Intraobserver and Interobserver Reproducibility in the Evaluation of Optic Disc Stereometric Parameters by Heidelberg Retina Tomograph

Stefano Miglior, MD,[1] Elena Albé, MD,[2] Magda Guareschi, MD,[2] Luca Rossetti, MD,[2] Nicola Orzalesi, MD[2]

Purpose: To assess intraobserver and interobserver reproducibility of the measurement of stereometric parameters of the optic disc by means of the Heidelberg Retina Tomograph I (HRT).

Study design: Observational study, with interobserver variability.

Participants: Fifty-five volunteers (healthy subjects and patients with glaucoma).

Methods: HRT examination of the optic disc was repeated on 3 consecutive days on 1 eye of each of the 55 subjects. During each session, five single images were randomly acquired by two independent observers. One mean topography image (MTI), based on three single images, was then built at each session. For the intraobserver intraimage evaluation, the two observers traced their own contour line on one randomly chosen MTI. This procedure was repeated three times. For the intraobserver interimage and interobserver intra/interimage evaluations, the two observers traced their own contour line on the MTI of the first session, which was then automatically superimposed on the MTIs of the other two sessions.

Main Outcome Measures: Reproducibility of the 12 stereometric parameters was calculated for each observer by means of the intraclass correlation coefficient (ICC). The expected range of variability between two independent evaluations was calculated by the scatter-plots of each test—retest difference versus their mean. The standard deviation of the mean test—retest score differences was used to describe the spread of score differences.

Results: The ICC ranged between 0.79 and 0.99 for intraobserver intraimage and between 0.56 and 1 for intraobserver interimage evaluation. The ICC ranged between 0.54 and 0.99 for interobserver intraimage and between 0.65 and 0.97 for the interobserver interimage evaluation. ICC was almost perfect to perfect for planimetric measures ($0.81 < \text{ICC} \leq 1$), substantial to almost perfect for volumetric and cup measures ($0.61 < \text{ICC} \leq 0.99$), and moderate to almost perfect for retinal nerve fiber layer related measures ($0.41 < \text{ICC} < 0.99$). The expected variability was low (95% confidence interval, $<\pm 9\%$). Interimage evaluation showed a higher variability than intraimage evaluation in both of interobserver ($P = 0.012$) and intraobserver evaluation ($P = 0.028$ and $P = 0.031$ for the two observers).

Conclusions: Measurement of optic disc stereometric parameters by HRT is highly reproducible. However, the use of retinal nerve fiber layer–related parameters should be taken cautiously. The image acquisition–induced variability seems larger than the operator-induced variability. *Ophthalmology 2002;109:1072–1077 © 2002 by the American Academy of Ophthalmology.*

Primary open-angle glaucoma consists of a progressive loss of retinal ganglion cells that leads to subsequent visual field damage. Retinal ganglion cell loss is clinically detectable as a structural modification of the optic disc or as a loss of retinal nerve fiber layer (RNFL) reflectivity, particularly when using a blue or green light.[1] Because the optic disc and RNFL are usually evaluated by means of subjective techniques (ophthalmoscopy, retinography, etc.) that clearly rely on the examiner's experience, it seems advantageous to introduce objective techniques to obtain quantitative measurements. The results of such quantitative analysis should improve the ability not only to distinguish a normal from an abnormal (glaucomatous) optic disc and/or RNFL but also to distinguish true from false progressive optic disc/RNFL changes, which is as important as making an accurate early diagnosis in the correct clinical management of the disease.

This has led to the development of a number of new

Originally received: July 13, 2001.
Accepted: September 20, 2001. Manuscript no. 210482.

[1] Dipartimento di Neuroscienze e Tecnologie Biomediche, Universita' degli Studi di Milano Bicocca, Milan, Italy.

[2] Dipartimento di Medicina, Chirurgia ed Odontoiatria, Universita' degli Studi di Milano, Milan, Italy.

Presented in part at the annual meeting of the Association for Research in Vision and Ophthalmology (ARVO), Ft. Lauderdale, Florida, May 2001.

Funded by a grant from the Ministero dell'Universita' e della Ricerca Scientifica, Italy.

Reprint requests to Stefano Miglior, MD, Clinica Oculistica, Ospedale S. Paolo, Via di Rudini' 8, 20142 Milano, Italy.

