Comparison of Optic Nerve Imaging Methods to Distinguish Normal Eyes from Those with Glaucoma

Michael J. Greaney, Douglas C. Hoffman, David F. Garway-Heath, Mamdouh Nakla, Anne L. Coleman, and Joseph Caprioli

PURPOSE. To compare the ability of qualitative assessment of optic nerve head stereophotographs (ONHPs), confocal scanning laser ophthalmoscopy (CSLO), scanning laser polarimetry (SLP), and optical coherence tomography (OCT) to distinguish normal eyes from those with early to moderate glaucomatous visual field defects.

METHODS. Eighty-nine eyes (63 normal, 63 age-matched with glaucoma) of 89 subjects more than 40 years of age were studied. Receiver operating characteristic (ROC) curves were generated from discriminant analysis of CSLO, SLP, and OCT measurements and from ONHP scores. Sensitivity at 80% and specificity at 90% were calculated. Differences between individual methods and combinations of methods were assessed for statistical significance. Agreement on categorization between methods (κ) was assessed.

RESULTS. The average visual field mean deviation (MD ± SD) in patients with glaucoma was -3.9 ± 2.2 dB, and the average pattern standard deviation (PSD) was 4.7 ± 3.4 dB. In normal subjects the average MD was 0.1 ± 0.9 dB and the average PSD was 1.5 ± 0.3 dB. Optimal sensitivities, specificities, and areas under ROC curves were, respectively: ONHP (0.94, 0.87, 0.93), CSLO (0.84, 0.90, 0.92), SLP (0.89, 0.87, 0.94), and OCT (0.82, 0.84, 0.88). Best agreement on categorization (κ) was between ONHPs and CSLO (0.70). The ROC area for the combination of methods was 0.99, higher than for any method alone. The ROC area for the combination of methods was significantly better than the CSLO rim area (P = 0.012) and the OCT retinal nerve fiber layer (RNFL) thickness (P = 0.002).

CONCLUSIONS. The quantitative methods CSLO, SLP, and OCT were no better than qualitative assessment of disc ONHPs by experienced observers at distinguishing normal eyes from those with early to moderate glaucoma. A combination of the imaging methods significantly improves this capability. (Invest Ophthalmol Vis Sci. 2002;43:140–145)

A lteration in the structural appearance of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL) usually precedes the development of reproducible glaucomatous achronic1–5 and blue-on-yellow6 visual field defects. Identification of these changes is important in the diagnosis and treatment of glaucoma at an early to moderate stage and in monitoring its clinical course. Variability in the size and appearance of the ONH of normal eyes7,8 accounts for much of the difficulty in detecting the presence of early to moderate glaucomatous optic nerve damage. Assessment of ONH and RNFL topographies based on clinical evaluation is subject to disagreement between even experienced observers9–11 and may not be sufficiently sensitive to distinguish normal from glaucomatous optic nerves5 or may not detect subtle changes in appearance over time.12 Structural measurement of the ONH and RNFL offers the prospect of improved early detection and monitoring of glaucoma.

In recent years, various methods have become available to produce digital images of the ONH and RNFL from which to compute structural measurements. They appear to offer objectivity, rapidity, and reproducibility. Their ability to detect glaucoma has been described.13–17 However, the utility of each of these techniques in isolation and in combination for the detection of early to moderate glaucomatous optic neuropathy requires further study. In this study we evaluated the same sample of patients with qualitative evaluation of ONH stereoscopic color photographs (ONHPs), confocal scanning laser ophthalmoscopy (CSLO), scanning laser polarimetry (SLP), and optical coherence tomography (OCT) to compare their relative abilities to distinguish normal eyes from those with early to moderate glaucomatous visual field loss and to evaluate agreement between the techniques in categorizing eyes.

METHODS

Eighty-nine eyes (39 normal, 50 age-matched with glaucoma) of 89 subjects more than 40 years of age were examined. The demographic details are given in Table 1. All eyes had 5 D or less spherical equivalent of ametropia and best corrected visual acuity of 20/40 or better. ONH appearance was not considered during subject selection, to avoid sample bias. Subjects who had undergone cataract extraction were included only if the procedure had been performed more than 1 year before examination. All subjects underwent visual field testing followed by slit lamp biomicroscopic ocular examination before and after pupillary dilation with 1% tropicamide eye drops (Mydriacyl, Alcon Laboratories, Inc., Fort Worth, TX) and 2.5% phenylephrine eye drops (Bausch & Lomb Pharmaceuticals, Inc., Tampa, FL). All tests on individual subjects were performed within a 6-month period. The methods described adhered to the tenets of the Declaration of Helsinki for the use of human subjects in biomedical research. The Human Subject Protection Committee of the University of California at Los Angeles approved the study. Informed consent was obtained from each subject before enrollment.

