

SCIENTIFIC REPORT

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

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

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 mm^2), improving further with only good quality images (repeatability coefficient 0.08 mm^2). For the longitudinal data, a significant ($p < 0.0001$) reduction in the RSD from 0.10 to 0.05 mm^2 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 mm^2 .

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,^{1–5} 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to be reproducible,^{6–8} and capable of detecting structural changes before repeatable visual field (VF) defects.⁹ A newer version of the HRT, HRT-II, has been shown to have good reproducibility.¹⁰ It is a semi-automated device designed for use by several different operators in the clinic.¹¹

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.¹² Of these two parameters, RA is the more clinically meaningful. RA is also useful in discriminating between normal, ocular hypertensive, and glaucoma eyes.^{13–15} 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).¹² 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.¹²

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.¹⁶ Eyes with poor quality images were then excluded, and repeatability was assessed using only good quality images (MPHSD <25).

Estimation of measurement error

Repeatability coefficient (RC) is the method for defining measurement error that has been adopted by the British Standards Institution,¹⁷ and is calculated as:

$$RC = 2 * \left(\sqrt{\frac{\sum(\text{observation 1} - \text{observation 2})^2}{n(\text{observations})}} \right)$$

where 95% of inter-test differences will be within this error. Subject eyes were classified according to image quality: good

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

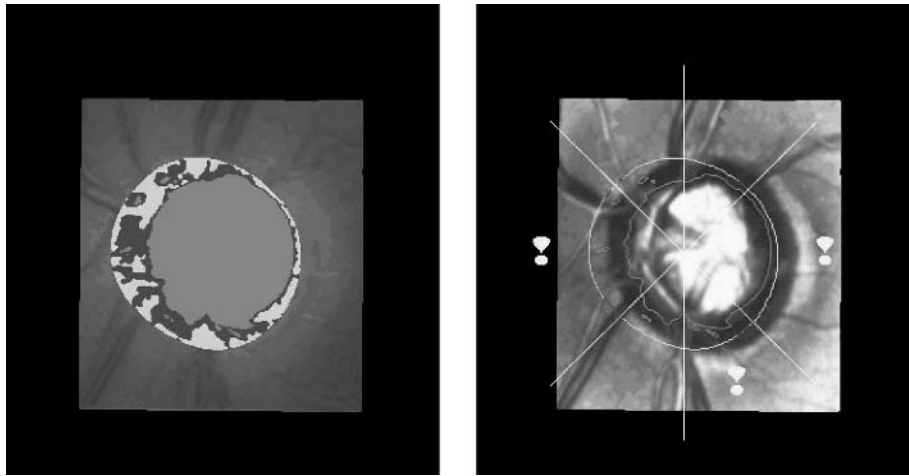


Figure 1 An example of a misplaced contour line on a follow up HRT image in the Explorer software. Manual alignment is achieved by selecting a minimum of four common landmarks between the baseline and follow up images to attain the best approximation of alignment. A flicker chronoscopy enables the closeness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

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.¹⁸ 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.¹⁹ 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.²⁰

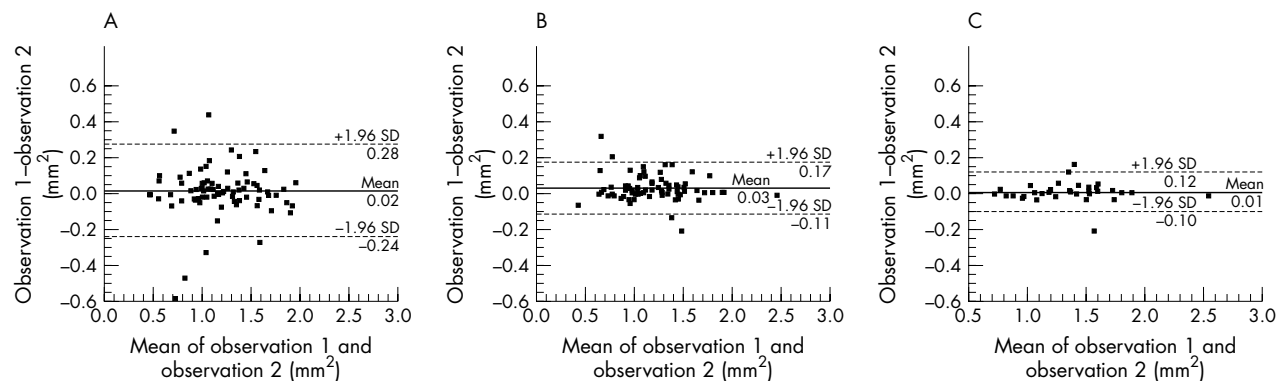


Figure 2 Bland-Altman plots of HRT Explorer interobserver/intervisit rim area (mm^2) using the Standard reference plane (A), the 320 μm reference plane (B) and the 320 μm reference plane with inclusion of only good quality (MPHSD <25) images (C).