© 2002 by the American Academy of Ophthalmology
Published by Elsevier Science Inc.
ISSN 0161-6420/02/$–see front matter
PII S0161-6420(02)01032-1

devices over the last few years, some of which provide quantitative measures of a variety of optic disc parameters,[2,3] and others of which are designed to measure only RNFL[4,5] or entire retinal thickness.[5,6]

The Heidelberg Retina Tomograph (HRT) is a scanning laser ophthalmoscope that allows three-dimensional topographic analysis of the optic disc and the retina by providing rapid measurements of optic disc topography on a pixel-by-pixel basis and analyzing a number of optic disc parameters. However, this requires manually defining the optic disc by tracing a contour line along the disc margin and thus introduces a potential source of variability that may negatively affect the result of the examination. It also needs to be borne in mind that various observers may be involved in the process of image acquisition and processing; this introduces another source of error that is particularly important during the long-term follow-up typical of glaucomatous patients.

For this reason caution is necessary when considering serial quantitative HRT data, unless reproducibility can be precisely estimated. Although the reproducibility of the topographic measures provided by HRT has been widely investigated, most of the studies have addressed intraobserver interimage reproducibility.[2,7–11] Only Orgül et al[12] have evaluated the intraobserver intraimage and interimage reproducibility, and they found a wide range of values for the coefficient of variation of the repeated measures ($<0.01-0.24$, depending on the topographic parameter). Only Hatch et al[13] have evaluated interobserver reproducibility. This last study was limited to the intraimage assessment of a single original topography image, and they found a substantial to almost perfect agreement in the evaluation of intraclass correlation coefficient (ICC) (0.67–0.94, depending on the topographic parameter).

Because no comprehensive data have yet been published, concerning the reproducibility of the measurement of HRT stereometric parameters of the optic disc (including intraobserver and interobserver, as well as intraimage and interimage evaluations), this study was designed with the following aims: (1) to assess reproducibility of HRT optic disc measurements as they are usually made in the clinical setting (thus including all the possible sources of variability inherent in the procedure); and (2) to estimate the HRT measurements that can be expected in eyes believed not to have experienced optic disc progressive changes.

Material and Methods

The study population consisted of 55 volunteers aged between 25 and 79 years attending the Outpatient Service at the S. Paolo Hospital. The protocol of the study was approved by the local ethical review committee. After informed consent was obtained, one eye was randomly chosen for each subject. The sample included 11 eyes affected by primary open angle glaucoma at different stages and eyes with a wide range of disc, rim, and cup sizes and different morphologic characteristics. The glaucomatous eyes were pooled with normal eyes to have a group of eyes with completely different morphologic features (very narrow or very thick neuroretinal rims, enlarged cups or absence of cups, cups with temporal flat slopes, steep "punched-out" cups, etc.), which may affect the reproducibility of HRT topographic disc measurements.[14,15]

Inclusion criteria were 20/20 visual acuity, a clear lens, and normal retina. Exclusion criteria were myopia >-6 diopters, optic disc abnormality (i.e., drusen or tilted disc), and history of neurophthalmologic diseases. All the patients included in the study underwent a complete ophthalmologic evaluation. Biomicroscopy of the anterior segment and intraocular pressure measurement by means of Goldmann tonometry were made before pupillary dilatation. The indirect ophthalmoscopic fundus oculi evaluation and HRT examination were made in mydriasis. Visual field was determined using the standard 30/II Humphrey full-threshold strategy, and primary open angle glaucoma was defined on the basis of the criteria suggested by the European Glaucoma Society.[16]

The HRT (Heidelberg Instruments, Heidelberg, Germany-version 2.01) is a confocal scanning laser ophthalmoscope that uses a diode laser (wavelength 670 nm) to scan the retinal surface in three dimensions. A topographic image consists of a series of 32 confocal images (i.e., 32 optical sections) at 32 consecutive focal planes, each made up of 256×256 pixels. The HRT examination was performed using a 10° angle view. The reference plane was set at the standard value of 50 μm below the mean retinal surface at the temporal sector between 350° and 356°.

