Improving the repeatability of Heidelberg retina tomograph and Heidelberg retina tomograph II rim area measurements

N G Strouthidis, E T White, V M F Owen, T A Ho, D F Garway-Heath

Aims: To devise and test strategies for improving Heidelberg retina tomograph (HRT and HRT-II) rim area (RA) repeatability and assess the benefit of the strategies in time series of HRT images.

Methods: The effect of the standard and 320 μm reference planes and image quality on RA repeatability was assessed in a test-retest HRT image dataset from 74 subjects. A longitudinal HRT image dataset from 30 ocular hypertensive subjects was analysed by linear regression of RA over time, with each of the reference planes and using a manual image alignment facility. RA variability was estimated by comparing the standard deviation of residuals (RSD) generated by each linear regression.

Results: RA repeatability was better with the 320 μm reference plane (repeatability coefficient 0.17 mm²), improving further with only good quality images (repeatability coefficient 0.08 mm²). For the longitudinal data, a significant (p < 0.0001) reduction in the RSD from 0.10 to 0.05 mm² was obtained with the 320 μm reference plane. Manual alignment led to a further significant (p < 0.0001) reduction in the RSD to 0.04 mm².

Conclusions: The findings support the use of a 320 μm reference plane and manual image alignment to analyse RA over time. The estimates of RA repeatability may be used to define thresholds for glaucomatous change.

Monitoring of progression is a key aspect in the management of glaucoma. As structural changes to the optic nerve head (ONH) may precede detectable functional deficits, an objective method for measuring these changes is desirable. Scanning laser tomography by the Heidelberg retina tomograph (HRT) is an established technique for ONH imaging. This technique has been shown to be reproducible and capable of detecting structural changes before repeatable visual field (VF) defects. A newer version of the HRT, HRT-II, has been shown to have good reproducibility. It is a semi-automated device designed for use by several different operators in the clinic.

To detect reliably the progression of morphometric parameters, measurement error should be determined and minimised. In the companion paper, rim area (RA) and mean cup depth consistently demonstrated high repeatability and reliability. Of these two parameters, RA is the more clinically meaningful. RA is also useful in discriminating between normal, ocular hypertensive, and glaucoma eyes. RA therefore constitutes an appropriate candidate for monitoring glaucomatous progression.

RA variability was influenced by the inter-test reference height difference and by image quality (mean pixel height standard deviation, MPHSD). The authors propose that RA variability may be improved with the use of an alternative to the Standard reference plane and by taking into account image quality. This reference plane is located parallel to the retinal surface and delineates the “neuroretinal rim” (structures anterior to the plane) from the “cup” (space posterior to the plane). The Standard reference plane is located 50 μm posterior to the height of the contour line at the temporal disc margin. The 320 μm reference plane is a plane located 320 μm posterior to a reference ring located in the image periphery. This reference ring is used to define the stable, zero plane of the topography image.

In this study, strategies to reduce RA measurement error were devised and tested in a series of longitudinal HRT data to determine if they may be used to improve the detection of progressive RA loss.

METHODS

Test-retest data acquisition

The test-retest acquisition protocol is described in the companion paper.

Image analysis

Heidelberg Eye Explorer (Version 1.7.0; Heidelberg Engineering, Heidelberg, Germany) was used to generate and analyse the mean topography images for both HRT and HRT-II images (HRT Explorer and HRT-II Explorer). In addition, the HRT images were analysed using the HRT Classic option on the Explorer platform (which simulates the older DOS software). The default Standard reference plane was applied.

Improving repeatability

The test-retest RA values were recalculated using the 320 μm reference plane, the effect of which on RA repeatability, compared with the standard reference plane, was analysed using Bland-Altman plots. Eyes with poor quality images were then excluded, and repeatability was assessed using only good quality images (MPHSD < 0.25).

Estimation of measurement error

Repeatability coefficient (RC) is the method for defining measurement error that has been adopted by the British Standards Institution, and is calculated as:

$$RC = 2\times\left( \frac{\sum (observation\ 1 - observation\ 2)}{n(observations)} \right)^{0.5}$$

where 95% of inter-test differences will be within this error. Subject eyes were classified according to image quality; good eyes were excluded. For this study, the MPHSD was used as a measure of image quality.