Subjects

Patients with glaucoma were recruited from patients attending the Jules Stein Eye Institute’s Glaucoma Division. Criteria for the patient’s inclusion as having early to moderate glaucoma were an open anterior chamber, absence of other eye disease, and an early to moderate reproducible glaucomatous visual field defect in the absence of any other abnormalities to explain the defect (defined later) and with a
TABLE 1. Gender, Race, Age, Mean Deviation, and Diagnosis of Normal Subjects and Patients With Glaucoma

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Glaucoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>Female</td>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td>Black</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Age (y ± SD)</td>
<td>58.4 ± 10.2</td>
<td>61.8 ± 10.5</td>
</tr>
<tr>
<td>MD (dB ± SD)*</td>
<td>0.0 ± 0.9</td>
<td>-3.9 ± 2.3</td>
</tr>
<tr>
<td>Range (dB)</td>
<td>-1.49-2.08</td>
<td>-8.79-0.00</td>
</tr>
<tr>
<td>PSD (dB ± SD)†</td>
<td>1.5 ± 0.3</td>
<td>4.7 ± 3.4</td>
</tr>
</tbody>
</table>

Glaucoma diagnoses

POAG                  | 36     |
POAG, normal tension type | 15     |
Pigment dispersion      | 5      |
Angle recession         | 3      |
Nevus of Ota            | 1      |

* Significantly different (P < 0.001).
† Significantly different (P < 0.001).

mean deviation (MD) of more than −10 dB. Patients’ diagnostic subtypes and the number of patients in each group are shown in Table 1. Subjects with normal eyes were recruited from spouses of patients with glaucoma or volunteers. Criteria for inclusion as normal were no history of eye disease, no history of glaucoma in a first-degree relative, intraocular pressure less than 21 mm Hg when measured by Goldmann applanation tonometry, normal findings in an eye examination, and a reliable glaucoma hemifield test with normal findings.

Visual Field Testing

All subjects underwent achromatic automated static perimetry with the Swedish interactive threshold algorithm (SITA) standard of the Humphrey Field Analyzer 750 (Allergan Humphrey, San Leandro, CA). A reliable test was defined as having fewer than 20% false-positive or false-negative scores and fewer than 33% fixation losses.

A glaucomatous visual field defect was defined in a SITA standard 24-2 program (Allergan Humphrey) as two or more contiguous points with a pattern deviation P < 0.01 sensitivity loss or more, or three or more contiguous points with P < 0.05 sensitivity loss or more, in the superior or inferior arcuate areas (compared with that in perimeter-defined age-matched control subjects), or a 10-db difference across the nasal horizontal midline at two or more adjacent locations and an abnormal result in a glaucoma hemifield test.

One of the authors (MJG) identified the patients with early to moderate glaucoma and normal subjects with these criteria while masked to the ONH’s appearance and data.

Imaging

A fundus camera (Fundus Flash III; Carl Zeiss, Oberkochen, Germany) mounted with a ×2 magnification adaptor was used to acquire a pair of sequential ONHs in each subject. Subjects’ pupils were dilated, and photographs were taken at the leftmost and rightmost position of the pupil, to maximize the stereoscopic base. Each stereoscopic pair of transparencies was diffusely retroilluminated by a horizontally mounted, color-corrected, discharge-tube light box (Logan Electric Spec Mfg. Co., Chicago, IL) and viewed through a stereo viewer (Carl Zeiss). Three experienced observers (ALC, JC, DFG), who were masked to the patients’ identities and diagnoses by concealment of written information on the photographs, independently made qualitative assessments of the ONH and RNFL for glaucomatous damage. One observer had not seen the photographs for more than 6 months and two of the observers had never seen them. Each OHNP was graded as 1 (definitely normal), 2 (probably normal), 3 (undecided), 4 (probably glaucoma), or 5 (definitely glaucoma). The cumulative score for each ONHP was the sum of the scores assigned by the three observers.

CSLO (Heidelberg Retina Tomograph [HRT]; Heidelberg Engineering GmbH, Heidelberg, Germany) was performed in each patient with a 10° × 10° image field. Three 10° topographic images taken at the same sitting were used to generate a mean topographic image. A good-quality image was defined as one in which the mean SD of height measurements was less than 50 μm. The optic disc was defined by a contour line drawn along the inner margin of Eisingh’s ring by the concordance of two operators (DFG, MJG) referring to the ONHs. HRT proprietary software (ver. 2.01b; Heidelberg Engineering GmbH) was applied, with the standard reference plane, to calculate global cup area, rim area, cup-to-disc area ratio, rim volume, cup volume, cup shape measure, height variation contour, RNFL height, and cross-sectional area of the NFL. In addition, a proprietary HRT software program (HRTclc10 ver. 2.01; Heidelberg Engineering GmbH) was used to calculate rim area in each of twelve 30° sectors.

