Inter- and intraobserver variation in the analysis of optic disc images: comparison of the Heidelberg retina tomograph and computer assisted planimetry

D F Garway-Heath, D Poinoosawmy, G Wollstein, A Viswanathan, D Kamal, L Fontana, R A Hitchings

Abstract

Aims—The development of imaging and measurement techniques has brought the prospect of greater objectivity in the measurement of optic disc features, and therefore better agreement between observers. The purpose of this study was to quantify and compare the variation between observers using two measurement devices.

Methods—Optic disc photographs and images from the Heidelberg retina tomograph (HRT) of 30 eyes of 30 subjects were presented to six observers for analysis, and to one observer on five separate occasions. Agreement between observers was studied by comparing the analysis of each observer with the median result of the other five, and expressed as the mean difference and standard deviation of differences between the observer and the median. Inter- and intraobserver variation was calculated as a coefficient of variation (mean SD/mean × 100).

Results—For planimetry, agreement between observers was dependent on observer experience, for the HRT it was independent. Agreement between observers (SD of differences as a percentage of the median) for optic disc area was 4.0% to 7.2% (planimetry) and 3.3% to 6.0% (HRT), for neuroretinal rim area it was 10.8% to 21.0% (planimetry) and 5.2% to 9.6% (HRT). The mean interobserver coefficient of variation for optic disc area was 8.1% (planimetry) and 4.4% (HRT), for neuroretinal rim area it was 16.3% (planimetry) and 8.1% (HRT), and (HRT only) for rim volume was 16.3%, and reference height 9.1%. HRT variability was greater for the software version 1.11 reference plane than for version 1.10. The intraobserver coefficient of variation for optic disc area was 1.5% (planimetry) and 2.4% (HRT), for neuroretinal rim area it was 4.0% (planimetry) and 4.5% (HRT).

Conclusions—Variation between observers is greatly reduced by the HRT when compared with planimetry. However, levels of variation, which may be clinically significant, remain for variables that depend on the subjective drawing of the disc margin.

Material and methods

Optic disc photographs and images from a scanning laser ophthalmoscope (SLO), from...
Inter- and intraobserver variation in the analysis of optic disc images

665

One eye of 30 subjects, were presented to six different observers for analysis. In addition, one of the observers (observer 1) analysed the images five times, on separate occasions at least a week apart. Photographs and SLO images were selected on the basis of image quality rather than on the presence of any particular optic disc features.

Optic Disc Photography

Subjects’ pupils were dilated with 1% tropicamide. Photographs of the optic disc were taken with the Canon CF60U at the 30 degree setting. Four sequential photographs of each eye were taken, with a lateral shift in camera position after two pictures to obtain a stereo effect when the images are viewed stereoscopically. Clear images with a fair to good stereoscopic impression were selected for the study.

Planimetry

Photographs were analysed by computer-assisted planimetry using the DISC-DATA, Thot Informatique (Pr Bechettoille, Angers, France) program. For each subject three photographs were supplied—two forming a stereoscopic pair for reference, and a third for analysis. The planimetry system is calibrated at the start of each session by placing a grid with marks of known spacing beneath the video camera of the system. The grid is then replaced by the optic disc photograph. The observer views the image on a computer screen and outlines the optic disc and cup margins using a computer mouse while referring to the stereo pair to identify contour changes. Each observer was instructed to define the optic disc anatomy according to the following conventions: the area of the disc is the area within Elschnig's ring, the cup is defined on the basis of contour, not pallor, and the neuroretinal rim/optic cup border is taken as the level at which the slope of the rim steepens. Vessels were considered to be part of the cup if there was no underlying rim tissue. Change in direction of vessels in the optic disc was used as a guide to the neuroretinal rim edge in those photographs where the stereo impression was not good. No training in standardisation was given.

The output of the DISC-DATA program includes optic disc, cup and neuroretinal rim areas, and cup/disc ratio.

