A Statistical Model for the Evaluation of Sensory Tests in Glaucoma, Depending on Optic Disc Damage

Andrea Stroux,1,2 Matthias Korth,3 Anselm Jünemann,3 Jost B. Jonas,4 Folkert Horn,3 Andreas Ziegler,5 and Peter Martus1

PURPOSE. To analyze the sensitivity of various sensory tests adjusted for glaucomatous optic disc damage.

METHODS. In a cross-sectional study, the results of testing of 196 control subjects (age range, 18–69 years) and 308 patients with chronic open-angle glaucoma (age range, 18–70 years) were included. The perimetric mean defect (MD), a temporal contrast sensitivity test (TCS), a spatiotemporal contrast sensitivity test (STCS), the peak time of a blue-on-yellow visual evoked potential (BYVEP), and the amplitude of a pattern-reversal electroretinogram (PERG) were evaluated by a specific logistic regression model. This model included glaucomatous damage, quantified by neuroretinal rim area corrected for disc size, as a covariate of sensitivity.

RESULTS. Sensitivity of diagnostic tests increased for all procedures with increasing loss of neuroretinal rim area. With progressing optic disc damage, MD and STCS showed higher sensitivity than did TCS. BYVEP showed a higher sensitivity than PERG in all disease stages. In general, the psychophysical tests were more sensitive than the electrophysiological ones.

CONCLUSIONS. The specific model used in this study was an appropriate tool to analyze the sensitivity of several sensory glaucoma tests in relation to disease stage. Moreover, tests that were more sensitive in early disease stages (TCS) and others that were more sensitive in more advanced stages (MD, STCS) were identified. (Invest Ophthamol Vis Sci. 2003;44: 2879–2884) DOI:10.1167/iovs.02-0419

Chronic glaucomas are a heterogeneous group of eye diseases characterized by increased intraocular pressure (IOP) in many patients and by progressive damage to the optic disc and to the visual field. In these diseases, the IOP is an important risk factor, whereas optic disc changes and perimetric defects are definite signs of the disease and typically progress slowly over years or decades. An elevated IOP without any of the just mentioned signs of the disease is a state called ocular hypertension (OHT). Although it has been shown that IOP levels are important in the course of the disease and that IOP reduction slows down disease progression, the progression of OHT to the glaucomatous disease seems to be a relatively rare event, estimated to occur at a rate of approximately 1% to 2% per year.1–3 If a conversion from OHT to glaucoma takes place, the optic disc or retinal nerve fiber layer changes often occur before the visual field defects,4,5 and thus may be considered useful signs for an early glaucoma diagnosis. Visual field loss, determined by conventional white-on-white perimetry, seems to be a late defect in the disease process and may occur only when a considerable number of optic nerve fibers have been lost.6 Thus, perimetry may be considered a useful tool for the follow-up of patients with advanced stages of glaucoma. Apart from detecting rather advanced glaucomatous damage, perimetry is a time-consuming and fatiguing procedure burdened with high variability of the patient's performance and learning effects.8–10 Therefore, during recent years, other sensory procedures, psychophysical as well as electrophysiological, have been devised that can indicate glaucomatous damage earlier and in a more reliable and quicker way than perimetry. The diagnostic value of such methods has been described previously.11–13

In general, diagnostic tests become more sensitive with progressing damage. In addition, between the individual test types, there seems to be a great difference in which test performs best in which stage of the disease. The question of how the sensitivity of such tests changes during the disease process is important, because it may advise the ophthalmologist which of the experimental tests to use in different stages of glaucoma.