For study purposes, five image series were acquired each day for 3 consecutive days (a total of 15 image series per eye) by two independent and well-trained observers. All the acquired image series were recorded only after the built-in quality control of the HRT software confirmed that the optimal settings were found. Each day a mean topographic image (MTI) was then computed from three of the five original image series (a total of 3 MTIs per eye), by excluding the image series in which the optic disc was not well centered in the monitor.[17]

Intraobserver intraimage and *interimage* reproducibility for each of the 12 predefined stereometric parameters was tested according to the following procedure:

- For the intraimage comparison, each observer manually drew the contour line on the same MTI three times. This procedure was performed on one MTI only, randomly chosen from the three MTIs.
- For the interimage comparison, first observer E drew the contour line and then observer M did the same on MTI-1. The two contour lines were then automatically superimposed by the HRT software on the MTI-2 and MTI-3.

Interobserver intraimage and *interimage* reproducibility for each of the 12 parameters was tested according to the following procedure, using the same contour lines previously drawn by both observers for the intraobserver/interimage evaluation:

- For the intraimage comparison, we analyzed the three combinations obtained at the same sessions by observer E and observer M (E1–M1, E2–M2, E3–M3).
- For the interimage comparison, we analyzed the six combinations obtained during the different sessions by observer E and observer M (E1–M2, E1–M3, E2–M1, E2–M3, E3–M1, E3–M2).

Reproducibility was evaluated by means of the ICC using the formula:

$$r_1 = \sigma_a^2/(\sigma^2 + \sigma_a^2)$$

where: r_1 = ICC; σ^2 = QM(e) or variability intraclass; σ_a^2 = [QM(a) − QM(e)]/n or (interclass–intraclass variability)/n° of classes.[18–20]

The ICC was calculated for each independent test-retest evaluation, after which the mean ICCs were calculated with their ranges. The agreement was judged to be poor (an ICC of between

Table 1. Descriptive Statistics of Topographic Parameters. Mean ± Standard Deviation (Range).

	Observer E	Observer M	
Disc area	1.856 ± 0.378 (0.917–2.796)	1.888 ± 0.398 (1.017–2.896)	mm^2
Cup area	0.636 ± 0.434 (0.001–1.915)	0.653 ± 0.426 (0.000–1.857)	mm^2
Cup/disc area	0.319 ± 0.184 (0.001–0.784)	0.323 ± 0.175 (0.001–0.737)	mm^2
Rim area	1.220 ± 0.289 (0.526–1.890)	1.234 ± 0.274 (0.609–1.914)	mm^2
Height variation contour	0.409 ± 0.107 (0.210–0.648)	0.402 ± 0.109 (0.208–0.621)	mm^2
Cup volume	0.170 ± 0.174 (0.001–0.831)	0.168 ± 0.167 (0.000–0.786)	mm^3
Rim volume	0.323 ± 0.143 (0.065–0.749)	0.318 ± 0.134 (0.083–0.707)	mm^3
Mean cup depth	0.242 ± 0.107 (0.030–0.504)	0.241 ± 0.106 (0.023–0.487)	mm
Maximum cup depth	0.617 ± 0.204 (0.085–1.069)	0.619 ± 0.204 (0.069–1.078)	mm
Cup shape measure	−0.163 ± 0.083 (−0.360–+0.004)	−0.166 ± 0.082 (−0.340–−0.017)	
Mean RNFL thickness	0.245 ± 0.079 (0.087–0.426)	0.242 ± 0.078 (0.097–0.432)	mm
RNFL cross-sectional area	1.164 ± 0.377 (0.481–2.064)	1.143 ± 0.346 (0.524–1.875)	mm^2

RNFL = retinal nerve fiber layer.

−1 and 0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), almost perfect (0.91–0.99), or perfect (equal to 1).

The expected range of variability between two independent evaluations of each stereometric parameter was obtained by assessing test-retest variability. This was calculated using the paired score differences of the test minus the retest (i.e., first minus second) for each patient. A scatterplot was then built by plotting each test-retest difference versus its mean. The mean and standard deviation of the test-retest score differences were then calculated. The standard deviation from the mean test-retest score differences was used to describe the spread of score differences, which is an estimate of their dispersion between repeated measurements. The 95% confidence intervals of the mean difference are thus the boundaries of the expected range of variability.[21] Because the mean difference between the two repeated measurements performed on our 55 cases tends to zero, the 95% confidence intervals are expressed as percent values around the mean measurement.

Statistical comparisons were performed by means of Wilcoxon's nonparametric test.

Results

The study is based on a total of 165 MTIs, whose mean standard deviation was 27.87 ± 10.56 (range, 11.51–66.03).

