The Agreement between the Heidelberg Retina Tomograph and a Digital Nonmydriatic Retinal Camera in Assessing Area Cup-to-Disc Ratio

Ecosse L. Lamoureux, Karyan Lo, John G. Ferraro, Mario Constantinou, Jill E. Keeffe, Andreas Müller, and Hugh R. Taylor

PURPOSE. To determine the repeatability and agreement between a digital camera (monoscopic and stereoscopic images) and the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany) in determining cup-to-disc ratio. A secondary purpose was to determine the monoscopic and stereoscopic thresholds that maximize specificity and sensitivity when compared with the HRT.

METHODS. Community living participants aged between 70 and 79 years had their optic discs imaged with a digital nonmydriatic retinal camera (NMRC) and the HRT. Intraclass correlation coefficients (ICCs) and 95% tolerance limits of change were used to determine repeatability characteristics of the instruments. The agreement between the HRT- and NMRC-derived area cup-to-disc ratios was assessed using weighted κ statistics and receiver operator characteristic (ROC) curves.

RESULTS. The agreements between the monoscopic and stereoscopic images and HRT were assessed for 1238 and 1173 eyes, respectively. The reliability measures for both NMRC techniques and HRT were almost perfect (ICC = 0.84 – 0.99) with narrow tolerance limits of change (9.2%–18.4%) and very small systemic biases (P < 0.05). The agreement between the HRT- and NMRC techniques was substantial, with a weighted κ = 0.83. The HRT gave a marginally larger area cup-to-disc ratio than the monoscopic and stereoscopic images by 0.008 and 0.006, respectively (P < 0.001). The areas under the ROC curves for both NMRC techniques were 0.98, indicating excellent discriminating characteristics (P < 0.001). An area cup-to-disc ratio cutoff of ≥0.5 for the monoscopic and stereoscopic NMRC was highly specific (94.1% and 91.6%) and sensitive (87.5% and 97.2%, respectively) in determining an HRT-derived area cup-to-disc ratio >0.6.

CONCLUSIONS. The monoscopic and stereoscopic digital images showed excellent repeatability and demonstrated substantial agreement with the HRT. The results indicate that the digital NMRC could be a reliable and useful instrument for assessing area cup-to-disc ratio and screening for glaucoma-suspect eyes in the community. (Invest Ophtalmol Vis Sc 2006;47:93–98) DOI:10.1167/iovs.05-0936

Glaucoma is ranked the third most common eye disease in the world and has been predicted to become the most common cause of blindness in the early years of this millennium. One of the major challenges in the management of glaucoma is the timely detection of the disease. In Australia alone, it is estimated that there are approximately 200,000 people with glaucoma. However, half of all cases are not diagnosed and consequently are not treated. There is therefore a need to develop and promote effective programs to detect glaucoma that may include community screening.

The diagnosis of glaucoma is usually based on a combination of variables. In large population-based investigations, an enlarged vertical cup-to-disc ratio has been used as one of the critical parameters in determining glaucoma prevalence rates. Over the past decade, the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany), a confocal scanning laser ophthalmoscope, has been commonly used to assess cup-to-disc ratio in patients with glaucoma. The assessment of area cup-to-disc ratio by the HRT is greatly reproducible with a high sensitivity and specificity in the diagnosis of glaucoma when its results are combined with perimetry and family history. However, the HRT is costly and is not readily available for use in a community screening setting. An ideal community-based screening instrument should not only be accurate, but also convenient and cost effective. A digital nonmydriatic retinal camera (NMRC; CR6-45NM; Canon, Tochigiken, Japan) can be used to help assess the rates of the most common causes of visual impairment, including cataract and interobserver repeatability and the agreement in cup-to-disc ratio assessed by a digital optic disc camera and the HRT. Sung et al undertook a comparable investigation but used pharmacological dilation on a relatively small sample (n = 39). However, in a community-screening environment pharmacologic mydriasis increases examination time, may induce angle closure glaucoma, and produces discomfort in some people. These can affect the compliance rate and act as a barrier to screening. Also, glaucoma-suspect eyes are often characterized by a symmetrically enlarged cup-to-disc ratio higher than 0.6. However, to date, there is little information about the accuracy and agreement between the assessment of cup-to-disc ratios of digital images and the HRT.