It was the purpose of the present investigation to answer this question by analyzing data from the Erlangen Glaucoma Registry, which was started in 1991 and has accumulated a considerable number of patients and control subjects who have been examined at regular intervals with various experimental sensory tests. Ideally, such an analysis should be undertaken by individual follow-ups. However, that is difficult because of the low rate of disease progression. In addition to the chronicity inherent in the disease, therapeutic interventions probably further delayed an advancement of the disease. For these reasons, the progression of the glaucomatous process was studied in cross-sectional analyses of patients who entered the glaucoma registry at different stages of glaucoma. In the present investigation, perimetry was treated equivalent to other experimental sensory tests to compare them with each other. In several studies, a highly significant correlation (P < 0.001, r = −0.81) was shown between the neuroretinal rim (NRR) area and glaucoma stage according to optic disc topography.14,15 This leaves the optic nerve head damage, measured as size of the neuroretinal rim (NRR) area, as the appropriate reference for scaling the glaucomatous damage.

METHODS

The methods applied in the study adhered to the tenets of the Declaration of Helsinki for the use of human subjects in biomedical research. Informed consent was obtained from each subject before enrollment.
Table 1. Subject Characteristics

<table>
<thead>
<tr>
<th>Diagnostic Tests</th>
<th>Gender</th>
<th>Age (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POAG (201 subjects, 385 eyes)</td>
<td>Male</td>
<td>49.5 ± 11.8</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>51</td>
</tr>
<tr>
<td>NPG (107 subjects, 201 eyes)</td>
<td>Male</td>
<td>53.5 ± 10.3</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>55</td>
</tr>
<tr>
<td>Control (196 subjects, 383 eyes)</td>
<td>Male</td>
<td>44.0 ± 13.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>45</td>
</tr>
</tbody>
</table>

Data are the mean ± SD [median; range].

Institutional Review Board and Ethics Committee approvals were achieved at baseline of the ongoing glaucoma study (the “Erlangen Glaucoma Registry”).

Excluded from the study were eyes with more than 12% false-positive and false-negative responses in visual field testing, and visual acuity of 0.6 or better was a requirement. Fixation was automatically controlled by the perimetric device (Octopus-501, G1, three-phases, 59 measure points; Interzeag, Schlieren, Switzerland). In addition, eyes without available results of NRR measurements had to be excluded in view of the study’s rationale. All subjects had clear optic media, and no disease was found by slit lamp examination, gonioscopy, and funduscopy.

Patients

Examination was performed on 586 eyes from 308 patients (155 women; age range, 18–70 years) from the Erlangen Glaucoma Registry, fulfilling the criteria of optic nerve head damage (see Diagnostic Reference Criteria). Of the 586 examined eyes, 585 had primary open-angle glaucoma (POAG) and 201 had normal-pressure glaucoma (NPG).

Control Subjects

One hundred ninety-six control subjects (89 women; age range, 18–69 years) were recruited from the staff of the hospital and the university administration, for a total of 385 eyes. The control subjects had to have normal IOP, normal-appearing optic disc, and no visual field loss (Table 1).

Diagnostic Reference Criteria

Visual field loss and IOP were not inclusion criteria for the patient group. Glaucomas were defined through the following criteria for optic disc damage: glaucomatous changes of the optic nerve head, such as unusually small NRR in relation to the optic disc size; cup-to-disc ratios higher vertically than horizontally; an abnormal shape of the neuroretinal rim; and/or localized or diffuse retinal nerve fiber layer defects.15 Stereo optic disc photographs were taken with a telecentric fundus camera (Carl Zeiss Meditech, Oberkochen, Germany). For the morphometric analysis of optic disc and optic cup, outlines of the projected slides (scale, 1:15) were plotted on paper.14,16 Anterior corneal curvature and refractive error were considered for the correction of the ocular and camera magnification.17 The standard protocol, used for evaluation of all optic disc photographs, consisted of a checklist including the following variables: size and shape of the optic disc, shape of the optic cup, size of the optic cup in relation to the size of the optic disc; shape and size of the NRR (calculated as the difference between the expected area of the NRR adjusted for optic disc area minus cup area); presence of disc hemorrhages, location and extent of the a and b zones of parapapillary atrophy, diameter of the retinal arterioles, and visibility of the retinal nerve fiber layer. The evaluations were performed in a blinded fashion by two experienced examiners. In case of doubt, other reasons for optic nerve damage were excluded by additional neurologic, neuroradiologic, or medical examinations. Perimetry was performed in all subjects with a computerized static projection perimeter.18