The descriptive statistics values of the 12 parameters for each observer are summarized in Table 1; there was no statistically significant difference between the two observers for any of the parameters.

ICC

The analysis of the intraobserver intraimage reproducibility showed a mean ICC ranging from substantial to almost perfect (between 0.79 and 0.99) for each of the 12 parameters (Table 2).

The analysis of the intraobserver interimage reproducibility showed a mean ICC ranging from moderate to perfect (between 0.56 and 1.00) for each of the 12 parameters (Table 2).

The analysis of the interobserver intraimage reproducibility showed a mean ICC ranging from moderate to almost perfect (between O.54 and 0.99) for each of the 12 parameters (Table 3).

The analysis of the interobserver interimage reproducibility showed a mean ICC ranging from substantial to almost perfect (between 0.65 and 0.97) for each of the 12 parameters (Table 3).

Scatterplots

The expected measurements for the intraobserver interimage and interobserver interimage evaluations are reported in Tables 4 and 5. All the mean expected measurements were within ±10%, most being within ±5%. Most of the expected measurements coming from the single test–retest comparisons were within ±10%, with

Table 2. Mean (Range) Intraclass Correlation Coefficient of Intraobserver Evaluations

	Intraimage		Interimage	
	Observer E	*Observer M*	*Observer E*	*Observer M*
Disc area	0.95 (0.90–0.98)	0.96 (0.93–0.98)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Cup area	0.98 (0.97–1.00)	0.97 (0.95–0.99)	0.99 (0.96–1.00)	0.98 (0.97–0.99)
Cup/disc area	0.99 (0.98–0.99)	0.98 (0.96–0.99)	0.99 (0.98–0.99)	0.97 (0.96–0.99)
Rim area	0.96 (0.93–0.99)	0.99 (0.98–1.00)	0.99 (0.99–0.99)	0.89 (0.82–0.99)
Height variation contour	0.96 (0.93–0.99)	0.80 (0.66–0.99)	0.74 (0.52–0.94)	0.81 (0.66–0.97)
Cup volume	0.97 (0.96–1.00)	0.95 (0.93–0.99)	0.82 (0.68–0.99)	0.96 (0.93–0.99)
Rim volume	0.98 (0.97–0.99)	0.96 (0.93–0.99)	0.79 (0.65–0.98)	0.89 (0.80–0.96)
Mean cup depth	0.99 (0.99–1.00)	0.99 (0.99–0.99)	0.94 (0.90–0.99)	0.91 (0.85–0.99)
Max cup depth	0.90 (0.84–0.99)	0.99 (0.98–0.99)	0.88 (0.78–0.92)	0.91 (0.83–0.96)
Cup shape measure	0.96 (0.92–0.98)	0.99 (0.98–0.99)	0.92 (0.86–0.97)	0.86 (0.73–0.96)
Mean RNFL thickness	0.92 (0.85–0.98)	0.79 (0.60–0.91)	0.56 (0.29–0.97)	0.77 (0.61–0.97)
RNFL cross-sectional area	0.93 (0.87–0.97)	0.83 (0.67–0.91)	0.64 (0.43–0.99)	0.83 (0.69–0.96)

RNFL = retinal nerve fiber layer.

Table 3. Mean (Range) Intraclass Correlation Coefficient of Interobserver Evaluations

	Intraimage	Interimage
Disc area	0.69 (0.69–0.69)	0.69 (0.69–0.69)
Cup area	0.91 (0.85–0.96)	0.91 (0.86–0.94)
Cup/disc area	0.97 (0.93–1.00)	0.97 (0.91–0.99)
Rim area	0.84 (0.74–0.94)	0.85 (0.77–0.91)
Height variation contour	0.77 (0.64–0.89)	0.65 (0.22–0.97)
Cup volume	0.96 (0.90–0.99)	0.90 (0.70–0.96)
Rim volume	0.94 (0.83–0.99)	0.84 (0.69–0.98)
Mean cup depth	0.99 (0.98–0.99)	0.92 (0.83–0.99)
Max cup depth	0.99 (0.98–0.98)	0.89 (0.73–0.98)
Cup shape measure	0.85 (0.77–0.90)	0.81 (0.60–0.99)
Mean RNFL thickness	0.54 (0.39–0.72)	0.83 (0.55–0.99)
RNFL cross-sectional area	0.82 (0.63–0.97)	0.65 (0.28–0.93)

RNFL = retinal nerve fiber layer.