SLP (GDx Nerve Fiber Analyzer; Laser Diagnostics Technologies, San Diego, CA) was used to estimate the peripapillary RNFL thickness. The mean of three good-quality aligned images was evaluated. The proprietary software (ver. 1.0.16) calculates summary measurements: symmetry, superior ratio, inferior ratio, superior-to-nasal ratio, maximum modulation, ellipse modulation, the number (a score generated by a neural network in this system), average thickness, ellipse average, superior average, inferior average, and superior integral. The RNFL thickness was measured along an annular ellipse 10 pixels wide, concentric with, and 1.75 times the diameter of the ellipse drawn over the scleral ring. In addition, the average RNFL thickness on this ellipse was recorded in each of twelve 30° sectors. The SLP parameter, the number, was recorded.

OCT (Humphrey Instruments, Dublin, CA) was used to measure the thickness of the peripapillary RNFL. Measurements were made at 100 points along a circle concentric with the ONH at a radius of 1.69 to 1.75 mm. These were used to calculate the average RNFL thickness for each of twelve 30° sectors.

Statistical Analyses

Statistical analysis was performed on computer (SPSS for Windows, ver. 9.0; SPSS, Chicago, IL). Differences between ages of subjects in the two groups were compared by Student’s t-test. Discriminant analysis was used to identify and combine the most useful parameters of each imaging method. Discriminant analysis is a technique that helps identify what characteristics best distinguish the differences between predefined groups. Discriminant analysis combines the original variables to generate a new variable in such a way that the measurable differences between the groups are maximized. In all our discriminant analyses, a diagnostic score of 0 for normal or 1 for glaucoma was entered as the dependent variable. To evaluate diagnostic categorization, measurement data were entered together as the independent variables included in the analysis (‘group discriminant analysis’). In each analysis, each subject was classified by the functions derived from all the other subjects using the ‘leave-one-out’ method. The relative importance of each of a set of independent variables was assessed by stepwise discriminant analysis.

For ONHs, chance-corrected agreement between pairs of ONH grader (κ) was calculated by a weighted κ algorithm.²° Software developed by one of the authors (DH) was applied to the cumulative score derived from assessment of the ONHP pairs to generate a receiver operating characteristic (ROC) curve.²² For CSLO, the following variables were entered into group and stepwise discriminant analyses: cup area, rim area, cup-to-disc area ratio, rim volume, cup volume, cup shape measure, height variation contour, RNFL height and cross-sectional area, and cup shape measure. A similar but separate analysis was performed on rim area measured for each 30° sector. For SLP, the following summary parameters were entered into group and stepwise...
discriminant analyses: symmetry, superior ratio, inferior ratio, superior-to-nasal ratio, maximum modulation, ellipse modulation, average thickness, ellipse average, superior average, inferior average, and superior integral. Group and stepwise discriminant analyses were performed separately on the mean RNFL thickness measured by the SLP for each 30° sector. For OCT, group and stepwise discriminant analyses were applied to the mean RNFL thickness measured by OCT for each 30°.

Comparison of Methods

For each method, the analysis producing the largest area under the ROC curve was chosen as the best. The differences between the ROC curves derived from the best analysis by each method were tested for significance. The agreement (k) of categorization between pairs of methods was tested for the best analyses and for the percentile analyses. An ROC curve was generated for the combined best analyses for each method and was compared with the single best analysis.

Results

Demographic Data

The average age (±SD) of patients with glaucoma was 61.0 ± 10.9 years (Table 1). The average age (±SD) of normal subjects was 58.5 ± 9.7 years. There was no statistically significant difference between the groups (P = 0.18).

Perimetry

The average MD in patients with glaucoma (±SD) was −3.88 ± 2.24 dB (range, −0.72 to −10.43 dB). The average MD for normal subjects was 0.11 ± 0.88 dB (range, −2.15 to 2.37 dB). The mean deviations of the two groups were significantly different (P < 0.001). The average PSD in patients with glaucoma was 4.7 ± 3.4 dB and in normal subjects was 1.5 ± 0.3 dB. The PSDs of the two groups were significantly different (P < 0.001).