Diagnostic Tests

Temporal Contrast Sensitivity Test (TCS). A white flicker light is presented to individuals who are seated in front of a 58-cm-diameter full-field bowl. At a constant frequency of 37 Hz at a time-averaged luminance of 10 cd/m², the flicker threshold is determined (Erlangen flicker test). No fixation by the subject is required. Corresponding to the pupil diameter and the Stiles-Crawford effect, each measurement is preceded by a correction of the mean luminance of the full-field bowl. The mean value of at least six threshold crossings enters the evaluation.19

Spatiotemporal Contrast Sensitivity Test (STCS). An alternating vertical sinusoidal stripe pattern is projected onto an upper temporal retinal area (13.8° horizontally, 4.2° vertically), using a white, square-shaped piece of cardboard surrounding a television screen with a high frame rate (DM2; Joyce Electronics Ltd., Cambridge, UK) that is illuminated by a bluish light of similar color and the same luminance as the screen. To determine contrast sensitivity, a two-alternative forced-choice procedure combined with a staircase-tracking procedure is used.20,21

Blue-on-Yellow Visually Evoked Potential (BYVEP). Using a two-channel Maxwellian view system with a Xenon arc lamp as the light source, a high-contrast 0.88 cyc/deg stripe pattern of blue light (460 nm, 3.3 × 10⁷ td) in one channel, superimposed on a homogeneous yellow adaptation light (570 nm, 1.3 × 10⁸ td) in the other channel, is presented to the individuals. Stimulation takes place in the onset-offset mode (200–500 ms). After amplification, 150 sweeps are averaged. Monopolar recordings are made from the inion against the left earlobe, with the right earlobe kept grounded. To check for reproducibility, two recordings are made. Peak time measurements of the onset responses are included in the statistical evaluation.22

Pattern-Reversal Electroretinogram (PERG). Only one channel of the same two-channel Maxwellian view system is used to record the PERG. At a frequency of 7.8 Hz a vertical high-contrast black-and-white, square-wave-stripped pattern with mean luminance of 4263 photopic td and a spatial frequency of 0.88 cyc/deg of the fundamental component is presented. A carbon-glide electrode hooked over the patient’s lower eyelid is taken to record the responses. After amplification, the responses are averaged and stored in a digital computer. Within one sweep of a 256-ms, four pattern-reversal responses are analyzed. Using discrete Fourier analysis, the amplitude of the second harmonic component of a total of 120 pattern reversal responses is evaluated.23

Static Automatic Perimetry. In our study, the mean perimetric defect (MD) quantified the extent of the visual field defect. Fixation losses are controlled through pupil videos taken by an integrated camera. If fixation loss occurs, an acoustic signal indicates that the measurement has to be repeated at the respective position.18 For each patient, all tests were explained in detail and performed on the same day and in the same sequence (MD, STCS, TCS, PERG, and BYVEP). Learning tests were not performed. With the exception of perimetry with a duration of approximately 25 minutes per eye, the diagnostic procedures took approximately 5 to 10 minutes per eye.

Statistical Methods

For all five diagnostic tests, age dependency was adjusted for linear regression analysis in the control group. The measure for disease severity was the loss of the neuroretinal rim (LNRR) area—that is, the difference between the expected area of the NRR adjusted for optic disc size by a linear regression analysis in the control group. Each test was dichotomized by fixing the specificity in the control group to 80%. Descriptive analysis included means, standard deviations, medians, and ranges for continuous variables (Tables 1, 2, 3). In significance tests, the generalized estimating equations (GEE) method was used to account for the dependency arising from multiple measurements at the same subject.24,25 The Bonferroni correction was used for multiple testing.

