Optic Disc and Visual Field Progression in Ocular Hypertensive Subjects: Detection Rates, Specificity, and Agreement

Nicholas G. Strouthidis, Andrew Scott, Neena M. Peter, and David F. Garway-Heath

PURPOSE. To examine the detection rates, specificity, and agreement between visual field (VF) progression and Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, GmbH, Heidelberg, Germany) rim area (RA) progression in subjects with ocular hypertension (OHT).

METHODS. One hundred ninety-eight OHT and 21 control subjects were examined prospectively (1994-2001) with regular Humphrey VF (Carl Zeiss Meditec, Inc., Dublin, CA) and HRT testing. Point-wise linear regression (PLR) of sensitivity/time was used to assess VF progression, using standard and three-omitting (less stringent and stringent) criteria. The change in HRT-detected progression was assessed by linear regression of sectoral RA/time, defined as slope >1%/year, with significance level tailored according to series variability. Less stringent and stringent criteria were tested. Specificity was estimated by the proportions of control subjects with disease progression and significantly improving subjects (all). Agreement between disc and field progression in the subjects with OHT was assessed with specificities matched for both VF and HRT.

RESULTS. Specificity for VF PLR was estimated to be 85.7% to 95.4% when standard criteria were used, and for RA/time to be 88.1% to 90.5% with the less-stringent criteria. In this comparison, 21.2% progressed by RA alone and 20.2% by VF alone, and 12.1% progressed by both RA and VF. Specificity was estimated to be 95.2% to 98.2% for both VF PLR and RA/time, using the three-omitting criteria and the stringent RA/time criteria, respectively. In this comparison, 8.6% progressed by RA alone, 15.1% by VF alone, and 3.5% by both RA and VF.

CONCLUSIONS. A relatively high frequency of detected disease progression was observed with either method, with progression by VF occurring at least as frequently as progression by RA. Poor agreement between RA and VF progression was observed regardless of the specificity of the progression criteria. The results indicate that, in patients with ocular hypertension, monitoring of both VF and optic disc is necessary, as agreement between optic disc and VF progression is the exception rather than the rule. (Invest Ophtalmol Vis Sci. 2006; 47:2904–2910) DOI:10.1167/iovs.05-1584
come redundant, an alternative test should provide the investiga-
tor with, at a minimum, the same information on progression.

MATERIALS AND METHODS

Subject Selection

Subjects were selected from a cohort of 255 patients with OHT origi-

nally recruited to a betaxolol-versus-placebo study that took place at

Moorfields Eye Hospital between 1992 and 1997. Eligibility criteria for

this study are described in detail elsewhere. Briefly, OHT was defined

as in the OHT group). Control subjects were excluded if there

been randomized in the betaxolol-versus-placebo study. Briefly, ran-

involvement, and subsequently at 4-month intervals until September

24-2 program; Carl Zeiss Meditec, Inc.). Subjects had visual acuity at

21 mm Hg and normal findings in a baseline VF test (the same

criteria as in the OHT group). Control subjects were excluded if there

was a family history of glaucoma or any coexistent ocular or neurologic
disease.

Control subjects were selected from a cohort of 30 subjects who

were recruited among senior citizens and retirees, or were the spouses

or friends of subjects in the OHT cohort. These subjects had a baseline

IOP < 21 mm Hg and normal findings in a baseline VF test (the same
criteria as in the OHT group). Control subjects were excluded if there

was a family history of glaucoma or any coexistent ocular or neurologic
disease.

The same eye was analyzed in the present study as had originally

been randomized in the betaxolol-versus-placebo study. Briefly, ran-

domization was stratified according to risk of conversion to glaucoma,

were included for analysis, so that a minimum amount of potentially useful data was discarded.

The baseline visit was taken as the first visit in which both VF testing

and HRT imaging were performed. The study adhered to the tenets of

the Declaration of Helsinki and had local ethics committee approval, as

well as subjects’ informed consent.

Visual Field Analysis

A PLR of each subject’s VF series was performed using PROGRESSOR

(Institute of Ophthalmology, London, UK). Two strategies were used to

identify progression, one less stringent and the other stringent:

1. Standard criteria: A point is flagged as progressing if it shows a

significant negative slope of at least −1 dB/year, with a signif-

icance level of P < 0.01. This is the most commonly used PLR

method in published studies.11–13

2. Three-omitting criteria: A point is flagged as progressing if it

satisfies the standard criteria in each of three slopes. The first

slope is constructed using all time points up to time point n, the

second slope is constructed omitting point n and including the

next point in the series (n + 1), and the final slope is con-

structed omitting points n and n + 1 with the inclusion of the

next point in the series (n + 2). Increased specificity, compared

with the standard criteria, has been demonstrated applying the

three-omitting criteria to simulated VF data.16

VF series were identified as demonstrating significant improvement

using the same two strategies, except with a positive slope direction

(≥ +1 dB/year).