Optic Nerve Head Stereophotographs

Sensitivities and specificities for detecting glaucoma were, respectively, 76% and 85% for the first ONHP grader, 86% and 90% for the second, and 84% and 92% for the third. Agreement between pairs of ONHP graders (weighted k) was 0.66, 0.68, and 0.76. The combined assessment of the three graders is represented by an ROC curve shown in Figure 1, along with the ROC curves derived from the best analysis obtained with each of the other methods. For each ROC curve, the corresponding areas, sensitivities, and specificities, along with the specificities at the 80% and 90% sensitivity levels, are given in Table 2.

Confocal Scanning Laser Tomography

The average of the mean SDs of height for the mean topographic images was 19.2 μm. The ROC area, optimal sensitivity and specificity, and sensitivities at the 80% and 90% specificity levels for the analyses performed are shown in Tables 2 and 3. The parameters identified by stepwise analysis of global parameters (i.e., those having the greatest discriminant value) were rim area and cup shape measure (normalized third moment of the frequency distribution of the depth values for the disc structures). The ROC curves of the latter two were not significantly different (P = 0.77). The best analysis with the CSLO was obtained from group discriminant analysis of rim area in 30° sectors. Stepwise discriminant analysis of rim area in 30° sectors identified the following five sectors listed in order of the most discriminating: nasal of inferior (240–270°), temporal of superior (60–90°), inferior of temporal (330–360°), superior of temporal (0–30°), and superotemporal (30–60°).

Comparison among Best Analyses for Each Method

ROC curves generated from the discriminant analyses and the assessment of disc ONHPs are shown in Figure 1. For CSLO, SLP, and OCT, the best analysis was obtained by group discriminant analysis of 30° sectoral data. Stepwise discriminant analysis identified the following four sectors in order of the most discriminating: temporal of inferior (265–295°), superior of temporal (0–30°), superior of nasal (150–180°), and nasal of inferior (240–270°).

Scanning Laser Polarimetry

The ROC areas, optimal sensitivities and specificities, and the sensitivities at the 80% and 90% specificity levels from the analyses of SLP summary parameters are shown in Tables 2 and 3. The parameters identified by stepwise discriminant analysis as having the greatest discriminant value were inferior ratio, maximum modulation, ellipse modulation, and superior ratio. The best analysis with the SLP was obtained from group discriminant analysis of 30° sectoral data. Stepwise discriminant analysis of RNFL thickness in 30° sectors identified the following four sectors listed in order of the most discriminating: temporal of inferior (270–300°), superior of temporal (0–30°), superior of nasal (150–180°), and nasal of inferior (240–270°).

Optical Coherence Tomography

The ROC areas, optimal sensitivities and specificities, and the sensitivities at the 80% and 90% specificity levels from the analyses of OCT summary parameters are shown in Tables 2 and 3. The best analysis was derived from group discriminant analysis of RNFL thickness in 30° sectors. Stepwise discriminant analysis identified the following four sectors in order of the most discriminating: temporal of superior (45–75°), inferior (265–295°), temporal (345–15°), and superior of temporal (15–45°).

Comparison among Best Analyses for Each Method

ROC curves generated from the discriminant analyses and the assessment of disc ONHPs are shown in Figure 1. For CSLO, SLP, and OCT, the best analysis was obtained by group discriminant analysis of 30° sectoral data. Differences between other combinations of pairs of methods were not significant (P > 0.05). Agreement on categorization (k) between pairs of methods was assessed (Table 4). Each method showed a good ROC curve, but the agreement between methods (k) ranged from substantial (0.73 for ONHPs and CSLO), to moderate (0.58 for ONHPs and OCT, 0.58 for CSLO and OCT), to poor (0.37 for CSLO and SLP, 0.36 for SLP and OCT, 0.34 for ONHPs and SLP).
Combination of Methods

The ROC curve derived from group discriminant analysis of the combination of the best analyses for each method (the 12 sectors from CSLO, OCT, and SLP plus clinical disc score, i.e., 37 parameters in all) is shown in Figure 1 and the corresponding data are given in Table 3. The inclusion of 37 parameters in a discriminant analysis may, by chance, lead to a function capable of discriminating between two groups. To assess this possible artifact, 37 random variables were entered into a group discriminant analysis, with 63 cases in each of two diagnostic categories, 0 and 1. This resulted in an ROC curve with an area of 0.62, a sensitivity of 54%, and a specificity of 67%.