Imaging with the Scanning Laser Ophthalmoscope

All the subjects were imaged using the Heidelberg retina tomograph (HRT) (software version 1.11) in the 10 × 10 degree frame. All images were obtained by one of two trained technicians. Imaging was performed at the 1.5 cm imaging head/eye distance recommended in the instruction manual as the subject viewed a distant fixation target. Each patient had three high quality scan series recorded at one sitting. The quality of images was assessed with the aid of the HRT software, and by the experience of the technician. The HRT software is able to correct for small eye movements by aligning consecutive images within a scan series. Scan series with movements occurring within a single image in the series, which caused distortion of the image that could not be corrected, were excluded from the study. The mean topography of the three scan series was used for the analysis.

Each observer was instructed to outline the margin of the optic disc (as defined above). Two choices to make the outline were given: (1) to make the outline directly on the mean topography image, or (2) to make the outline on one of the “optical sections” from an aligned series and to “export” the outline to the mean topography image. The second option allows better identification of Elschnig’s ring in some eyes. The optic disc photographs were not available to aid the observer for this part of the study. The images were then analysed by the HRT software to generate the morphological variables using both the standard reference plane (50 µm below the temporal disc margin at the papillomacular bundle, software version 1.11) and a reference plane placed 320 µm below the mean retinal height at the reference ring (as in software version 1.10). After each observer’s analysis, the outline was deleted and replaced with a randomly placed circle, so that the next observer was not influenced by the previous observer’s opinion.

Subjects

Photographs and HRT images from 15 normal subjects (mean age 56.0 (SD 12.8) years) and 15 patients with early glaucomatous field defects (mean age 58.9 (8.8) years) were taken. Subjects had been recruited prospectively as part of a study on the early detection of glaucoma (approved by the hospital advisory research committee). All subjects gave informed consent to the investigations performed.

Normal subjects

Restriction criteria included ametropia <6 dioptres, visual acuity of 6/9 or better, normal visual fields, and intraocular pressure of <21 mm Hg.

Glaucoma patients

Restriction criteria included ametropia <6 dioptres, visual acuity of 6/9 or better, a visual field defect (scoring 1–5 as in the Advanced Glaucoma Intervention Study protocol) reproduced on at least three successive occasions, and intraocular pressure >21 mm Hg at diagnosis.

Observers

Five clinical or research glaucoma fellows and one experienced glaucoma technician took part. Observers 1 and 2 had considerable experience with planimetry, observers 3 to 6 had no previous experience. Observer 1 had considerable experience with the HRT, observers 2 and 3 had moderate experience, observers 4 to 6 had little or no previous experience.

Statistical Analysis

The expression “interobserver variation” is used to describe the variation among observers.
Table 1  Summary of morphological variables. The figures given are the means of the six observations for each subject. The mean for each variable is the mean for the 30 subjects.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Planimetry Mean (range)</th>
<th>HRT (standard reference plane) Mean (range)</th>
<th>HRT (&quot;320 µm&quot; reference plane) Mean (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc area (mm²)</td>
<td>2.18 (1.67–3.92)</td>
<td>2.05 (1.59–2.78)</td>
<td>2.05 (1.59–2.78)</td>
</tr>
<tr>
<td>Rim area (mm²)</td>
<td>1.29 (0.59–1.83)</td>
<td>1.24 (0.32–2.28)</td>
<td>1.29 (0.63–2.32)</td>
</tr>
<tr>
<td>Rim volume (mm³)</td>
<td>—</td>
<td>0.282 (0.25–0.704)</td>
<td>0.327 (0.100–1.603)</td>
</tr>
<tr>
<td>Cup shape measure</td>
<td>—</td>
<td>−0.135 (−0.322–0.102)</td>
<td>−0.135 (−0.322–0.102)</td>
</tr>
<tr>
<td>Height variation contour</td>
<td>—</td>
<td>0.343 (0.145–0.637)</td>
<td>0.343 (0.145–0.637)</td>
</tr>
<tr>
<td>Reference height (mm)</td>
<td>—</td>
<td>0.287 (0.14–0.452)</td>
<td>0.320</td>
</tr>
</tbody>
</table>

for measuring the various disc variables. Variation is expressed as a coefficient of variation. This is calculated for each subject as the standard deviation of the six observers’ estimation divided by the mean of the six observers’ estimation multiplied by 100 (to give a percentage). The mean value for the 30 subjects was used as a summary measure for each variable.