Table 5. Mean Expected Variability (Range [%]), Expressed in Terms of ±95% Confidence Interval Interobserver Evaluations

	Intraimage	Interimage
Disc area	±1.01 (1.01–1.01)	±1.01 (1.01–1.01)
Cup area	±1.96 (1.78–2.24)	±3.90 (3.40–4.67)
Cup/disc area	±2.17 (2.01–2.45)	±3.99 (3.44–4.89)
Rim area	±1.58 (1.53–1.66)	±2.37 (2.19–2.76)
Height variation contour	±1.98 (1.53–2.18)	±4.54 (4.21–4.79)
Cup volume	±5.73 (2.09–12.35)	±8.98 (5.71–14.05)
Rim volume	±3.30 (2.95–3.96)	±6.74 (6.02–7.45)
Mean cup depth	±0.64 (0.57–0.71)	±2.28 (1.83–2.62)
Max cup depth	±0.86 (0.40–1.78)	±2.73 (2.18–3.02)
Cup shape	±3.35 (2.09–5.27)	±5.72 (4.71–7.25)
Mean RNFL thickness	±5.92 (2.46–11.66)	±7.41 (5.22–11.50)
RNFL cross-sectional area	±2.85 (2.52–3.39)	±5.75 (5.02–6.56)

RNFL = retinal nerve fiber layer.

only "cup volume" and "mean RNFL height" having a maximum value between ±10% and ±15%. The results arising from the intraobserver intraimage evaluation and interobserver intraimage evaluation were similar to those given in the tables.

The standard deviations of the differences of the intraobserver interimage evaluations of observers E and M were not significantly different ($P = 0.436$ Wilcoxon's test).

The standard deviations of the differences between the intraobserver intraimage and interimage evaluations were significantly different for both observers E ($P = 0.028$) and M ($P = 0.031$).

The standard deviations of the differences between the intraimage and interimage analysis in the interobserver evaluation were significantly different ($P = 0.0128$).

Discussion

Primary open-angle glaucoma patients are affected by optic nerve head modifications and retinal nerve fiber loss, as well as by progressive visual field alterations. To plan a therapy designed to preserve visual function for as long as possible, it seems clinically appropriate to use instruments capable of detecting progressive structural changes.

An important step in the follow-up of a chronic disease is to know whether the observed changes are related to test variability or to real disease progression. For this reason, it is useful to know the major sources of variability, which parameters may be negatively affected, and what value could be expected solely from the "background noise" of the measurement. This study evaluated the observer-related and image-related sources of HRT variability when two independent observers analyze the same (intraobserver intraimage) or different images (intraobserver interimage), as well as the reproducibility of the two independent observers (interobserver intraimage and interimage).

The reproducibility of the measurements was high for most of the stereometric parameters, with the mean ICC ranging from 0.81 and 0.99 for cup- and rim-related parameters, and from 0.54 and 0.96 for RNFL-related parameters (height variation contour, RNFL cross-sectional area and mean RNFL thickness). In the single test-retest evaluations, the ICC ranged from 0.60 and 1.0 for cup- and rim-related parameters, and from 0.28 and 0.99 for RNFL-related parameters. The mean expected variability, estimated by means of the scatterplot (at a confidence level of 95%), was between ±0.55% and ±8.98% for cup- and rim-related parameters, and between ±2.25% and ±9.65% for RNFL-

Table 4. Mean Expected Variability (Range) (%), Expressed in Terms of ±95% Confidence Interval Intraobserver Evaluations