**Discussion**

In this study, ONHP, CSLO, SLP, and OCT were used to measure different characteristics of glaucomatous optic nerve damage with the goal of diagnosing early to moderate glaucoma in the same population sample. Neither CSLO, SLP, nor OCT was better than ONHP assessed by experienced observers to distinguish normal eyes from those with early to moderate glaucoma. For CSLO, SLP, and OCT, the best analysis was achieved by group discriminant analysis of 30° sectoral data. Agreement on categorization (κ) between methods was assessed using the most discriminating analysis for each.

The parameters found with CSLO to discriminate best between normal eyes and those with early to moderate glaucoma were rim area and cup shape measure. For SLP the best discriminants were inferior ratio, inferior average, and ellipse modulation. For OCT they were average RNFL thickness along the measurement circle in the following 30° sectors: temporal of superior (45–75°), inferior (265–295°), temporal (345–15°), superior of temporal (15–45°).

A combination of imaging methods significantly improves the ability to distinguish normal eyes from those classified by visual fields as having early to moderate glaucoma. The ROC curve derived from the 37 parameters from the best analyses with each method (Fig. 1, Table 2) was significantly better than that of the individual methods CSLO (P = 0.012) and OCT (P = 0.002). This should not be surprising, because the methods evaluated in this study probably assess different aspects of the ONH and RNFL and are therefore likely to complement each other, as long as the measurements are reliable and reproducible.

Our finding that the most discriminating CSLO and SLP analyses were obtained from sectoral rather than global measurements agrees with a previous similar finding for the field analyzer (HRT; Allergan Humphrey). A possible explanation is that if glaucoma is defined by the presence of adjacent visual field test points at which retinal sensitivity values are below a certain threshold, it may favor focal defects,
and these will be less apparent in averaged or global measurements. The most discriminating CSLO parameters identified by this study were rim area and cup shape measure. This is in agreement with previous studies.\textsuperscript{15,22}

We compared the efficacy of qualitative assessment of ONHPs with that of three methods that allow objective measurement of disc topographic features (CSLO) or RNFL thickness (OCT, SLP) in the same population sample. Our analyses were conducted to identify the best discriminating parameters available with each technique. The results reported herein apply to the sample studied and should be extrapolated to other population samples with caution.

The appearance of the ONH or RNFL was not used as a restriction criterion for the entry of subjects into either the normal or glaucoma groups. This avoids sample bias that might prejudice the outcome and may allow our findings to be extrapolated to an unselected population.\textsuperscript{25,26} Differences between the abilities of quantitative imaging methods to distinguish glaucomatous from normal eyes should be more apparent in cases of glaucoma that are neither very early, when the ONH structural measurements more closely resemble normal,\textsuperscript{27} nor too advanced, when all methods are more likely to classify subjects correctly. This study demonstrates that currently available imaging devices perform no better than ONHPs to distinguish normal eyes from those with early to moderate glaucoma. However, a consensus of expert opinion about disc photographs does not reflect routine clinical practice. Clinical assessment of the optic disc is more commonly performed, not by a panel of experts, but by individual, sometimes less experienced, observers, and these clinicians may perform less well than a panel of experts. In such circumstances, the relative performance of quantitative assessment of disc photographs and quantitative methods may differ from those presented in this report.

Each of the imaging methods compared in this study has limitations. CSLO would benefit from a stable reference plane, independent of the RNFL, from which to make measurements\textsuperscript{18,28,29} and from automated detection of the optic disc margin. The integrated corneal compensator of the nerve fiber analyzer (GDx; Laser Diagnostics) does not neutralize corneal polarization in all cases.\textsuperscript{18,50} A more accurate definition of the boundaries of the RNFL may improve the accuracy of OCT.\textsuperscript{51}

No single quantitative imaging technique, CSLO, SLP, or OCT, was better than qualitative assessment of ONHPs at distinguishing normal eyes from those with early to moderate glaucoma. A combination of the best parameters from the four imaging methods, however, improved the ability to distinguish normal eyes from those with early to moderate glaucoma and was significantly better than CSLO or OCT alone. Currently, the extensive testing required to create a combination of the best features of each technique is cumbersome and too time consuming to be clinically practical. Future work will determine an optimal testing strategy that produces high sensitivity at high specificity and is clinically practical.

### Table 4. Agreement on Categorization (x) between the Best Analysis by Each Method

<table>
<thead>
<tr>
<th></th>
<th>CSLO</th>
<th>SLP</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONHPs</td>
<td>0.73</td>
<td>0.34</td>
<td>0.58</td>
</tr>
<tr>
<td>CSLO</td>
<td>0.37</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>SLP</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Acknowledgments

The authors thank Gang Li, Associate Professor, Division of Public Health Medicine, University of California Los Angeles, for statistical advice.

### References