HRT Analysis

HRT mean topographies were generated and analyzed using Heidel-

berg Eye Explorer (ver. 1.7.0; Heidelberg Engineering, GmbH). Contour

lines were drawn by a single observer (NGS) onto the baseline

mean topographies, and these were exported automatically to the

subsequent images. A manual alignment facility was used to correct

contour line position if the automatically placed contour line was

misplaced or if there was a magnification change.17 When a satisfac-

tory contour line position could not be achieved, the mean topography

was excluded from the analysis. Eight mean topographies were excluded

from the analysis, as a result either of a double image being present or

if the image was so grainy as to preclude visualization of Elschnig’s

ring.

RA measurements were calculated for the six Explorer-defined disc

sectors (nasal, temporal, superotemporal, inferotemporal, superonasal,

and inferonasal). RA was selected in favor of other stereometric pa-

rameters, as it is consistently repeatable and reproducible and consti-

tutes a clinically meaningful parameter.3,4 All analyses were performed

using the 320-μm reference plane as it results in less RA variability

than the standard reference plane.17,18 Linear regression of sectoral RA over
time was performed for each subject’s HRT series. The standard devi-

ation of the residuals (RSD) was calculated for each linear regression

analysis. Residuals refer to the difference between the values observed

and those predicted by the regression equation. The RSD may be used to

estimate the variability of the series.19 High RSDs equate to high

variability and vice versa (Fig. 1).

The RSDs generated were ranked in order of magnitude within each

HRT sector. The 50th percentile RSD for each sector was used as a
cutoff to define the variability of the subject’s HRT series. RSDs less

than the 50th percentile were defined as having low variability, and

those greater than the 50th percentile as having high variability. The

probability at which a linear regression was considered significant was

adjusted according to the RSD, with tighter significance levels selected

for the high-RSD series. This approach was adopted to account for the

inclusion of all image qualities in the analysis, as poorer quality images

tend to generate higher RA variability.9 Progression in any disc sector

was determined whether a slope exceeding −1% of baseline RA/year

was observed, with significant improvement occurring where the

slope exceeds +1%/year. The slope value was selected as it is approx-
imately double the value of age-related RA loss estimated histologically

and using cross-sectional HRT data.20,21

Comparing Disc and Field Progression

In the absence of any established independent reference standard for

measuring glaucoma progression, the estimated specificity of both

\[ \text{Figure 1. Two scatterplots with linear regression slopes constructed using simulated numerical data to demonstrate different levels of variability. Left: data demonstrate large residuals (arrows), equating to a high RSD and high variability for the series. Right: data demonstrate much smaller residuals, indicating a low RSD and low variability for the series.} \]
tests should be closely matched (or anchored) for the comparison between them to be valid. Two proxy measures were used to estimate specificity: the number of subjects progressing in the control cohort and the number (in both the OHT and control cohorts) demonstrating significant improvement.

Specificity estimates were made for the PLR standard and three-omitting criteria. Two HRT progression strategies (a less stringent and a stringent strategy) were devised by adjustment of the significance levels of the regression analysis, so that the estimated specificities closely matched those of the two PLR VF techniques.

Agreement between VF and RA progression at the end of the study period was assessed in the OHT group, with specificity anchored at the two different levels. Where agreement existed, congruity between the disc and field sectors was assessed with a map designed to compare the anatomic relationship between VF test points in the Humphrey 24-2 test pattern and regions of the ONH (Fig. 2).

A further comparison between HRT RA progression and VF progression in the OHT group was made using matched numbers of VF and HRT tests, performed on the same day. This process was executed by using only the standard PLR criteria, as sufficient data points may not have been available to perform the confirmatory tests required of the three-omitting criteria. The thinned standard VF progression was compared against the less-stringent HRT criteria.

All statistical analyses were performed on computer (Medcalc, ver. 7.4.2.0; Medcalc Software, Mariakerke, Belgium).

### RESULTS

One hundred ninety-eight subjects with OHT and 21 control subjects were included for analysis in the study. The male-female ratio was 105:93 in the OHT group, compared with 14:7 in the control group. Ninety-five right eyes were used in the OHT group and 11 in the control group. The demographics of subjects in both groups are depicted in Table 1.

