The Ocular Hypertension Treatment Study: Reproducibility of Cup/Disk Ratio Measurements Over Time at an Optic Disc Reading Center

WILLIAM J. FEUER, MS, RICHARD K. PARRISH II, MD, JOYCE C. SCHIFFMAN, MS, DOUGLAS R. ANDERSON, MD, DONALD L. BUDENZ, MD, MARIA-CRISTINA WELLS, MPH, DITTE J. HESS, CRA, MICHAEL A. KASS, MD, MAE O. GORDON, PhD, AND THE OCULAR HYPERTENSION TREATMENT STUDY GROUP

• PURPOSE: To determine the reproducibility over time of visual estimates of the horizontal cup/disk ratio by trained technicians from optic disk stereophotographs.
• METHODS: Baseline optic disk stereophotographs are graded at entry and regraded annually in a masked fashion. The 1,636 participants in the Ocular Hypertension Treatment Study (OHTS) undergo stereoscopic optic disk photography at study entry and annually thereafter. Stereophotographs are graded independently by two technicians at the Optic Disc Reading Center. If the readers' estimates of horizontal cup/disk ratio differ by more than 0.2 disk diameters (DD), they attempt to reach a consensus; if they cannot, the horizontal cup/disk ratio is adjudicated by a glaucoma specialist.
• RESULTS: The percent of regradings differing by 0.2 DD or more from the estimate of horizontal cup/disk ratio made at entry was 4%, 6%, and 7%, respectively at years 1, 2, and 3. The percent differing by more than 0.2 DD was 1% or less at all years. Intraclass correlation coefficients were 0.93, 0.92, and 0.92, respectively. Estimates of horizontal cup/disk ratio from sequential full-frame photographs and simultaneous split-frame photographs appeared comparable and equally reproducible. Gradings by technicians were comparable to gradings by glaucoma specialists.
• CONCLUSIONS: High reproducibility between repeated gradings of baseline horizontal cup/disk ratio was achieved by trained technicians adhering to a rigorous protocol. Horizontal cup/disk ratio measurements in OHTS are sufficiently reproducible to provide information about the relationship of cup/disk ratio to the prognosis of individuals with ocular hypertension. (Am J Ophthalmol 2002;133:19–28. © 2002 by Elsevier Science Inc. All rights reserved.)

THINNING OF THE NEURORETINAL RIM, AN IMPORtant indication of glaucomatous optic neuropathy, can be quantified as the ratio of the diameter of either the horizontal or vertical excavation (cup) to the optic nerve head (disk), the cup/disk ratio. The validity of studies [longitudinal and cross-sectional studies are both affected by reliability] of the incidence and progression of glaucoma which assess disease with the cup/disk ratio depend on the reproducibility of these measurements. Lichter[1] concluded that significant variability existed between glaucoma specialists in the evaluation of cup/disk ratios from stereoscopic photographs and that numerical methods were not reliable in judging small changes. More recently, Tielsch and associates[2] demonstrated that 17%–19% of cup/disk estimates made by two different glaucoma specialists differed by 0.2 disk diameters (DD) or more. Varma[3,4] and Zangwill[5] have reported similar results for estimates by glaucoma specialists.

Accepted for publication Oct 10, 2001.
From the Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami School of Medicine, Miami, Florida 33136 (W.J.F., R.K.P., II, J.C.S., D.R.A., D.L.B., M.C.W., D.J.H.); Department of Ophthalmology and Visual Sciences, Washington University, St. Louis, Missouri 63110 (M.A.K., M.O.G.).
A complete list of the participating clinics, committees, and resource centers in the Ocular Hypertension Treatment Study is at the end of this article.
Supported by grants EYO9341 (Dr. Gordon) and EY09307 (Dr. Kass) from the National Eye Institute, National Institutes of Health, Bethesda, MD, by the Office of Research of Minority Health, National Institutes of Health, by Merck Research Laboratories, West Point, PA (Dr. Kass), and by unrestricted grants from Research to Prevent Blindness Inc, New York, NY.
Presented as a poster at the Association for Research in Vision and Ophthalmology meeting, Fort Lauderdale, FL, May 10, 1999.
Reprint requests to Mae O. Gordon, PhD, OHTS Coordinating Center, Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, Box 8203, 660 So. Euclid, St. Louis MO 63110