	Intraimage		Interimage	
	Observer E	*Observer M*	*Observer E*	*Observer M*
Disc area	±0.72 (0.68–0.77)	±1.28 (1.15–1.42)	±0.00 (0.00–0.00)	±0.00 (0.00–0.00)
Cup area	±1.85 (1.30–2.24)	±2.71 (2.37–3.12)	±3.51 (3.25–3.85)	±3.49 (3.25–3.82)
Cup/disc area	±1.77 (1.33–2.08)	±2.49 (2.30–2.63)	±3.48 (3.27–3.83)	±3.49 (3.25–3.87)
Rim area	±1.24 (1.17–1.35)	±2.07 (1.94–2.19)	±1.83 (1.72–2.00)	±2.32 (1.70–2.76)
Height variation contour	±2.84 (1.68–3.48)	±2.56 (2.52–2.62)	±4.12 (3.86–4.32)	±4.30 (4.10–4.50)
Cup volume	±3.20 (2.55–3.66)	±4.59 (3.93–5.34)	±3.79 (1.93–13.69)	±2.07 (0.87–6.41)
Rim volume	±2.58 (2.04–3.02)	±3.90 (3.71–4.25)	±5.89 (5.47–6.11)	±6.24 (5.50–6.71)
Mean cup depth	±0.54 (0.51–0.59)	±0.73 (0.63–0.80)	±2.13 (1.75–2.49)	±2.25 (1.80–2.58)
Max cup depth	±3.07 (0.87–4.25)	±0.55 (0.53–0.56)	±2.66 (2.17–2.92)	±2.75 (2.28–3.01)
Cup shape measure	±1.52 (1.39–1.62)	±0.55 (2.21–2.47)	±6.45 (4.67–7.25)	±4.29 (2.68–5.56)
Mean RNFL thickness	±2.25 (1.66–2.56)	±3.16 (3.09–3.27)	±5.02 (4.46–5.54)	±9.65 (5.42–11.82)
RNFL cross-sectional area	±2.27 (1.62–2.61)	±3.22 (3.14–3.29)	±4.92 (4.43–5.49)	±5.38 (4.64–6.00)

RNFL = retinal nerve fiber layer.

related parameters. In relation to the single test-retest evaluations, the expected variability at the same confidence level ranged from ±0.40% and ±14.05% for cup- and rim-related parameters and from ±1.53% and ±11.82% for RNFL-related parameters.

Of the two major sources of variability (the tracing of the contour line and the mean image construction, which includes the acquisition of the three original image series and the software processing to build the mean image), the first was less influential insofar as the mean expected variability of all the stereometric parameters was lower for intraimage than for interimage evaluation. The difference was statistically significant in the case of both the intraobserver ($P = 0.028$ for observer E; $P = 0.031$ for observer M) and interobserver evaluation ($P = 0.012$). It should be noted that the mean expected variability in disc area was 0% in the case of the intraobserver interimage evaluation, because the first contour line was automatically superimposed on the next images (and thus theoretically did not affect their analysis); conversely, the mean expected variability in disc area ranged from ±0.72% to ±1.28% in the case of the intraobserver intraimage and interobserver evaluations, thus being extremely low. These results confirm the recent findings of Iester et al,[22]who reported good agreement in the interobserver intraimage evaluation of disc area when four glaucoma experts independently drew their own contour line on the same HRT MTI (Kendall's W coefficient = 0.92).

Reproducibility in the evaluation of HRT stereometric parameters has been tested in a number of previous studies, but these did not attempt to make a comprehensive assessment of the different sources of variability that may occur in a clinical setting.[2,7–13] Our results are comparable with those of Azuara-Blanco et al[11]in terms of the intraobserver evaluation and those provided by Hatch et al[13] in terms of the interobserver evaluation. We confirmed the previous findings of Orgül et al,[12]who reported that the contour line tracing induced less variability than image acquisition/processing. The other previous studies used the coefficient of variation to evaluate reproducibility, which cannot appropriately be compared with the ICC we used.[2,7–10]

The figures of the interimage evaluations, represent a common clinical situation in which the same observer may use the same contour line to analyze follow-up images or two or more different observers may use their own contour line for the same type of analysis. On the basis of these figures it seems possible to speculate that the "background noise" induced by test-retest variability is extremely low in the case of HRT stereometric parameters of those eyes believed not to have experienced any structural optic disc changes. This holds true for most of the parameters, the 95% confidence limits of variability of which are within 10% of the mean value.

New statistical analyses have recently been developed to assess longitudinal changes of the optic disc by HRT. Chauhan et al[23] have developed software that can adjust each follow-up image to the baseline image to detect subtle changes in the topography of each calculated "megapixel," with the aim of avoiding the use of a standard reference plane (a potential major source of error) and standard stereometric parameters. This technique has greatly increased the specificity of detecting longitudinal optic disc changes. Although Chauhan's approach is extremely interesting and valuable, other comparative methods for the longitudinal HRT follow-up of glaucoma are readily available, such as the direct comparison of baseline and subsequent follow-up stereometric parameters, which is simple and easy to perform, and allows investigators to work on mean topography images and standard parameters.