### Estimation of Specificity

A specificity range of 85.7% to 95.4% was estimated with the standard PLR criteria and a range of 95.2% to 98.2% with the three-omitting criteria (Table 2).

The specificity estimate for the less-stringent HRT progression strategy (significant slope >1%/year with significance levels of $P < 0.05$ for low-variability series and $P < 0.01$ for high-variability series) was 88.1% to 90.5% (Table 3). The specificity estimate for the stringent HRT progression strategy (significant slope, >1% per year with significance levels of $P < 0.001$ for the low-variability series and $P < 0.0001$ for the high-variability series) was 95.2% to 98.2% (Table 3).

### Comparison of Disc and Field Progression

Using the less-stringent criteria (specificity for both tests anchored at approximately 90%), there was agreement between HRT and VF for progression in 24 subjects within the OHT group. A further 42 subjects showed progression by HRT alone compared with 40 by VF alone (Fig. 3).

When the stringent criteria were used (specificity anchored at approximately 97%), seven subjects within the OHT cohort showed progression by both HRT and VF. A further 17 subjects showed progression by HRT alone, compared with 30 subjects by VF alone (Fig. 4).

Of the 24 subjects with disease progression shown by disc and field with the less-stringent criteria, there was congruity in at least one disc and field sector in 14 (58.3%), compared with 7 (100%) of 7 subjects when the stringent criteria were applied. There was congruity in four sectors in one subject, in two sectors in four subjects, and one sector in eight subjects with the less-stringent criteria. With the less-stringent criteria, congruity was detected in one sector in all seven subjects.

When an equal number of HRT and VF examinations was performed on the same day (i.e., thinned VF data), progression was shown in 22 subjects (11.1%) within the OHT cohort by both HRT and VF. A further 44 subjects (22.2%) showed progression by HRT alone, compared with 25 subjects (12.6%) by VF alone.

### Example Patient

Subject 9 was a man aged 58 years who was followed up for 6.5 years; the left eye was analyzed. He underwent 15 VF examinations (Fig. 4).
nations (baseline MD, −1.17 dB) and 11 HRT examinations (baseline global RA, 1.27 mm²). The range of MPHSD values was 13 to 26 (median 20). Figure 5 depicts the HRT mean topographies at baseline and at the conclusion of the study. Scatterplots show significant decay of sectoral RA over time (less-stringent criteria). The following slopes of RA/time were generated: ST sector, 3.8%/year ($P = 0.005, R^2 = 0.6$); IT sector, 3.8%/year ($P < 0.0001, R^2 = 0.9$); SN sector, 3.9%/year ($P < 0.0001, R^2 = 0.9$); and IN sector, 5.0%/year ($P = 0.003, R^2 = 0.6$). Within the final VF gray-scale, points progressing by standard PLR criteria are highlighted. There was complete congruity between the four disc and field sectors identified as progressing.

### Discussion

An inevitable shortcoming with any study that compares methods of detecting glaucoma progression is the absence of a reference standard method based on which the diagnostic accuracy of the tests used may be defined. The choice of progression criteria is often arbitrary. The relative merits and shortcomings of PLR analysis have been well documented. The adoption of linear regression is supported by modeling of sensitivity/time in series of fields from normal tension glaucoma subjects. A linear model demonstrates an adequate data fit and also generates more accurate prediction of future perimetric change than do higher-order models. PLR using PROGRESSOR may detect progression earlier than STATPAC 2, an event-analysis technique native to the Humphrey Field Analyzer system (Carl Zeiss Meditec, Inc.). However, sensitivity, specificity, and time to detect progression are influenced by the progression criteria selected and, in the absence of an independent reference standard for progression, performance may be judged by data modeling. In such a model, PLR techniques have been shown to have high specificity and to detect progression in more patients but they may require more tests (or time) than event-analysis techniques.

However, the longer time required by PLR may, in part, reflect the greater number of patients in whom progression is detected, with patients in whom it is not detected at all, by “event analyses” taking longer to detect by PLR. To achieve satisfactory levels of sensitivity and specificity, PLR requires a minimum of seven or eight VFs. This requirement was comfortably exceeded in the majority of subjects in both groups of the present study.