Abbreviations: HRT, Heidelberg retina tomograph; OHT, ocular hypertension; ONH, optic nerve head; POAG, primary open angle glaucoma; RA, rim area; RSD, residual standard deviation; VF, visual field
quality (MPHSD <21), medium quality (MPHSD 21–35), and poor quality (MPHSD >35). For each patient, follow up in a clinical setting was simulated using the following strategy. The mean RA obtained from two images taken on the same day by a single observer (ETW) served as the baseline reference value, which was compared with the “follow up” RA from an image taken on a subsequent day by a different observer (NGS). An RA repeatability coefficient, derived for images analysed with the 320 μm reference plane, was calculated for each of the three image quality categories.

Longitudinal HRT series: patient selection

The improvement strategies were also tested on a longitudinal dataset obtained from a cohort of OHT subjects recruited to take part in a betaxolol versus placebo study, the details of which have been described elsewhere. Unlike the test-retest set, where RA was expected to remain constant, the RA of the longitudinal dataset was expected to change over time. The study adhered to the tenets of the Declaration of Helsinki and had local ethics committee approval and subjects’ informed consent. The subjects underwent regular HRT imaging, initially at yearly intervals for the first 2 years of their involvement, and 4 monthly subsequently. HRT imaging at this latter frequency was continued until 2001, when imaging was continued using the HRT-II. Thirty subjects identified as converting to glaucoma (consistent Advanced Glaucoma Intervention Study visual field score >0) were selected. The subjects’ mean age was 61.8 years (range 43.4–71.7 years), mean number of HRT tests was 11 (range 9–16), and mean length of follow up was 6.3 years (median 6.2 years, range 4.8–7.2 years).

Application of “improvement” strategy to longitudinal HRT series

For each image in a series, the global RA was calculated using both the standard and 320 μm reference planes. All three image quality categories were included because the exclusion of poorer quality images (MPHSD >25) from a longitudinal series of approximately 11 images would have resulted in a substantial loss of valuable data. When using the Explorer software, the contour line is manually drawn around the disc margin of the baseline image and then automatically exported to all follow up images. This automatic procedure sometimes gives rise to misalignment between the contour line and the ONH, especially when follow up images are of poor quality (high MPHSD values or magnification changes). A novel manual alignment facility, which adjusts for these anomalies, has been introduced (fig 1). Contour line alignment is achieved by selecting a minimum of four common landmarks (for example, vessel bifurcations) between the baseline and follow up images. A flicker chronoscopy, or “toggle,” enables the closeness of fit to be assessed.

Linear regression of global RA over time (years) was plotted using non-aligned and aligned images obtained with both reference planes. The residual standard deviation (RSD) of the linear regression was used to gauge the variability of each series.

Figure 1 An example of a misplaced contour line on a follow up HRT image in the Explorer software. Manual alignment is achieved by selecting a minimum of four common landmarks between the baseline and follow up images to attain the best approximation of alignment. A flicker chronoscopy enables the closeness of fit to be assessed before accepting the changes.

Figure 2 Bland-Altman plots of HRT Explorer interobserver/intervisit rim area (mm²) using the Standard reference plane (A), the 320 μm reference plane (B) and the 320 μm reference plane with inclusion of only good quality (MPHSD < 25) images (C).
for the HRT-II was minimal (0.07–0.09 mm²), compared with the Standard reference plane were 0.24 mm² (HRT Classic), 0.16 mm² (HRT Explorer), and 0.10 mm² (HRT-II Explorer) when only good quality images (MPHSD <25) were used (table 1).

Measurement error
The measurement error for good quality images (MPHSD <21) was similar for all three methods, with a mean repeatability coefficient of 0.06 mm² (table 2).

A stepwise increase in measurement error was first seen with medium quality images (MPHSD 21–35) and then with poor quality images (MPHSD >35). The incremental change of measurement error from good to medium quality images for the HRT-II was minimal (0.07–0.09 mm²), compared with the HRT (0.05–0.16 mm², using HRT-Classic). By contrast, the measurement error for poorer quality images was greater for the HRT-II compared with the HRT (0.27 mm² and 0.20 mm², respectively).