0002-9394/02/$22.00
PII S0002-9394(01)01338-1

The Ocular Hypertension Treatment Study (OHTS) is a multicenter randomized clinical trial to evaluate the safety and efficacy of topical ocular hypotensive therapy in preventing or delaying the onset of visual field loss or optic nerve damage due to primary open-angle glaucoma in individuals with ocular hypertension.[6] OHTS has 1,636 participants who were randomized to either the close observation group or the medication group. Determination of optic nerve damage is performed at the OHTS Optic Disc Reading Center by masked technicians because clinic personnel know the randomization assignment and clinical status of participants. Several multicenter clinical trials in ophthalmology, such as the Early Treatment Diabetic Retinopathy Study,[7] Cryotherapy for Retinopathy of Prematurity,[8] Glaucoma Laser Trial,[9] and Central Vein Occlusion Trial[10] have employed photography reading centers to provide unbiased, standardized assessment of eligibility and outcome measures.

Cost, efficiency, and reliability are key considerations in the assessment of stereoscopic optic disk photographs by a centralized reading center. Several studies[2–5] have demonstrated fair-to-good agreement between observers when the observers are glaucoma specialists. However, using glaucoma specialists to evaluate stereoscopic optic disk photographs would have been prohibitively costly given the large sample size of OHTS. Klein and associates[11] described a method for measuring cup/disk ratio using a template with circles that was specifically designed for use by technicians. However, the Klein protocol is time consuming, approximately 8 minutes/eye, or approximately 16 minutes/patient.

The OHTS Optic Disc Reading Center adopted a protocol in which trained and certified technicians visually estimate horizontal cup/disk ratio. To our knowledge, this is the first report to describe the reproducibility of visual estimates of cup/disk ratio by trained technicians. This protocol is described in further detail in the Methods section.

This report is limited to data on horizontal cup/disk ratio which are collected in OHTS primarily for the purpose of numeric description and for use in statistical analysis. The estimation of horizontal cup/disk ratio is only one aspect in the determination of progressive optic disk damage in OHTS. The occurrence of progressive optic disk damage in OHTS is determined by a masked side-by-side comparison of the baseline and follow-up stereoscopic optic disk photographs. The determination of a difference between the two sets of stereophotographs yields a "yes" or" no" response, not a numeric value.

Data collected by the OHTS provide a unique opportunity to study the reproducibility of horizontal cup/disk ratio estimates using the Optic Disc Reading Center protocol. We report on agreement of cup/disk ratio estimates from baseline stereoscopic photographs graded at entry and regraded annually thereafter in a masked fashion. The large sample permits us to examine several factors for their potential effect on agreement including camera type (sequential full-frame vs simultaneous split-frame), observer (trained technician readers vs glaucoma specialists), and magnitude of cup/disk ratio (small-to-large cup/disk ratio). We also examine initial interobserver agreement between independent estimates by two primary readers before consensus and/or adjudication.

METHODS

THE PROTOCOL AND BASELINE CHARACTERISTICS OF 1,636 participants enrolled in the OHTS have been reported elsewhere.[6] We report on the reproducibility of horizontal cup/disk estimates of stereoscopic optic disk photographs which were graded at entry for eligibility determination and regraded in a masked fashion annually during follow-up. Only OHTS certified photographers were permitted to take study optic disk stereophotographs. Certification of photographers required completion of stereo sets, both right and left eyes of two patients, with adequate exposure, stereoscopic quality, proper labeling, and completion of forms.

OHTS eligibility criteria required normal optic discs in both eyes on clinical examination and on stereoscopic photographs as determined by the Optic Disc Reading Center, Bascom Palmer Eye Institute, University of Miami, Miami, Florida. Exclusion criteria included the inability to visualize or photograph the optic discs, the presence of a disk hemorrhage, notching, localized pallor, asymmetry in cup/disk ratio of the two eyes > 0.2 DD, or inability to adequately evaluate photographs due to poor photographic quality. Optic disk stereophotographs used to determine eligibility serve as the baseline. Horizontal cup/disk ratio of baseline optic disk stereophotographs was estimated at study entry and annually thereafter in the process of a masked, side-by-side comparison with follow-up photographs. This report only includes horizontal cup/disk ratio data from baseline photographs graded at entry and regraded at years 1, 2, and 3.