The pattern of RA loss in glaucoma is not yet clearly established. Airaksinen et al., described three patterns of RA change over time—linear, episodic, and curvilinear—with approximately half of both OHT and glaucoma subjects demonstrating a linear decay. This observation supports the adoption of a linear model to detect HRT RA progression, although linear regression may miss some episodic change, particularly if it occurs in a highly variable series. In addition, it should be appreciated that measurement noise may mimic stepwise and nonlinear change. As with PLR, detection of progression is improved, with a greater amount of data included in the analysis. In this study, there were fewer HRT examinations, compared with VF examinations, in the OHT group, because of less frequent HRT imaging. Some subjects have a particularly high number of VF examinations, as VF conversion was the study end point. These subjects were tested intensively over 3-month periods when VF conversion was suspected, which may introduce some bias favoring the detection of VF progression, as subjects deemed to have converted to glaucoma within the original study have a greater number of VF examinations relative to HRT examinations. This finding has been confirmed by repeating the comparison using a matched number of HRT and VF examinations, which resulted in a much lower detection rate by VF (12.6% compared with 20.2%), although it had very little effect on the level of agreement as regards progression status (11.1% compared to 12.1%). A much higher detection rate of progression by VF was observed in this study compared with PLR.

### Table 2. Estimation of Specificity for Visual Field Point-Wise Linear Regression for Both Standard and Three-Omitting Criteria

<table>
<thead>
<tr>
<th></th>
<th>Standard Criteria</th>
<th>Three-Omitting Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>No significant</td>
<td>209/219 = 95.4%</td>
<td>215/219 = 98.2%</td>
</tr>
<tr>
<td>improvement</td>
<td>95% CI</td>
<td>95.4%–99.3%</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No progression in</td>
<td>18/21 = 85.7%</td>
<td>20/21 = 95.2%</td>
</tr>
<tr>
<td>control cohort</td>
<td>95% CI</td>
<td>77.3%–99.2%</td>
</tr>
<tr>
<td>Specificity estimate</td>
<td>85.7%–95.4%</td>
<td>95.2%–98.2%</td>
</tr>
</tbody>
</table>

### Table 3. The Estimation of Specificity for Two Different HRT Progression Techniques

<table>
<thead>
<tr>
<th></th>
<th>Less Stringent*</th>
<th>Stringent†</th>
</tr>
</thead>
<tbody>
<tr>
<td>No significant</td>
<td>197/219 = 98.1%</td>
<td>215/219 = 98.2%</td>
</tr>
<tr>
<td>improvement</td>
<td>95% CI</td>
<td>95.4%–99.3%</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No progression in control</td>
<td>19/21 = 90.5%</td>
<td>20/21 = 95.2%</td>
</tr>
<tr>
<td>cohort</td>
<td>95% CI</td>
<td>77.3%–99.2%</td>
</tr>
<tr>
<td>Specificity estimates</td>
<td>88.1%–90.5%</td>
<td>95.2%–98.2%</td>
</tr>
</tbody>
</table>

* Slope >1% baseline sector RA/year; $P < 0.05$ for low-variability series; $P < 0.01$ for high-variability series.
† Slope >1% baseline sector RA/year; $P < 0.001$ for low-variability series; $P < 0.0001$ for high-variability series.
with previously published reports with similar length of follow-up.\textsuperscript{28,29} The higher rate is most likely a reflection of the use of linear regression-based techniques that have been shown to have higher detection rates than event-analysis techniques when sufficient data and length of follow-up are available.\textsuperscript{24} In addition, as optic disc appearance was not an entry criterion for the study, there may have been more subjects with OHT in the present study with early glaucomatous damage than there were in the similar studies. High rates of structural progression may be explained by the detection method. Analysis of structural progression in the present study was a quantitative analysis of CSLO images, which has been shown in a primate model of glaucoma to be more sensitive than the subjective evaluation of stereo ONH photographs.\textsuperscript{30}

The estimation of specificity in this study was derived from proxy measures based on two assumptions. First, significant RA and differential light sensitivity (DLS) improvement over time should not occur. Second, RA and DLS loss at a much greater rate than the age-related estimate should not be observed in control subjects. One may argue that these assumptions are not wholly valid. VF sensitivity may improve because of a learning effect. However, significant learning effects are unlikely in this study because it was required that subjects have reproducibly normal and reliable VFs at baseline.\textsuperscript{12} When the topographical change analysis is used to monitor HRT progression, positive topographical changes may be seen when there are also areas of topographic depression.\textsuperscript{31} However, isolated improvement is rare. It is uncertain whether such changes exert any sustained influence over the longitudinal RA trend in a given sector. Both RA and VF sensitivity probably decline with ageing. Our rate criteria were set to be higher than average (estimated) rates, but individuals are likely to age at different rates, and so it is possible that some true, age-related, change was detected. Thus, our estimates of specificity may, in fact, constitute an underestimation of true specificity.