The expression “intraobserver variation” is used to describe the variation of one observer for repeated estimations. Variation is expressed as a coefficient of variation. This is calculated for each subject as the standard deviation of the observer’s five estimations divided by the mean of the five estimations multiplied by 100 (to give a percentage). The mean value for the 30 subjects was used as a summary measure for each variable.

The expression “agreement” is used to describe the difference between each observer and the other five. This is calculated for each observer by comparing the analysis of the observer with the median result of the other five observers’ analyses for the 30 subjects. Agreement is expressed as the mean difference and standard deviation of differences (SD differences). The SD differences was expressed as a percentage of the mean value for each variable and the (magnitude of differences was unrelated to the size of the variable measured in all cases). The mean difference represents the bias of one observer compared with the others. The SD differences represents the random error of an observer.

Observer agreement was compared in the two diagnostic categories (normal and glaucomatous) by means of Student’s t test.

Comparisons of the performance (differences from median) between each observer were made by a one way ANOVA, using the Bonferroni correction for making multiple comparisons.

**Results**

The mean value and range of values for the morphological variables measured by planimetry and the HRT are summarised in Table 1.

**INTEROBSERVER VARIATION**

The mean interobserver variation for each variable and each measurement technique is summarised in Table 2.

For the HRT, the differences in the coefficient of variation between the two methods of reference plane definition were significant for rim area (paired t test, p <0.03) and rim volume (paired t test, p = 0.01). For the variable “cup shape measure” (CSM), the mean interobserver variation (mean SD of measures), for all observers, represents 2.8% of the range of values in the study. There was no difference between the normal and glaucoma subjects in the magnitude of differences found between the observers.

For the variable “reference height” (version 1.11), the mean interobserver coefficient of variation was 7.2%. There was no difference between the normal and glaucoma subjects in the magnitude of differences found between the observers.

**INTRAOBSERVER VARIATION**

The intraobserver variation for each variable and each measurement technique is summarised in Table 3. The variability in estimating the disc area and rim area was approximately the same for planimetry and the HRT.

**OBSERVER AGREEMENT**

The performance of each observer (agreement with the median) is shown graphically for measurement of optic disc area (Fig 1) and neuroretinal rim area (Fig 2). The central line in the box represents the bias of the observer compared to the other five. The size of the box equates with the magnitude of the random error. Observer 2 made a calibration error at the start of the planimetry (optic disc, cup and rim area measurements are all larger than those made by other observers). That this was a calibration error was confirmed by comparing the cup/disc ratio with other observers.

For planimetry, agreement (SD differences) with the median for disc area was 4.0% (best observer (observer 1)) to 7.2% (worst observer (observer 2)). For rim area it was 10.8% (best observer (observer 1)) to 21.0% (worst observer (observer 6)).
The magnitude of differences between observers and the median estimate, taking all observers, was significantly greater for defining the optic cup in the glaucoma group (mean difference 0.21 mm²) than in the normal group (mean difference 0.16 mm²), p = 0.04. No significant differences were found for optic disc area or rim area.

For the HRT, agreement (SD differences) with the median for disc area was 3.3% (best observer (observer 1)) to 6.0% (worst observer (observer 3)). For rim area it was 5.5% (best observer (observer 1)) to 9.7% (worst observer (observer 3)) using the standard reference plane and 5.2% (best observer (observer 1)) to 9.6% (worst observer (observer 3)) using the “320 µm” reference plane.

