2.0–2.7)	2.7 (2.3–3.1)	<0.001	0.53 (0.07)	0.08	0.23
DilMaxVelL	2.5 (2.1–2.7)	2.8 (2.4–3.1)	<0.001	0.54 (0.08)	0.08	0.14
DilMaxVelLatR	1121 (1075–1175)	1110 (1056–1141)	0.018	0.52 (0.05)	0.08	0.23
DilMaxVelLatL	1120 (1075–1173)	1101 (1066–1130)	0.006	0.54 (0.04)	0.14	0.29
absRAPDscore	0.17 (0.06–0.30)	0.09 (0.04–0.15)	<0.001	0.60 (0.05)	0.20	0.42

The best performing parameters are shown in bold.

absRAPDscore = absolute relative afferent pupillary defect score; ConMaxVel = maximum velocity of constriction; ConMaxVelLat = latency of the maximum velocity of constriction; DilMaxVel = maximum velocity of dilation; DilMaxVelLat = latency of the maximum velocity of dilation; L = left; R = right; RespAmp = response amplitude; RespLat = response latency.

*Wilcoxon signed-rank test.

Table 5

Pupil Response Parameters to Dim and Bright Peripheral and Central Stimuli Using the White, Blue, Yellow, Green, or Red Light Stimuli

	Glaucoma Subjects (n = 66)	Healthy Subjects (n = 50)	*P* Value*	Area under the Receiver Operating Characteristic Curve (Standard Error)
Dim peripheral				
DilMaxVelLatL-G	1263 (1200–1325)	1200 (1125–1250)	<0.001	0.75 (0.07)
ConMaxVelLatR-W	413 (375–425)	375 (363–413)	<0.001	0.73 (0.08)
ConMaxVelLatL-B	413 (388–450)	388 (363–400)	<0.001	0.74 (0.07)
RespAmpL-R	0.19 (0.15–0.22)	0.20 (0.17–0.23)	0.100	0.75 (0.09)
Bright peripheral				
ConMaxVelLatR-Y	400 (363–425)	375 (363–400)	<0.001	0.73 (0.06)
ConMaxVelLatL-W	388 (375–425)	375 (363–400)	<0.001	0.74 (0.08)
RespLatR-B	193 (172–210)	174 (164–189)	<0.001	0.75 (0.07)
DilMaxVelLatR-B	1213 (1150–1275)	1163 (1113–1225)	0.001	0.73 (0.07)
Dim central				
DilMaxVelLatL-W	1281 (1213–1363)	1244 (1188–1325)	0.043	0.70 (0.08)
RespAmpL-W	0.26 (0.21–0.28)	0.29 (0.25–0.32)	<0.001	0.71 (0.09)
RespAmpL-B	0.22 (0.19–0.25)	0.25 (0.21–0.29)	<0.001	0.72 (0.09)
RespAmpR-G	0.25 (0.21–0.29)	0.27 (0.24–0.31)	0.003	0.72 (0.10)
Bright central				
ConMaxVelLatR-W	444 (413–463)	413 (388–438)	<0.001	0.71 (0.06)
DilMaxVelLatR-W	1241 (1138–1300)	1188 (1113–1250)	0.002	0.75 (0.05)
DilMaxVelLatR-B	1188 (1025–1275)	1063 (988–1200)	0.002	0.72 (0.06)
DilMaxVelLatL-B	1163 (1050–1275)	1000 (938–1125)	<0.001	0.75 (0.05)

-B = blue; ConMaxVel = maximum velocity of constriction; ConMaxVelLat = latency of the maximum velocity of constriction; DilMaxVelLat = latency of the maximum velocity of dilation; -G = green; L = left; -R = red; R = right; RespAmp = response amplitude; RespLat = response latency; -W = white. Only the 3 parameters for each stimulus area with the greatest area under the receiver operating characteristic curves are shown.

* Wilcoxon signed-rank test.

Table 6

Results of the Receiver Operating Characteristic Regression Model to Evaluate the Effect of Disease Severity and Disease Asymmetry on the Diagnostic Ability of the Best White Full-Field Pupil Response Parameter, with Age Included as a Covariate

Parameter	Coefficient	Estimate	95% Confidence Interval	*P* Value
Severity				
Intercept	α_1	0.346	(−0.055 to 0.746)	0.091
$\Phi^{-1}(q)$	α_2	0.928	(0.724–1.131)	<0.001
Severity (average MD)	β_1	−0.056	(−0.102 to 0.010)	0.058
Asymmetry				
Intercept	α_1	0.265	(−0.723 to 0.603)	0.124
$\Phi^{-1}(q)$	α_2	0.945	(0.771–1.118)	<0.001
Intereye MD difference	β_1	0.098	(0.050–0.145)	<0.001

MD = mean deviation (decibels); Φ = normal cumulative distribution function.