• **PROTOCOL FOR OPTIC DISK STEREOPHOTOGRAPHY:** A 2X or 1.6X magnification lens is used or the highest magnification for the fundus camera. All photographs are taken on Ektachrome or Fujichrome 100 film. For sequential full-frame photography, starting with the right eye, the participant is instructed to follow the fixation light until the optic nerve is centered on the cross-hairs. The joystick is tilted to the right at the 3 o'clock position just outside the pupillary crescent focusing at the junction of the retinal pigment epithelium and the neuroretinal rim. This step is repeated with the joystick at the 9 o'clock position. The same protocol is repeated for the left eye. For simultaneous split-frame photography, the participant's head is positioned in the chin rest and the participant is instructed to fixate on the fixation light. The photographer focuses on the participant's eye.

Fundus cameras used by clinical centers include Zeiss, Kowa, or Topcon cameras to obtain full-frame stereoscopic images sequentially or Topcon or Nidek cameras to obtain split-frame images simultaneously. The Donaldson Viewer (George Davco, Holbrook, MA 02343) is used to read two full-frame 35 mm by 25 mm transparencies and the Asahi Stereo Viewer II (Pentax Corp, Englewood, CO 80112) is used to read single split-frame 35 mm by 25 mm transparencies.

• **CERTIFICATION OF OPTIC DISK READERS AT OPTIC DISC READING CENTER:** To be certified as an optic disk photography reader, technicians have to successfully grade a test set of stereophotographs (provided by Joseph Caprioli, MD) side by side with a glaucoma specialist. Training of readers was performed largely by DRA. Upon the successful completion of the test set, the trainee completed independent readings of 50 consecutive sets of slides taken from OHTS. To complete certification, the trainee had to demonstrate the ability to agree to within 0.1 DD of the official reading on all 50 eyes. Complete details of the training of readers at the Optic Disc Reading Center can be found in the OHTS Manual of Procedures.[12]

• **MEASUREMENT OF HORIZONTAL CUP/DISK RATIO:** Stereophotographs are sent to the OHTS Optic Disc Reading Center where they are logged, labeled with random identifiers, and graded. Readers are masked as to randomization, clinic, patient ID, visit, prior gradings, fellow eye grading, and clinical status of the eye. Two primary readers first independently grade the photographs for quality (clarity and stereo). If the two primary readers disagree, the stereophotographs are reviewed by a senior reader whose quality grading becomes the official decision. Photographs whose quality is too poor for assessment are not evaluated further and a new set of photographs is requested.

The primary readers independently conduct side-by-side comparisons of technically adequate baseline and follow-up stereophotographs masked as to their order. The readers estimate horizontal cup/disk ratio of both baseline and follow-up optic disk stereophotographs thereby regrading the baseline stereophotographs with each new set of annual follow-up stereophotographs. The primary readers visually estimate the horizontal cup/disk ratio (3:00 to 9:00 meridian) to the nearest 0.1 DD. The cup is determined by contour. Assignment is straightforward when the orientation of both the disk and the cup are cylindrical and parallel to the optical axis. When the cup is conical, the plane midway between the surface of the disk and the depth of the cup is used as the standard reference plane. When the optic nerve enters the sclera obliquely and the anatomic configuration is tilted with respect to the optical axis, an estimate is made of the horizontal cup/disk ratio at the plane perpendicular to the axis of the insertion of the optic nerve to the eye. This conceptual midplane perpendicular to the optic nerve axis is chosen to provide the greatest consistency of the reading for the frequent condition in which the nasal edge of the cup is steep and the temporal edge is sloping.

If the estimates of horizontal cup/disk ratio of the two primary readers are within 0.2 disk diameters (DD) of each other, the official horizontal cup/disk ratio is the average of the two estimates. If the primary readings differ by more than 0.2 DD, the primary readers attempt to arrive at a consensus and the official horizontal cup/disk ratio becomes the average of the two post-consensus estimates. If agreement within 0.2 DD is not possible, the senior reader, who is a glaucoma specialist, determines the official horizontal cup/disk ratio. The Optic Disc Reading Center protocol was prepared by DRA.