The mean magnitude of differences between an observer and the median, taking all observers, was significantly greater for defining the optic disc in the normal group (mean difference 0.10 mm²) than in the glaucoma group (mean difference 0.06 mm²), p = 0.003. Similarly, the magnitude of differences was significantly greater for defining the neuroretinal rim in the normal group (mean difference 0.09 mm²) than in the glaucoma group (mean difference 0.06 mm²), p = 0.014.

The estimation of optic disc area by the more experienced observers (1, 2, and 3) was significantly greater than that of the less experienced observers (one way ANOVA): observer 1 greater than observers 3, 5, and 6, p=0.000 to 0.004, observer 2 greater than observer 6, p=0.000, and observer 3 greater than observer 6, p=0.018.

**Discussion**

Previous studies have demonstrated that the repeatability of the determination of the topography of the optic nerve head and surrounding retina is very high with scanning laser ophthalmoscopes.18 19 21 22 35–38 Little attention, however, has yet been given to the question of the variability that arises in variable definition when different observers analyse the same images. This study was designed to address this question and compare the results with the established method of planimetry.

The results will be discussed for variables that have been shown to be useful to distinguish glaucomatous from normal optic discs39–41: optic disc, optic cup and neuroretinal rim areas, CSM, and rim volume.

**OPTIC DISC AREA**

The observer variability in estimating the disc area was approximately the same for the HRT and planimetry (Fig 1), for all observers except observer 2. The mean interobserver variation was consequently higher for planimetry than for the HRT (Table 2) as this includes observer 2.

In the estimation of disc edge in the HRT images, the magnitude of differences between the observers was greater for normal than for glaucomatous optic discs. This is likely to result from difficulty in identifying Elschnig’s ring in parts of the optic disc where the nerve fibre layer is thickest, at the poles and nasal part of the disc. In glaucoma, as the nerve fibre layer thins, Elschnig’s ring becomes progressively more visible. This may also account for the larger disc size estimations by the more experienced observers, as they estimated the probable position of the disc edge by experience when Elschnig’s ring is not visible.

The interobserver coefficient of variation is slightly higher, and the agreement similar, to that previously reported for planimetry42 (3.0% to 4.2%) and for the ONHA (6.2%).27

Intraobserver variation was very similar for planimetry and the HRT (coefficient of variation 1.5% and 2.4% respectively), and similar to that previously reported for planimetry (coefficient of variation 1% to 2.3%).18 32 42 The Humphrey retinal analyser (mean coefficient of variation 0.8%)30 and the ONHA (SD differences of 0.03 to 0.09 mm²).27

![Figure 1](http://bjo.bmj.com/first-published-as/10.1136/bjo.83.6.664/1-Jun-1998)

**Figure 1** Optic disc area. Box plot of differences between each observer’s analysis and the median analysis of the other five observers for the HRT (dark grey) and planimetry (light grey). The box represents the interquartile range (central 50% of values) and the heavy horizontal line is the median value. The whiskers extend to observations that are less than one interquartile range from the end of the box. The circles represent “outliers” (cases with values more than three box lengths from the upper or lower edge of the box). The asterisks represent “extremes” (cases with values between 1.5 and 3 box lengths from the upper or lower edge of the box).

![Figure 2](http://bjo.bmj.com/first-published-as/10.1136/bjo.83.6.664/1-Jun-1998)

**Figure 2** Neuroretinal rim area. See Figure 1 for details.
OPTIC CUP AREA
Planimetry requires the subjective definition of the “edge” between the neuroretinal rim and optic cup, which is itself an arbitrary concept. Overall, for planimetry, agreement between observers was better when assessing the cup size in the normal subjects than in the glaucoma patients. This may result from a loss of features, such as small vessel detail, in the neuroretinal rim of glaucoma patients, which makes edge definition more difficult.