The specificity estimates relate to the analysis performed in this study, with progression rates calculated only at the end of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Disc and field progression in subject 9. (a) Visual field grayscale at the conclusion of the study; the \textit{hatched squares} indicate the test-points flagged as progressing according to the standard PLR criteria; (b) HRT baseline mean topography; (c) HRT mean topography at the conclusion of the study period. (d-g) Scatterplots with regression line for sectoral rim area against time.}
\end{figure}
the observation period. Were the same criteria applied in the clinical setting, lower specificity would result because the clinician analyzes data for progression at each patient visit, rather than at the end of a specified period. The result is reduced specificity from multiple testing.

In this study, there was poor agreement between ONH progression and VF progression in subjects with OHT, regardless of the specificity of the progression criteria applied. A similar result was reported in a group of subjects with established glaucomatous VF loss followed longitudinally.\(^\text{32}\) In that study, event- and trend-based progression analysis methods were assessed by SAP, high-pass resolution perimeter, and HRT. For the trend-based analysis, agreement between all three tests was observed in 2.4% to 7.1% of subjects, depending on the stringency of the criteria. This closely matches the results of the present study, with agreement for progression of 3.5% to 10.6%. The results of these two studies suggest that poor agreement is seen regardless of the stage of disease and criteria for monitoring progression. One may therefore conclude that, as both VF and HRT testing provide additive information on progression, the two cannot be used interchangeably. Therefore, a single test, either of structure or of function, is not yet sufficient for monitoring glaucoma progression in isolation.

Structural damage and loss of visual function are both features of glaucoma. There is a popularly held view that structural changes are detectable before functional changes in progressive glaucoma by using currently available techniques.\(^\text{7}\) Some structural changes may occur without concomitant changes in function. This theory was expounded by Fuchs as far back as 1916,\(^\text{33}\) and has been supported more recently by descriptions of lamina bowing and astrocytic remodeling in studies of experimental glaucoma.\(^\text{34,35}\) Another possibility is that functional loss may occur without concomitant structural alteration: electrophysiological evidence of IOP-mediated ganglion cell dysfunction in the primate retina has recently been described.\(^\text{36}\) An explanation for the high frequency of detected VF progression in the present study, in the absence of concomitant identifiable structural progression in many eyes, is that a functional reserve may have been expended before the time of entering into the study. This may be pertinent, as the appearance of the ONH was not taken into account at the time of recruitment. However, this theory is not supported by the results of Artes and Chauhan,\(^\text{32}\) who observed a similarly poor level of agreement in subjects with established glaucoma. A recent study, also at odds with the view that structural damage precedes function damage, has demonstrated evidence of electrophysiological changes detectable before structural changes in subjects with OHT.\(^\text{37}\)

A likely explanation for the apparent structure–function dissociation lies in measurement variability. Even if disc and VF progression rates were identical, differences in measurement noise between the two testing modalities may result in progression’s being detected by one modality, but not the other.

Steps taken to tailor the HRT progression criteria according to RA variability reduce the effect of differing image qualities and measurement noise on false-positive progression detection. Such a reduction has not been possible for the VF analysis, primarily because there is no satisfactory metric with which to measure visual field reliability.\(^\text{38,39}\) The issue of visual field variability over time is a complex one, primarily as variability increases in areas of established depressed sensitivity.\(^\text{40,41}\) and is beyond the scope of the present study.

It is possible that improved agreement may be observed with further refinement of imaging and visual function tests and data analysis techniques. Machine learning classifiers, both supervised and unsupervised, may be useful in detecting true progression from measurement noise.\(^\text{42,43}\) Monitoring topographic changes with a recently described technique, statistic image mapping, may be more useful than monitoring stereometric parameters (such as RA).\(^\text{44}\) The adoption of VF spatial filtering techniques may also ameliorate our ability to discriminate perimetric progression.\(^\text{45}\) However, until these developments are fully tested and validated, the results of our study indicate that both structural and functional tests should be monitored to identify progression in patients with OHT.

Acknowledgments

The authors thank Tuan Ho for editing the manuscript.

References

17. Strouthidis NG, White ET, Owen VM, Ho TA, Garway-Heath DF. Improving the repeatability of Heidelberg retina tomograph and