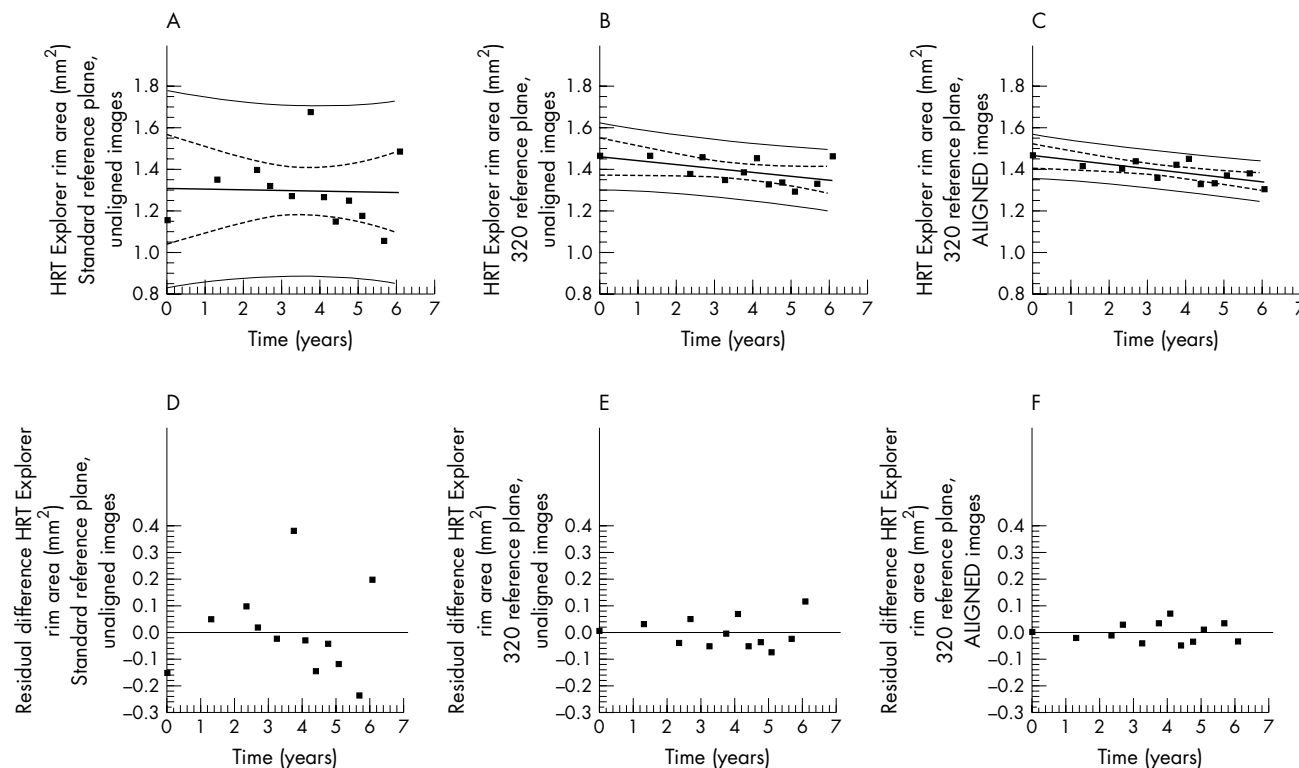


Figure 3 Linear regression of HRT Explorer global rim area over time using the Standard reference plane (A), 320 μm reference plane (B), and 320 μm reference plane with manual alignment (C). In these graphs, the solid line represents the regression line, the broken lines represent the 95% confidence limits, and the dotted lines represent the 95% prediction limits. Plots of the residuals from the linear regression analysis are shown in (D-F).

All statistical analyses were performed using Medcalc Version 7.4.2.0 (Medcalc Software, Mariakerke, Belgium) and SPSS Version 11.5 (SPSS Inc, Chicago, IL, USA).