Because the specific logistic model used in the present investigation has not been applied before in the ophthalmology literature, it is described briefly: Given that an eye is actually glaucomatous (accord-
RESULTS

Subjects in the POAG and, more distinctively, in the NPG group were older than those in the control group. Furthermore, the patients with NPG were predominantly women (Table 1). The values for intraocular pressure (IOP), neuroretinal rim area (NRR), loss of neuroretinal rim area (LNRR), and mean perimetric defect (MD) and for the remaining four diagnostic procedures were more pathologic in the glaucoma group than in the control group (Tables 2, 3).

The results of the logistic regression model are presented in Table 4. The diagonal entries include sensitivities of the individual procedures at average LNRR in the diseased population (LNRR = 0.44 mm²), and odds ratios for increase in sensitivity per unit of increase in LNRR area (0.2 mm²). The results reveal significant (after Bonferroni correction for multiple testing) differences in average sensitivity between the electrophysiological (BYVEP, PERG) and the psychophysical (MD, TCS, STCS) tests, the latter being more sensitive for average disease stage (upper right entries). The amount of increase in sensitivity with increasing disease damage significantly (after Bonferroni correction) differed between TCS and the remaining two psychophysical procedures (lower left entries in Table 4). The graphic representation of these results is given in Figure 1. In early glaucoma stages, TCS revealed the highest sensitivity, for a fixed specificity of 80%. In the course of disease progression, induced by increasing LNRR area, the increase in sensitivity was steeper for STCS and MD than for the other procedures, so that in the most progredient stages MD, followed by STCS and BYVEP, showed the highest sensitivity. Especially, it can be seen that the slope of MD is significantly steeper than that of TCS. PERG showed minor sensitivity in all glaucoma stages.

In an additional analysis for the detection of possible interactions between test performance, IOP, and disease severity, it turned out that there was no difference in test performance between the NPG and the POAG group for the electrophysiological tests. For the psychophysical procedures, sensitivity was lower in the NPG group (P < 0.01). No further interaction with disease progression was detected. We present only the graphic display of the psychophysical test performances for NPG and POAG (Fig. 2).

DISCUSSION

In the present study, five different tests were evaluated, including automated white-on-white projection perimetry. There are some other reports in the literature of investigations that have used batteries of different psychophysical and electrophysiological tests to determine their usefulness in glaucoma diagnosis. The discussion of their findings in comparison with the present results is difficult because different stimulus paradigms were used and heterogeneous groups of patients and patients with suspected glaucoma in different stages were examined. None of the studies included disease severity itself into the analyses to rank the various tests under explicit consideration of their differing performance in different glaucoma stages. In the present cross-sectional analysis, the degree of the glaucomatous damage was graded on a continuous scale of the

Table 2. Eye-Specific Characteristics

<table>
<thead>
<tr>
<th>Group</th>
<th>IOP (mm Hg)</th>
<th>Neuroretinal Rim Area (mm²)</th>
<th>Loss of Neuroretinal Rim Area (mm²)</th>
<th>Mean Perimetric Defect (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glaucoma (n = 586)</td>
<td>26.6 ± 8.0</td>
<td>1.11 ± 0.40</td>
<td>+0.44 ± 0.39</td>
<td>+4.33 ± 4.79</td>
</tr>
<tr>
<td>Control (n = 383)</td>
<td>17.7 ± 2.7</td>
<td>1.66 ± 0.33</td>
<td>0.00 ± 0.32</td>
<td>+1.01 ± 1.27</td>
</tr>
</tbody>
</table>

Entries are mean ± SD, [median; range].

* The mean value in the control group is due to the construction of the LNRR, which is based on a linear regression analysis in the control group.