Greater care should be taken when acquiring the images, because this was the greatest source of variability in our HRT examination. On the basis of our figures relating to a series of 55 normal and glaucomatous eyes with a standard deviation of MTIs ranging from 11.5 to 66 μm, progressive changes of more than 10% of the baseline values may be considered statistically significant (at 95% confidence limit) for most of the standard stereometric parameters. To increase specificity, the variability of each patient should be taken into account and calculated using a simple test-retest analysis made over a short time span. Whether these statistically significant changes are also clinically significant can only be addressed by means of ongoing longitudinal clinical trials.

References

1. Spaeth GL. Development of glaucomatous changes of the optic nerve. In: Varma R, Spaeth GL, Parker KW, eds. The Optic Nerve in Glaucoma. Philadelphia: JB Lippincott,1993; 63–81.
2. Kruse FE, Burk RO, Volcker HE, et al. Reproducibility of topographic measurements of the optic nerve head with laser tomographic scanning. Ophthalmology 1989;96:1320–4.
3. Dreher AW, Tso PC, Weinreb RN. Reproducibility of topographic measurements of the normal and glaucomatous optic nerve head with the laser tomographic scanner. Am J Ophthalmol 1991;111:221–9.
4. Weinreb RN, Dreher AW, Coleman A, et al. Histopathologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness. Arch Ophthalmol 1990;108:557–60.
5. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. Arch Ophthalmol 1995; 113:586–96.
6. Zeimer R, Shahaidi M, Mori M, et al. A new method for rapid mapping of retinal thickness at the posterior pole. Invest Ophthalmol Vis Sci 1996;37:1994–2001.
7. Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg Retina Tomograph. J Glaucoma 1993;2:101–3.
8. Rohrschneider K, Burk ROW, Volcker HE. Reproducibility of topometric data acquisition in normal and glaucomatous optic nerve heads with the laser tomographic scanner. Graefes Arch Clin Exp Ophthalmol 1993;231:457–64.
9. Rohrschneider K, Burk ROW, Kruse FE, Volcker HE. Reproducibility of the optic nerve head topography with a new laser tomographic scanning device. Ophthalmology 1994;101: 1044–9.
10. Janknecht P, Funk J. Optic nerve head analyser and Heidelberg retina tomograph: accuracy and reproducibility of topographic measurements in a model eye and in volunteers. Br J Ophthalmol 1994;78:760–8.

11. Azuara-Blanco A, Harris A, Cantor LB. Reproducibility of optic disk topographic measurements with the Topcon ImageNet and the Heidelberg Retina Tomograph. Ophthalmologica 1998;212:95–8.
12. Orgül S, Croffi GA, VanBuskirk EM. Variability of contour line alignment on sequential images with the Heidelberg Retina Tomograph. Graefes Arch Clin Exp Ophthalmol 1997; 235:82–6.
13. Hatch WV, Flanagan JG, Williams-Lyn DE, et al. Interobserver agreement of Heidelberg retina tomograph parameters. J Glaucoma 1999;8:232–7.
14. Brigatti L, Weitzman M, Caprioli J. Regional test-retest variability of confocal scanning laser tomography. Am J Ophthalmol 1995;120:433–40.
15. Chauhan BC, LeBlanc RP, McCormick TA, Rogers JB. Test-retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. Am J Ophthalmol 1994;118:9–15.
16. European Glaucoma Society. Terminology and Guidelines for Glaucoma. Savona, Italy. DOGMA, 1998;43–8.
17. Weinreb RN, Lusky M, Bartsch DU, Morsman D. Effect of repetitive imaging on topographic measurements of the optic nerve head. Arch Ophthalmol 1993;111:636–8.
18. Fleiss JJ, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 1973;33:613–9.
19. Snedecor GW, Cochran WG. Statistical Methods, 8th ed. Ames, IA: Iowa State University Press, 1989;242–3.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
22. Iester M, Mikelberg FS, Courtright P, et al. Interobserver variability of optic disk variables measured by confocal scanning laser tomography. Am J Ophthalmol 2001;132:57–62.
23. Chauhan BC, Blanchard JW, Hamilton DC, LeBlanc RP. Technique for detecting serial topographic changes of the optic disc and peripapillary retina using scanning laser tomography. Invest Ophthalmol Vis Sci 2000;41:775–82.