• **ANALYSIS:** We assessed reproducibility of horizontal cup/disk ratio over time with the intraclass correlation coefficient.[13–16] The intraclass correlation is an index of agreement that weights disagreement as the squared distance from perfect agreement. For descriptive purposes only, when the official horizontal cup/disk estimate was the average of two readers' estimates, we rounded the averaged value to the nearest 0.1 DD to allow direct comparisons to published studies. We report the mean and standard deviation of horizontal cup/disk ratio of baseline photographs graded at study entry and regraded at annual follow-up visits at years 1, 2, and 3. We report the mean difference between these readings (follow-up horizontal cup/disk ratio minus entry horizontal cup/disk ratio) and the 95% confidence intervals of the differences. A paired t test was used to determine if the difference between readings was greater than zero, that is, greater than or less than expected by chance alone.

To determine if technician readers differed from glaucoma specialists, we compared the horizontal cup/disk ratio of the same optic disk stereophotographs as graded by technician readers and glaucoma specialists. This analysis was conducted on a subset of stereophotographs which were graded by two glaucoma specialists who served as primary readers early in the study and which were regraded by two technician readers later in the study. We computed intraclass correlation coefficients to describe agreement between glaucoma specialists and technician readers and used a paired t test to determine if the difference was greater than expected by chance.

To estimate interobserver agreement between the two primary readers who graded each set of stereoscopic photographs independently, we sampled the primary readers' worksheets before consensus and/or adjudication. Left eyes of 100 OHTS participants were randomly selected with a computer uniform random number generator. The differences between the primary readers' horizontal cup/disk ratio measurements before consensus and adjudication

TABLE 1. Horizontal Cup/Disk Ratio (Mean ± SD) of Baseline Stereoscopic Optic Disk Photographs Graded at Entry and Regraded at Years 1, 2, or 3

Time of Reread	Regrade Sample N (Eyes)	Entry Horizontal Cup/Disk Ratio Mean ± SD	Regrade Horizontal Cup/Disk Ratio Mean ± SD	Difference in Horizontal Cup/Disk Ratio (Regrade Minus Entry) Mean ± SD	95% Confidence Interval of Difference	*p* Value
Year 1	2953	0.39 ± 0.19	0.39 ± 0.19	0.00 ± 0.08	−0.0009 to 0.005	.168
Year 2	2922	0.38 ± 0.19	0.39 ± 0.20	0.01 ± 0.08	.008 to 0.014	<.001
Year 3	2835	0.38 ± 0.20	0.41 ± 0.20	0.02 ± 0.08	.020 to .026	<.001

were tabulated. Agreement between primary readers was determined using the intraclass correlation coefficient.

RESULTS

BETWEEN FEBRUARY 1994 AND OCTOBER 1996, THE OPTIC Disc Reading Center reviewed stereoscopic optic disk photographs for 2,200 individuals taken by 102 OHTS certified photographers. Fourteen percent (626 of 4,400) of the stereoscopic optic disk photographs screened at baseline needed to be retaken due to poor quality. We report on the agreement in horizontal cup/disk ratio measurements for the 1,636 participants randomized to OHTS who have completed at least one follow-up visit through year 3. Out of a possible 3,272 eyes with stereoscopic optic disk photographs of acceptable quality at baseline, 90% (2,953 eyes) were reread at year 1, 89% (2,922 eyes) were reread at year 2, and 87% (2,835 eyes) were regraded at year 3. There is no evidence of bias in the horizontal cup/disk ratio of participants who are missing stereoscopic photographs for years 1, 2, or 3 compared with participants with stereoscopic photographs.

The median time to complete processing of stereoscopic photographs from the day of receipt, logging, masking, and grading is one day. On average, the time required for the primary reader to grade for photographic quality, estimate baseline horizontal cup/disk ratio, and check for progression is 2–3 minutes/eye. The time increases to 4–5 minutes/eye in cases of poor photographic quality, when disk hemorrhages are present or when horizontal cup/disk ratios are large.