Table 3. Results of Diagnostic Procedures

<table>
<thead>
<tr>
<th>Group</th>
<th>TCS (log %⁻¹)</th>
<th>STCS (log %⁻¹)</th>
<th>BYVEP (ms)</th>
<th>PERG (µV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glaucoma (n = 586)</td>
<td>1.26 ± 0.27</td>
<td>1.64 ± 0.48</td>
<td>126.1 ± 14.1</td>
<td>3.03 ± 1.12</td>
</tr>
<tr>
<td>Control (n = 383)</td>
<td>1.48 ± 0.18</td>
<td>1.93 ± 0.13</td>
<td>117.1 ± 9.7</td>
<td>3.55 ± 1.12</td>
</tr>
</tbody>
</table>

Data are the mean ± SD [median; range].
NRR area of the optic disc, and the concomitant performance in different sensory tests can be judged (Figs. 1, 2).

Figures 1 and 2 of the present study suggest that in early stages of glaucoma, in POAG and NPG, the TCS is the most sensitive of all. The STCS and the perimetric MD seem to have rather low value in early glaucoma diagnosis, and both take a similar course as the glaucomatous neuropathy increases (Figs. 1, 2).

The sensitivity of perimetry is comparably lower in the beginning of the disease but becomes very important in late stages. However, we are aware of the fact that the MD ignores spatial clustering and dispersion of local damages. Especially, early glaucomas might be indicated by a few clusters of damage that do not shift the entire MD into the pathologic range. Thus, it is not the sensitivity of the entire perimetric measurement we are analyzing. In addition, more elaborated perimetric techniques have been developed recently that are able to detect optic neuropathy in glaucomatous eyes with normal standard visual fields.29 The curve of the STCS has a similar shape as that of perimetry. Although the test uses a temporally alternating pattern the stimulus is localized like perimetric stimuli and probes an area known to have a high susceptibility for glaucomatous damage.32

In our study, the electrophysiological procedures in general showed lower sensitivity than the psychophysical ones. The BYVEP, which takes a course rather similar to the temporal contrast-sensitivity curve, is our most sensitive electrophysiological test in early glaucoma stages and ranks directly behind the TCS there.

Several PERG studies regarded the PERG as an early diagnostic test.27,28,33–36 Figure 1, however, does not confirm these findings. In contrast, our results show a comparably poor sensitivity of PERG in early stages that cannot be compensated through the relatively steep course of the curve. In fact, Arden30 states that ERG is not a screening technique, although

![Disease severity (measured by loss of neuroretinal rim area)](image)

**Figure 1.** The logistic model for diagnostic test evaluation. With increasing disease severity (measured by increasing loss of neuroretinal rim area in square millimeters, displayed on the x-axis), sensitivity of all diagnostic procedures increased, but the individual curves are of different shape. The crossover that takes place identifies some tests as more valuable for early glaucoma diagnosis (TCS, BYVEP) and others as more appropriate for the detection of more advanced glaucomas (MD, STCS). All tests were dichotomized by fixing specificity to 80%.

**Figure 2.** Sensitivity of the three psychophysical tests depending on disease stage, additionally stratified by NPG versus POAG. As in Figure 1, with increasing disease severity, the sensitivity of the diagnostic procedures increased, and TCS was comparably higher in sensitivity in early glaucoma, whereas STCS and MD were more sensitive with progressing optic disc damage. In addition, two nearly parallel but shifted curves can be identified for each procedure, indicating different performance in the POAG (TCS, STCS, MD) compared with the NPG (TCS*, STCS*, MD*) group, respectively.