The primary goal of this study was to determine the intra- and interobserver repeatability and the agreement in cup-to-disc ratio between a digital NMRC and the HRT in community-living older adults. A secondary purpose was to determine corresponding thresholds for monoscopic and stereoscopic values, which maximize specificity and sensitivity when compared with the HRT-derived cup-to-disc ratio >0.6.
METHODS

This study was undertaken as part of a survey of eye disease and eye care utilization in Victoria, Australia. The methods of recruitment and assessment procedures are reported in detail elsewhere. Briefly, all participants completed a self-administered questionnaire that included demographic details and relevant medical and family histories. Information pertinent to attitudes and knowledge of eye health was also obtained. The participants undertook an assessment that included presenting and best corrected visual acuities, automated visual field testing by frequency-doubling technology (FDT), lens and retinal assessment with an NMRC, and optic disc imaging with the HRT. Written informed consent was obtained from all participants at the start of the study, in accordance with the Declaration of Helsinki. The Department of Human Services, the Cancer Council and Human Research Ethics Committee of the Royal Victorian Eye and Ear Hospital granted ethics approval before the study.

Participants

All participants were aged between 70 and 79 years and resided in randomly selected metropolitan suburbs stratified by socioeconomic status. People who satisfied the inclusion criteria were identified by the Victoria Department of Human Services and were sent a letter inviting them to participate. Those accepting participation provided their contact details and were phoned to arrange an appointment. An information pack consisting of appointment details, a plain-language statement, a consent form, and a self-administered questionnaire were then sent out to each participant.

Protocol for Image Acquisition

A CR6-45NM (Canon) NMRC (with a 45° angle of view) was used with a digital camera (Canon EOS-D30; Canon) capable of taking images with a resolution of 2160 × 1440 pixels (large) and 1440 × 960 pixels (small) in true color. This gave an image size of 3.2 megapixels. Pharmacologic dilating agents were not used for the study, although participants spent a few minutes in a darkened area for their pupils to undergo physiological mydriasis before retinal images were taken. All images were taken in a darkened room. Four retinal images (two from each eye) were taken from each participant, as required for the stereoscopic analysis of the optic disc. An experienced technician obtained all images.

Calculation of Cup-to-Disc Ratio

Although the vertical cup-to-disc ratio is used clinically, in this study we measured the agreement between the HRT and NMRC by using the area cup-to-disc ratio. We chose the cup-to-disc ratio because it is a common variable that is provided by the HRT and NMRC software packages. After examination, the clearer of the two images of each eye was used for the monoscopic assessment of the optic disc. The disc and cup margins were traced with a computer mouse and the software computed the cup-to-disc ratio (Digital Healthcare Image Management Systems, Version 2.6). The disc was defined as the area inside the peripapillary scleral ring. The cup was determined on the basis of pallor and the changes in the course of the blood vessels. To induce stereopsis, the software manipulated the retinal images by converting one into red color and the other into green and then superimposing them. The superimposed images were viewed with red-green glasses. As with the monoscopic assessment, the disc was defined as the area inside the peripapillary scleral ring. However, for the stereoscopic assessment, the cup was defined on the basis of contour rather than pallor. The grader who demarcated the disc and cup margins was masked to the results of the monoscopic assessment.

The HRT images were assessed with the HRT software (ver. 1.4.1.0; Heidelberg Engineering), and one experienced technician graded all the images. Three optic nerve head (ONH) images were obtained from each eye. The grader drew the optic disc margin on the HRT topographic images while being masked from the results obtained from the retinal images analysis. The optic disc was defined as the area inside the peripapillary scleral ring, and the HRT software determined the cup and the cup-to-disc ratio automatically.

Statistical Analysis

Ten percent of the NMRC and HRT images were randomly selected and evaluated twice (±4 week interval), to assess the intraobserver repeatability of the two instruments. The same two graders, masked to the results of the other, assessed another 10% of randomly selected images to determine interobserver repeatability for both instruments. The intraclass correlation coefficient (ICC) was used to determine the reliability of the instruments. Weighted κ statistics (κw) were used to determine the agreement between the HRT and NMRC (Intercooled STATA, ver. 8.2; StataCorp, College Station, TX). The 95% tolerance limits for change were used to analyze measurement variability. The term change, as used herein, implies the difference between the two observations, not a change over time. The 95% tolerance limits for change were computed by multiplying the standard deviation of the difference between repeated measurements by 1.96. Dividing this result by the range of all measurements expressed the tolerance for change as a percentage of the data range, allowing for more meaningful comparisons between different measurement variables.