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), 0.26 mm² (HRT Explorer), and 0.23 mm² (HRT-II Explorer). These values were reduced to 0.16 mm² (HRT Classic), 0.16 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 (fig 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.²¹ This produces a colour coded probability map, indicating locations in the image where the change in

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

HRT type	Strategy	Rim area repeatability coefficient (mm ²)			
		I	II	III	IV
HRT Classic	Standard reference plane	0.21	0.29	0.19	0.24
	320 μm reference plane	0.10	0.14	0.16	0.16
HRT Explorer	320 μm reference plane; MPHSD <25	0.04	0.09	0.07	0.09
	Standard reference plane	0.28	0.31	0.30	0.26
	320 μm reference plane	0.13	0.16	0.16	0.16
HRT II Explorer	320 μm reference plane; MPHSD <25	0.07	0.09	0.09	0.11
	Standard reference plane	0.28	0.34	0.24	0.23
	320 μm reference plane	0.17	0.17	0.18	0.20
	320 μm reference plane; MPHSD <25	0.08	0.07	0.09	0.10

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

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, particularly in tilted optic discs.²⁶

It has been reported in our companion paper that there is variability in inter-test reference height and, therefore, in the height of the contour line at the temporal ONH margin.¹² This indicates that the Standard reference plane may change location between successive images, thereby causing variation in RA measurements. The current findings show that

using the 320 μm reference plane results in a reduction in test-retest RA variability. The advantage of the 320 μm reference plane is that it is stable in relation to changes in the disc margin contour line height. This helps to explain the improved repeatability in the test-retest data and in the longitudinal image series.

The 320 μm reference plane does, however, have weaknesses such as an underestimation of RA in ONHs with oblique insertion and an overestimation of RA in ONHs with advanced parapapillary damage, where the height of the reference ring is affected.²⁴ These factors are important for differentiating the stages of glaucoma.²⁷

A newer, experimental reference plane has been described by Tan *et al.*²⁸ It is intended to have a constant height relation with the ONH in an image series. The position of the reference plane is calculated from a combination of the mean height of the contour line in relation to the reference ring, the contour line's lowest region and the point below the lowest region with the least RA variability. This reference plane has demonstrated better reproducibility of HRT images compared with the Standard and 320 μm reference planes, and has been used in a novel approach to measure glaucomatous change.^{28–29} 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/intervisit 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 2 Interobserver/intervisit rim area repeatability coefficients (mm²) obtained with different image qualities and using different types of HRT and the 320 μm reference plane

Image quality	HRT Classic		HRT Explorer		HRT-II Explorer	
	No of subjects	RC	No of subjects	RC	No of subjects	RC
Good quality image (MPHSD <21)	28	0.05	28	0.08	38	0.07
Medium quality image (MPHSD 21–35)	27	0.16	27	0.13	23	0.09
Poor quality image (MPHSD >35)	19	0.20	19	0.19	13	0.27

MPHSD, mean pixel height standard deviation.

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

Strategy used for longitudinal image series	Standard deviation of residuals (mm ²)		
	Mean	Median	Range
Standard reference plane; non-aligned	0.10	0.10	0.01–0.40
Standard reference plane; aligned	0.09	0.08	0.01–0.40
320 µm reference plane; non-aligned	0.05	0.04	0.01–0.10
320 µm reference plane; aligned	0.04	0.03	0.01–0.10

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.

Authors' affiliations

N G Strouthidis, E T White, T A Ho, D F Garway-Heath, Glaucoma Research Unit, Moorfields Eye Hospital, London, UK

N G Strouthidis, D F Garway-Heath, Institute of Ophthalmology, London, UK

V M F Owen, School of Biomedical and Natural Sciences, Nottingham Trent University, Nottingham, UK

Competing interests: NGS was funded by a Friends of Moorfields research fellowship and through an unrestricted grant from Heidelberg Engineering.

Correspondence to: D F Garway-Heath, MD, FRCOphth, Glaucoma Research Unit, Moorfields Eye Hospital, 162 City Road, London EC1V 2PD, UK; david.garway-heath@moorfields.nhs.uk