In the HRT analysis the neuroretinal rim/optic cup border is fixed with reference to a predetermined “plane”. This plane is determined by the height of the retina, either at the papillomacular bundle at the defined optic disc edge (software version 1.11) or at the more peripheral “reference ring” (software version 1.10). The principal difference between these, in the context of this study, is that the height of the version 1.11 reference plane may vary according to the position of the observer defined optic disc margin, especially if there are large surface contour changes in this region. The version 1.10 reference plane is independent of the observer defined optic disc margin. The interobserver variation for the variable “reference height” (version 1.11) was quite large (coefficient of variation 7.2%). This variation in reference height is reflected by variability in cup area determination and, consequently, rim area determination. There is, by definition, no variability between observers in the version 1.10 reference height and, therefore, no variability in cup determination.

NEURORETINAL RIM AREA
The variability in rim estimates results from the combined effects of optic disc edge and cup edge estimates, and is much greater for planimetry than the HRT, principally as a result of the high variability in cup estimations. The agreement of the most experienced observer using planimetry was approximately the same as the least experienced with the HRT. With the HRT, the differences between the observers and the median are normally distributed, so that the 95% levels of agreement will be about plus or minus 15.8% from the mean.

The mean interobserver coefficient of variation with the HRT, at 8.2%, is similar to that previously reported for planimetry (7.8%),42 the ONHA 7.7%25 and slightly better than that for the IS 20009 (mean 11.6%).

Intraobserver variation was very similar for planimetry and the HRT (coefficient of variation 4.0% and 4.5% respectively), and similar to that previously reported for planimetry (coefficient of variation 2.9% to 4.3%).14 42 The Humphrey retinal analyser (mean coefficient of variation 2.2%),49 digital photogrammetry (mean coefficient of variation 5.3%)20 and the ONHA (coefficient of variation 6.1 to 6.7% and SD differences of 0.02 to 0.05 mm)14 21 27

CUP SHAPE MEASURE
This is a measure of the steepness and depth of cupping in an optic disc. Its value is unaffected by the height of the reference plane, but will be affected by the shape of the surface enclosed by the observer defined disc margin. Variability between observers was low for this measure, with the mean SD differences between each observer and the median representing about 2.8% of the range of values.

RIM VOLUME
The variation between observers was higher for this variable than any other assessed in this study. It was also particularly sensitive to differences in reference height, demonstrated by the difference between the coefficient of variation using the version 1.10 and 1.11 reference planes (8.7% and 13.6% respectively). A coefficient of variation of 9.75% using version 1.11 has been reported previously.43

The intraobserver variation determined in this study compares well with previously published data for planimetry,14 42 and demonstrates that it is possible to obtain the same level of consistency for an experienced observer using planimetry as with the (more objective) HRT. Intraobserver variation has less importance with analysis by the HRT, because the HRT software has the facility to “export” the disc margin definition from one image in a series to subsequent images. This may reduce variability in longitudinal analyses below that of the intraobserver variation found in this study. A recent study has investigated the variability in a number of variables that arises when the “export” option is used.44 The coefficient of variation for disc area, cup area, and rim volume was 0.05%, 2.2%, and 2.4% respectively. This compares with intraobserver variation of 2.4%, 3.8%, and 8.4% for the same variables in this study.

The analysis of HRT images is much less dependent on the experience of the observer than it is in planimetry and agreement was better with the HRT analysis for all variables, except for optic disc area. However, even with the improved agreement found with HRT, the variation in disc margin definition, together with the subsequent variation in reference height and cup definition, leads to a variation in rim area estimation which may be clinically significant in cross sectional (diagnostic) studies.

Mr Garway-Heath was supported in this research by a grant from the Guide Dogs for the Blind Association. The authors would like to thank Catey Bunce (medical statistician, Glaxo Department of Ophthalmic Epidemiology, Moorfields Eye Hospital, and Institute of Ophthalmology) for advice in the statistical analysis of the data in the study.

Inter- and intraobserver variation in the analysis of optic disc images