The receiver operating characteristic (ROC) curve procedure was selected to provide an evaluation of the performance of the NMRC (monoscopic and stereoscopic) against the HRT (grouped as area cup-to-disc ratios ≤0.6 and >0.6). The ROC curves plot true-positive rates against false-positive rates (sensitivity versus 1 − specificity). An ideal diagnostic instrument is one that has a sensitivity and specificity of 100% and approaches the top left-hand corner of the graph (see Fig. 3). An instrument with no prognostic value approximates a 45° slope from the origin. The farther the curve lies above the reference line, the more accurate the test. An alternate way of expressing the utility of the NMRC is to determine the area under the ROC curve. An ideal instrument would have an area under the curve of 100% and an instrument with no diagnostic utility would have an area under the curve of 50%. Paired t-tests were used to test for significant differences between measurements with a P < 0.05 considered significant. All statistical analyses, other than κ statistics, were performed on computer (Statistical Package for the Social Sciences for Windows, ver. 12.0; SPSS Inc, Chicago, IL).

RESULTS

Six hundred ninety-seven participants (58% males; mean age 74.7 years) took part in the study. Fifty-two (7.5%) participants reported a previous diagnosis of glaucoma and 67 (9.6%) had a positive family history of glaucoma. Six hundred and forty-five participants (92.5%) had presenting VA 6/12 or better, and 14 (2%) had bilateral best-corrected VA <6/12. Almost 91% of the participants obtained a normal screening test result by the range of all measurements expressed the tolerance for change as a percentage of the data range, allowing for more meaningful comparisons between different measurement variables. The receiver operating characteristic (ROC) curve procedure was selected to provide an evaluation of the performance of the NMRC (monoscopic and stereoscopic) against the HRT (grouped as area cup-to-disc ratios ≤0.6 and >0.6). The ROC curves plot true-positive rates against false-positive rates (sensitivity versus 1 − specificity). An ideal diagnostic instrument is one that has a sensitivity and specificity of 100% and approaches the top left-hand corner of the graph (see Fig. 3). An instrument with no prognostic value approximates a 45° slope from the origin. The farther the curve lies above the reference line, the more accurate the test. An alternate way of expressing the utility of the NMRC is to determine the area under the ROC curve. An ideal instrument would have an area under the curve of 100% and an instrument with no diagnostic utility would have an area under the curve of 50%. Paired t-tests were used to test for significant differences between measurements with a P < 0.05 considered significant. All statistical analyses, other than κ statistics, were performed on computer (Statistical Package for the Social Sciences for Windows, ver. 12.0; SPSS Inc, Chicago, IL).

Intraobserver Repeatability

There was almost perfect intraobserver repeatability on all three techniques in the measurement of area cup-to-disc ratio...
TABLE 1. Demographic and Clinical Characteristics of the 697 Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>382 (55)</td>
</tr>
<tr>
<td>Mean age, y (range)</td>
<td>74.7 (70–79)</td>
</tr>
<tr>
<td>Self-reported glaucoma</td>
<td>52 (7.5)</td>
</tr>
<tr>
<td>Family history of glaucoma</td>
<td>67 (9.6)</td>
</tr>
<tr>
<td>Best presenting visual acuity &gt;6/12</td>
<td>645 (92.5)</td>
</tr>
<tr>
<td>Best corrected visual acuity ≥6/12</td>
<td>683 (98)</td>
</tr>
<tr>
<td>FDT results for C-20-1 screening test (worse eye)</td>
<td></td>
</tr>
<tr>
<td>No point missed</td>
<td>633 (90.8)</td>
</tr>
<tr>
<td>One point missed</td>
<td>22 (3.2)</td>
</tr>
<tr>
<td>Two or more points missed</td>
<td>59 (5.6)</td>
</tr>
<tr>
<td>Unavailable</td>
<td>3 (0.4)</td>
</tr>
</tbody>
</table>

Data are expressed as the number of participants (% of entire sample) unless otherwise specified.

There were no significant differences between the means of repeated assessments (P > 0.05). The NMR tended to have slightly higher tolerance limits for change and ICCs than the HRT. The regression plots of the three techniques showed insignificant slopes for intraobserver agreement ($R^2$ = 0.000, 0.002, and 0.000; P > 0.05 for the monocscopic, stereoscopic, and HRT, respectively).

Interobserver Repeatability

The interobserver repeatability in the HRT was almost perfect (0.97), and stereoscopic and monocscopic interobserver agreements were substantial (Table 2). There were no significant interobserver differences on the mean scores of each technique. Similar to the intraobserver findings, the HRT tended to have smaller tolerance limits of change and higher ICC scores than did the NMRC (Table 2). The regression plots of all three techniques for interobserver repeatability showed insignificant slopes ranging between $R^2$ = 0.010 and 0.019 (P > 0.05).