Accepted for publication 25 June 2005

REFERENCES

- Motolko M, Drance SM. Features of the optic disc in preglaucomatous eyes. *Arch Ophthalmol* 1981;**99**:1992–5.
- Pederson JE, Anderson DR. The mode of progressive disc cupping in ocular hypertension and glaucoma. *Arch Ophthalmol* 1980;**98**:490–5.
- Sommer A, Pollack I, Maumenee AE. Optic disc parameters and onset of glaucomatous field loss. I. Methods and progressive changes in disc morphology. *Arch Ophthalmol* 1979;**97**:1444–8.
- Sommer A, Katz J, Quigley HA, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol* 1991;**109**:77–83.
- Quigley HA, Addicks EM, Green WR. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Arch Ophthalmol* 1982;**100**:135–46.
- Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg retina tomograph. *J Glaucoma* 1993;**2**:101–3.
- Miglior S, Albe E, Guareschi M, et al. Intraobserver and interobserver reproducibility in the evaluation of optic disc stereometric parameters by Heidelberg Retina Tomograph. *Ophthalmology* 2002;**109**:1072–7.
- Tan JC, Garway-Heath DF, Hitchings RA. Variability across the optic nerve head in scanning laser tomography. *Br J Ophthalmol* 2003;**87**:557–9.
- Kamal DS, Viswanathan AC, Garway-Heath DF, et al. Detection of optic disc change with the Heidelberg retina tomograph before confirmed visual field change in ocular hypertensives converting to early glaucoma. *Br J Ophthalmol* 1999;**83**:290–4.
- Sihota R, Gulati V, Agarwal HC, et al. Variables affecting test-retest variability of Heidelberg Retina Tomograph II stereometric parameters. *J Glaucoma* 2002;**11**:321–8.
- Heidelberg Engineering. *Operation manual for the Heidelberg retina tomograph II*. Vol 1.7.0. 2003.
- Strouthidis N G, White E T, Owen V M F, et al. Factors affecting the test-retest variability of Heidelberg retina tomograph and Heidelberg retina tomograph II measurements. *Br J Ophthalmol* 2005;**89**:1427–32.
- Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology* 1998;**105**:1557–63.
- Kiriya N, Ando A, Fukui C, et al. A comparison of optic disc topographic parameters in patients with primary open angle glaucoma, normal tension glaucoma, and ocular hypertension. *Graefes Arch Clin Exp Ophthalmol* 2003;**41**:541–5.
- Zangwill LM, van Horn S, de Souza Lima M, et al. Optic nerve head topography in ocular hypertensive eyes using confocal scanning laser ophthalmoscopy. *Am J Ophthalmol* 1996;**122**:520–5.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
- British Standards Institution. *Precision of test methods I: Guide for the determination and reproducibility for a standard test method*. London: BSI, 1979.
- Kamal D, Garway-Heath D, Ruben S, et al. Results of the betaxolol versus placebo treatment trial in ocular hypertension. *Graefes Arch Clin Exp Ophthalmol* 2003;**41**:196–203.
- Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology* 1994;**101**:1445–55.
- Altman DG. Statistics and ethics in medical research. VIP—Presentation of results. *BMJ* 1980;**281**:1542–4.
- Chauhan BC, Drance SM, LeBlanc RP, et al. Technique for determining glaucomatous visual field progression by using animation graphics. *Am J Ophthalmol* 1994;**118**:485–91.
- Chauhan BC, McCormick TA, Nicoletta MT, et al. Optic disc and visual field changes in a prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography with conventional perimetry and optic disc photography. *Arch Ophthalmol* 2001;**119**:1492–9.
- Patterson AJ, Garway-Heath DF, Strouthidis NG, et al. A new statistical approach for quantifying change in series of retinal and optic nerve head topography images. *Invest Ophthalmol Vis Sci* 2005;**46**:1659–67.
- Burk RO, Vihanninjoki K, Bartke T, et al. Development of the standard reference plane for the Heidelberg retina tomograph. *Graefes Arch Clin Exp Ophthalmol* 2000;**38**:375–84.
- Chen E, Gedda, Landau I. Thinning of the papillomacular bundle in the glaucomatous eye and its influence on the reference plane of the Heidelberg retina tomography. *J Glaucoma* 2001;**10**:386–9.
- Park KH, Caprioli J. Development of a novel reference plane for the Heidelberg retina tomograph with optical coherence tomography measurements. *J Glaucoma* 2002;**11**:385–91.
- Tuulonen A, Vihanninjoki K, Airaksinen PJ, et al. The effect of reference levels on neuroretinal rim area and rim volume measurements in the Heidelberg Retina Tomograph [abstract]. *Invest Ophthalmol Vis Sci (Suppl)* 1994;**35**:1729.
- Tan JC, Hitchings RA. Reference plane definition and reproducibility in optic nerve head images. *Invest Ophthalmol Vis Sci* 2003;**44**:1132–7.
- Tan JC, Hitchings RA. Approach for identifying glaucomatous optic nerve progression by scanning laser tomography. *Invest Ophthalmol Vis Sci* 2003;**44**:2621–6.