**RESULTS**

**Test-retest data**

Bland-Altman plots show the improvement of HRT Explorer interobserver/intervisit RA repeatability obtained with the 320 μm reference plane (fig 2B). Compared with the Standard reference plane (fig 2A), the distribution with the 320 μm reference plane was tighter, with narrower limits of agreement. Repeatability was further improved when only good quality images (MPHSD <25) were used (fig 2C), but the number of subjects was reduced from 74 to 35.

The interobserver/intervisit RA repeatability coefficients with the Standard reference plane were 0.24 mm² (HRT Classic), 0.16 mm² (HRT Explorer), and 0.23 mm² (HRT-II Explorer). These values were reduced to 0.16 mm² (HRT Classic), 0.10 mm² (HRT Explorer), and 0.20 mm² (HRT-II Explorer) with the 320 μm reference plane, and further reduced to 0.09 mm² (HRT Classic), 0.11 mm² (HRT Explorer), and 0.10 mm² (HRT-II Explorer) when only good quality images (MPHSD <25) were used (table 1).

**Measurement error**

The measurement error for good quality images (MPHSD <21) was similar for all three methods, with a mean repeatability coefficient of 0.06 mm² (table 2).

A stepwise increase in measurement error was first seen with medium quality images (MPHSD 21–35) and then with poor quality images (MPHSD >35). The incremental change of measurement error from good to medium quality images for the HRT-II was minimal (0.07–0.09 mm²), compared with the HRT (0.05–0.16 mm², using HRT-Classic). By contrast, the measurement error for poorer quality images was greater for the HRT-II compared with the HRT (0.27 mm² and 0.20 mm², respectively).

**Longitudinal analysis**

The use of a 320 μm reference plane significantly (paired t test, p<0.0001) reduced the RSD, compared with the Standard reference plane (table 3). There was a further small, but significant (paired t test, p<0.0001), reduction in RSD when the 320 μm reference plane was used with manual alignment of misaligned images, compared with automatic alignment alone. The reduction in RSD was small and significant (paired t test, p = 0.02) when the Standard reference plane was used with manual alignment. A mean of four (median 4, range 0–11 images) images required manual alignment in each longitudinal series.

The scatter plots obtained with regression lines and RSD plots of global RA over time for an example patient (fig 3) show the improvement gained with the 320 μm reference plane and manual alignment. These data were obtained with the left eye of a 71 year old female subject examined over a 6 year period with 12 HRT tests. A closer relation between RA and time was observed with the 320 μm reference plane compared with the Standard reference plane (figs 3A and B).

Manual alignment led to a further narrowing of the 95% confidence limits of the regression slope (figs 3C) and reduction of the RSD (figs 3D–F).

**DISCUSSION**

Although scanning laser tomography is an established technique for measuring ONH structure, a consensus does not yet exist as how to best monitor progression. Chauhan et al have described a method for detecting progression based on serial change in topographical height within a super pixel. This produces a colour coded probability map, indicating locations in the image where the change in...
topographical height between baseline and follow up is greater than that expected by chance. This “change probability” technique has been used to examine longitudinal data.22

Another pixel based technique, which establishes the significance of the trend of pixel height change over time, has been described recently and shows promise.23 Although this technique gives a useful graphical indication of the location of change, there is not yet a standardised method for quantifying the changes observed.

An alternative approach is to examine the values of stereometric parameters, generated by HRT image analysis, over time. It is possible to calculate a rate of change, which has important implications for patient management. The test-retest studies described in the companion paper have shown that RA has consistently high repeatability and reliability.12 It is therefore a good candidate for serial analysis to detect change.

A disadvantage of using stereometric parameters is the dependence on the position of the reference plane. The location of the Standard reference plane was selected on the basis of the mean surface inclination angle of the ONH,24 and because it coincides with the papillomacular bundle—a site predicted to be minimally affected until glaucoma is advanced. This has not been supported by retinal thickness measurements derived from optical coherence tomography (OCT) in glaucoma subjects.25 The mean retinal nerve fibre layer thickness estimated by OCT has been used to set a novel reference plane that may be useful to detect early glaucoma,26 however, because the reference plane height is linked to the height of the disc margin contour, it may be expected to shift posteriorly as glaucoma progresses. The rates of RA change with the various reference plane options have yet to be compared.