Agreement between the HRT and NMRC

The agreement between the HRT- and NMRC-derived area cup-to-disc ratios was substantial for both monocscopic and stereoscopic images (Table 3). The tolerance limits of change and percentage of area cup-to-disc ratio difference ≥0.2 between the two instruments were relatively small and very similar. The regression plots of the difference in area cup-to-disc ratios measured by the HRT versus the monoscopic and stereoscopic NMR against the average of the two instruments are shown in Figures 1 and 2, respectively. The slopes of the regression lines although statistically significant for both graphs showed a relatively low $R^2$ ($R^2$ = 0.06 and 0.03, for HRT versus the monoscopic and stereoscopic NMR, respectively). This indicates that the HRT tended to measure a marginally larger cup-to-disc ratio in discs with larger cup sizes and vice versa.

The ROC curve procedure was selected to provide an evaluation of the performance of the NMRC against the HRT grouped into two categories (i.e., area cup-to-disc ratio ≤0.6 and >0.6). Both graphs (monoscopic and stereoscopic) approached the top left-hand corner, suggesting that they discriminate well (Fig. 3). The areas under the curve were 0.98 and 0.97 (monoscopic and stereoscopic, respectively; P < 0.001) indicating that both NMRC techniques have excellent discriminating characteristics. An examination of the thresholds indicates that NMR ≥ 0.5 recorded high specificity (94.1% and 91.6%) and sensitivity (87.5% and 97.2%, monocscopic and stereoscopic, respectively) in determining cup-to-disc ratios >0.6 measured by the HRT.

**DISCUSSION**

The assessment of the cup-to-disc ratio is a critical criterion in determining eyes with suspected glaucoma, as it is generally believed that the progression of glaucomatous visual field defects follows the deterioration in the optic disc morphology.21–26 Traditionally, the stereoscopic mydriatic fundus camera has been used to obtain stereoscopic optic disc photographs, but this technology is cumbersome and time consuming.18 Similarly, the Heidelberg Retina Tomograph is an objective and highly reproducible device but it is also relatively expensive. A reliable, user-friendly, and cost-effective community method of assessing optic disc cupping would be helpful in detecting the early progression in glaucomatous optic nerve head damage.

![Table 2](image)

**Table 2. The Intra- and Interobserver Repeatability Results for Area Cup-to-Disc Ratios Derived from the NMRC and the HRT**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Mean Difference</th>
<th>95% CI</th>
<th>95% Tolerance Limits of Change (%)</th>
<th>ICC*</th>
<th>Cup-to-Disc Ratio Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraobserver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monoscopic</td>
<td>-0.01</td>
<td>-0.02-0.00</td>
<td>12.04</td>
<td>0.98</td>
<td>0.0</td>
</tr>
<tr>
<td>Stereoscopic</td>
<td>0.00</td>
<td>-0.00-0.01</td>
<td>12.32</td>
<td>0.97</td>
<td>0.0</td>
</tr>
<tr>
<td>HRT</td>
<td>0.00</td>
<td>-0.00-0.01</td>
<td>9.24</td>
<td>0.99</td>
<td>0.0</td>
</tr>
<tr>
<td>Interobserver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monoscopic</td>
<td>0.00</td>
<td>-0.00-0.02</td>
<td>18.48</td>
<td>0.84</td>
<td>2.6</td>
</tr>
<tr>
<td>Stereoscopic</td>
<td>0.01</td>
<td>-0.00-0.01</td>
<td>12.32</td>
<td>0.92</td>
<td>1.0</td>
</tr>
<tr>
<td>HRT</td>
<td>0.00</td>
<td>-0.03-0.00</td>
<td>10.36</td>
<td>0.97</td>
<td>2.6</td>
</tr>
</tbody>
</table>

* All are statistically significant (P < 0.001).

**Table 3. The Agreement between HRT- and NMRC-Derived Area Cup-to-Disc Ratios**

<table>
<thead>
<tr>
<th>Instruments</th>
<th>Mean Difference*</th>
<th>95% CI</th>
<th>95% Tolerance Limits of Change (%)</th>
<th>Weighted κ*</th>
<th>θ ≥0.2 Cup-to-Disc Ratio Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRT-monoscopic</td>
<td>0.01</td>
<td>0.00-0.01</td>
<td>15.96</td>
<td>0.83</td>
<td>2.3</td>
</tr>
<tr>
<td>HRT-stereoscopic</td>
<td>0.01*</td>
<td>0.00-0.01</td>
<td>15.68</td>
<td>0.83</td>
<td>1.7</td>
</tr>
</tbody>
</table>