<p>| Table 4. Results of the Logistic Regression Model |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>MD</th>
<th>TCS</th>
<th>STCS</th>
<th>BYVEP</th>
<th>PERG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>61% (56–66)</td>
<td>0.08</td>
<td>0.67</td>
<td>0.001*</td>
</tr>
<tr>
<td></td>
<td>1.61 (1.41–1.85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCS</td>
<td>&lt; 0.0001*</td>
<td>56% (50–61)</td>
<td>0.18</td>
<td>0.04†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.23 (1.11–1.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STCS</td>
<td>0.31</td>
<td>0.001*</td>
<td>60% (54–65)</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.52 (1.34–1.72)</td>
<td></td>
</tr>
<tr>
<td>BYVEP</td>
<td>0.04†</td>
<td>0.20</td>
<td>49% (43–54)</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.37 (1.20–1.56)</td>
<td></td>
</tr>
<tr>
<td>PERG</td>
<td>0.20</td>
<td>0.03†</td>
<td>0.56</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagonal data entries: sensitivities (with 95% confidence intervals) of the diagnostic procedures at average loss of neuroretinal rim area (0.44 mm²) and odds ratios (with 95% confidence intervals) for growth in sensitivity per 0.2 mm² loss of neuroretinal rim area. Upper right entries: uncorrected P values for pair-wise comparisons of regression constants, referring to differences in sensitivity at average neuroretinal rim area. Lower left entries: uncorrected P values for pair-wise comparisons of regression coefficients, referring to differences in odds ratios (slopes of curves).

* Significant result after Bonferroni correction.
† Significant result without Bonferroni correction.
many psychophysical tests appear promising. On the contrary, Graham et al.\textsuperscript{28} and Drance et al.\textsuperscript{27} used a series of different functional tests. In their studies, PERG alterations were found the most sensitive indicators of glaucomatous damage. A conclusive comparison between studies concerning the sensitivity of PERG would be possible, if the relationship between stage of the disease and sensitivity had been quantified by the authors. In the work of Weinstein et al.\textsuperscript{35} individual data are given both for disease severity in terms of cup-to-disc ratio and results of the PERG (P1/N1). We have reanalyzed these data and found a U-shaped dependency of sensitivity on cup-to-disc ratio—implying that this ratio may not have been the only relevant indicator for disease severity in that study.

Stratification for NPG and POAG led to different performance in the case of the three psychophysical procedures (Fig. 2), whereas the electrophysiological tests performed nearly identically in both subgroups. Although the damage in POAG is rather diffuse, NPG seems to be characterized by local, more heterogeneous defects of the NRR\textsuperscript{37–38} that may be detected more reliably through the electrophysiological procedures than through the psychophysical ones used in this study. In perimetry, for example, the corrected loss variance (CLV) is greater in eyes with NPG than in those with POAG and equal LNRR area, and the curve for MD would be shifted up with additional consideration of the CLV. Similar effects might occur for the other two psychophysical procedures.

One of the study’s advantages lies in the considerable sample size available from the Erlangen Glaucoma Registry. However, we could not use diagnostic procedures that have been developed most recently. Although somewhat striking differences were shown between the five tests according to optic disc damage, the sensitivity of every test was rather poor for early disease. However, the method we have introduced may be of general interest in vision research concerning the mechanisms of various electrophysiological and psychophysical tests in the course of disease progression.

**Conclusions**

The results of the present study indicate that the diagnostic sensitivity of a particular sensory test depends on the degree of the glaucomatous optic neuropathy. With progressing disease damage, sensitivity increases for all tests, but the curves describing the increase in sensitivity with increasing neuropathy are not at all the same shape but show different steepness and may show crossover. This behavior may partly explain some of the discrepant results described in the literature.

Our findings suggest that the specific approach used in this study may be a standard technique in the evaluation of diagnostic tests in glaucoma. It provides an analytical tool to compare performance and appropriateness of various diagnostic measurements, depending on disease severity. It may advise the ophthalmologist which of several tests to use in which stage of glaucoma; and, furthermore, it may improve the comparison of results obtained from different populations in different studies.

**References**


6. Kerriigan-Baumrind LA, Quigley HA, Pease ME, Kerriigan DF, Mitchel RS. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophtalmol Vis Sci.* 2000;41:741–748.