In this study, we have made estimates of interobserver/visit RA measurement error for different degrees of image quality. These values may be used to establish probability limits for change when comparing RA between two HRT or HRT-II images. An RA difference exceeding the repeatability coefficient at the appropriate mean image quality can be defined as surpassing a “threshold” for change, at which glaucomatous change is distinct from measurement noise. These thresholds could be incorporated

Table 1 Application of different strategies to test-retest data: rim area repeatability for HRT Classic, HRT Explorer, and HRT-II Explorer in various test-retest protocol permutations

<table>
<thead>
<tr>
<th>HRT type</th>
<th>Strategy</th>
<th>Rim area repeatability coefficient (mm²)</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRT Classic</td>
<td>Standard reference plane</td>
<td>0.21</td>
<td>0.29</td>
<td>0.19</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane</td>
<td>0.10</td>
<td>0.14</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane, MPHSD &lt;25</td>
<td>0.04</td>
<td>0.09</td>
<td>0.07</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>HRT Explorer</td>
<td>Standard reference plane</td>
<td>0.28</td>
<td>0.31</td>
<td>0.30</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane, MPHSD &lt;25</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>HRT II Explorer</td>
<td>Standard reference plane</td>
<td>0.28</td>
<td>0.34</td>
<td>0.24</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>320 µm reference plane, MPHSD &lt;25</td>
<td>0.08</td>
<td>0.07</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

MPHSD, mean pixel height standard deviation; I, intraobserver/intravisit; II, interobserver/intravisit; III, intraobserver/intervisit; IV, interobserver/intervisit.

Table 2 Interobserver/visit rim area repeatability coefficients (mm²) obtained with different image qualities and using different types of HRT and the 320 µm reference plane

<table>
<thead>
<tr>
<th>Image quality</th>
<th>HRT Classic</th>
<th>HRT Explorer</th>
<th>HRT-II Explorer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of subjects</td>
<td>RC</td>
<td>No of subjects</td>
</tr>
<tr>
<td>Good quality image (MPHSD ≤21)</td>
<td>28</td>
<td>0.05</td>
<td>28</td>
</tr>
<tr>
<td>Medium quality image (MPHSD 21–35)</td>
<td>27</td>
<td>0.16</td>
<td>27</td>
</tr>
<tr>
<td>Poor quality image (MPHSD &gt;35)</td>
<td>19</td>
<td>0.20</td>
<td>19</td>
</tr>
</tbody>
</table>

MPHSD, mean pixel height standard deviation.
into the Explorer software to facilitate detection of glaucomatous progression.

In summary, this study highlights the benefit of using the 320 μm reference plane and a manual alignment facility in reducing RA variability. Thresholds for change based on RA measurement error at different image qualities have also been proposed. These findings may be used as the basis for devising new strategies to detect RA progression.

---

**Table 3**  Standard deviations of residuals for longitudinal HRT image series analysed using HRT Explorer

<table>
<thead>
<tr>
<th>Strategy used for longitudinal image series</th>
<th>Standard deviation of residuals (mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard reference plane; non-aligned</td>
<td>Mean: 0.10  Median: 0.10  Range: 0.01–0.40</td>
</tr>
<tr>
<td>Standard reference plane; aligned</td>
<td>Mean: 0.09  Median: 0.08  Range: 0.01–0.40</td>
</tr>
<tr>
<td>320 μm reference plane; non-aligned</td>
<td>Mean: 0.05  Median: 0.04  Range: 0.01–0.10</td>
</tr>
<tr>
<td>320 μm reference plane; aligned</td>
<td>Mean: 0.04  Median: 0.03  Range: 0.01–0.10</td>
</tr>
</tbody>
</table>

---

**REFERENCES**

Improving the repeatability of Heidelberg retina tomograph and Heidelberg retina tomograph II rim area measurements

N G Strouthidis, E T White, V M F Owen, T A Ho and D F Garway-Heath

Br J Ophthalmol 2005 89: 1433-1437
doi: 10.1136/bjo.2005.067306

Updated information and services can be found at:
http://bjo.bmj.com/content/89/11/1433

These include:

References

This article cites 25 articles, 6 of which you can access for free at:
http://bjo.bmj.com/content/89/11/1433#BIBL

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/