* All are statistically significant (P < 0.001).
The digital nonmydriatic retinal camera had almost perfect intraobserver agreement for monoscopic and stereoscopic area cup-to-disc ratios with both techniques recording narrow tolerance limits of change and very small nonsignificant systemic biases. Similar results were found for the interobserver reliability measures. Our reliability measures for the digital camera compare favorably with similar studies that involved patients with glaucoma in a clinical setting. Shuttleworth et al.26 for example, reported ICCs of 0.955 and 0.921 for intra- and interobserver agreement for area cup-to-disc ratios derived from stereoscopic mydriatic digital images with very similar mean difference values and tolerance limits of change. Our findings confirm previous observations that the intraobserver agreement is higher than interobserver conformity.18,26–28 The HRT also demonstrated excellent reliability measures confirming that scanning laser tomography provides a reproducible measurement of the topography of the optic discs.29,30

We found a strong agreement between monoscopic NMRC and HRT-derived area cup-to-disc ratios. To our knowledge, this is the first study that has undertaken such an investigation and, although a comparison of findings is difficult, future research in this area is needed and should produce promising results. A similar substantial agreement was found between the stereoscopic NMRC and HRT. Identical studies have found only a fair to moderate agreement (mean ICC = 0.49–0.61),18,31 whereas similar investigations have reported a moderate to substantial agreement ($\kappa_w = 0.57–0.72$) for vertical cup-to-disc ratios and fair to moderate ($\kappa_w = 0.21–0.55$) for horizontal cup-to-disc ratios32 when comparing the two instruments. Our agreement between the HRT and stereoscopic NMRC was stronger and may be related to our substantial reliability measurements, and to differences in cameras, reference planes, and statistical methods used in the other studies.
An examination of the regression plots shows that the HRT tended to overestimate marginally the cup-to-disc ratio-derived monoscopic and stereoscopic NMRC with larger optic disc sizes and vice versa. Although other studies have also reported an overestimation of the HRT over the digital camera ranging between 0.07 and 0.11,18,33 our study found substantially smaller systemic biases for the monoscopic and stereoscopic NMRC, respectively. Although these biases were found to be statistically significant, it could be argued they are highly unlikely to be clinically important.

Both NMRC techniques were found to be excellent discriminating instruments with outstanding specificity and sensitivity characteristics in detecting suspect glaucomatous discs. However, the stereoscopic NMRC tended to have a marginally better specificity and sensitivity combination and a smaller overall bias than did the monoscopic. Considering that an enlarged cup-to-disc ratio is one of the critical determinants in diagnosing glaucoma, stereoscopic-derived cup-to-disc ratios appear to be the technique of choice of the two NMRC techniques. In contrast, two images are needed to produce stereopsis and that it takes longer to compute cup-to-disc ratios from stereoscopic images. Consequently, there is a need for further studies before a definite preference of the two methods is recommended.

An interesting component of this investigation was the nonutilization of pharmacologic pupil dilation for the eye examination. Although there was a relatively high rate of images that was deemed notgradable, only 0.3% and 2.3% of the HRT and NMRC images were attributable to a lack of adequate pupil dilation. The high rejection rate thus was mainly related to factors such as lens opacity or excessive eye movement. Given the advantages of natural physiological dilation, the utilization of a nonspecialist eye staff, the reduced possibility of adverse events to participants, the reduced costs, and the increased participants’ response, the relatively small rate of ungradable images related to poor pupil dilation appears to be a reasonable tradeoff. Still, the percentage of eyes that failed to produce an image with acceptable quality had been shown to be lower in similar studies. Sung et al.,16 for example, reported that only 2.6% and 7.7% of eyes had poor HRT and stereoscopic images, respectively. The discrepancy in image-rejection rate is perhaps related to the mean age of our participants, who were substantially older than those in Sung et al. (mean age = 75 and 61 years, respectively). Older participants tend to have smaller pupils or other conditions, such as cataract and corneal opacification, that in turn can reduce the quality of the images or images.

In conclusion, the excellent repeatability and agreement measures found between the HRT and the digital NMRC imply the usefulness of the latter as a valuable and reliable screening method to assess area cup-to-disc ratio in the community. Although we recorded an excellent interobserver reliability result, this study was limited by using only one observer to determine the agreement between the two instruments and with more training and standardization in cup margin determination, this agreement can be further improved. Given that the digital NMRC is already used for diabetic retinopathy screening in the community, is user friendly and relatively easy to transport, uses physiological dilation, and is relative less costly than the HRT, it is likely that it can be a useful community screening tool for glaucoma.

References