• **AGREEMENT OF REPEAT GRADINGS:** Table 1 reports the mean ± SD of horizontal cup/disk ratio estimates of baseline stereophotographs when graded at study entry, years 1, 2, and 3. The number of stereophotographs differs slightly from year to year reflecting missed follow-up visits. The mean differences between the horizontal cup/disk ratio estimated at entry and year 2 and year 3, while statistically significantly greater than zero ($P < .001$ at both years), are too small to be clinically meaningful. The upper limit of the 95% confidence intervals for the mean differences at all 3 years is less than 0.03 DD. The intraclass correlation coefficients for agreement between the horizontal cup/disk ratio estimated at entry and in successive years of the study are 0.93, 0.92, and 0.92, respectively.

TABLE 2. Distribution of Differences (n, Percent) in Horizontal Cup/Disc Ratio of Baseline Stereoscopic Optic Disk Photographs Graded at Entry and Regraded at Years 1, 2, and 3

Difference Between C/D Ratio Readings*	Time of Regrading		
	Year 1 n (%)	Year 2 n (%)	Year 3 n (%)
<−0.2	4 (0.1)	7 (0.2)	10 (0.4)
−0.2	63 (2.1)	85 (2.9)	130 (4.6)
−0.1	647 (21.9)	777 (26.6)	885 (31.2)
0.0	1591 (53.9)	1480 (50.7)	1348 (47.5)
+0.1	588 (19.9)	502 (17.2)	413 (14.6)
+0.2	50 (1.7)	59 (2.0)	43 (1.5)
>+0.2	10 (0.3)	12 (0.4)	6 (0.2)
Total n (%)	2953 (100%)	2922 (100%)	2835 (100%)

*(Difference = regrade estimate minus entry estimate). A positive difference indicates that the regrade estimate is larger than the entry estimate.

Since a difference of 0.2 DD is considered clinically significant, we determined how often differences between estimates of horizontal cup/disk ratio were 0.2 DD or more. The frequency distributions of the differences between estimates of horizontal cup/disk ratio of baseline stereophotographs graded at entry and at successive years are presented in Table 2. The percentage of rereads that differed by 0.2 DD or more from the estimate at entry are 4%, 6%, and 7% at years 1, 2, and 3, respectively. For each of the 3 years, the percent of rereads that differed by more than 0.2 DD is less than 1%.

• **THE EFFECT OF CUP/DISK RATIO ON REPRODUCIBILITY:** To determine if cup/disk ratio estimated at entry affected reproducibility of rereads, we plotted the difference between the estimate of horizontal cup/disk ratio made at entry and at year 1 by the estimate made at entry (Figure 1). Positive values indicate that the estimate of horizontal cup/disk ratio made at year 1 was larger than the

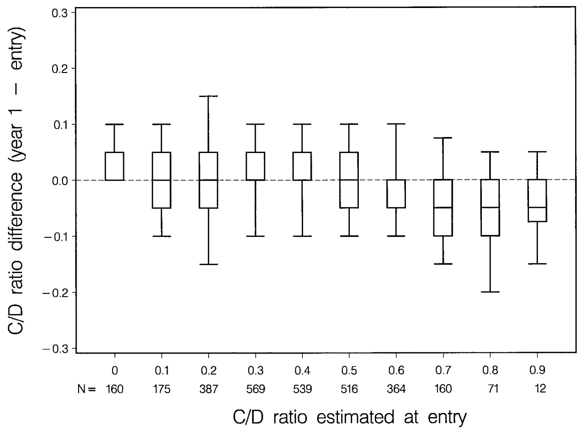

FIGURE 1. Difference between the horizontal cup/disk ratio estimate (year 1 - entry) by the entry horizontal cup/disk ratio. Median is the center of the box. The top and bottom of the box are the 75th and 25th percentiles. The ends of the lines extend to the 95th and 5th percentiles.

estimate made at entry; negative values indicate that the estimate made at year 1 was lower than the estimate made at entry. Agreement between the estimate made at entry and at year 1 is highest for eyes with entry horizontal cup/disk ratio values between 0.1 DD and 0.6 DD. Agreement is lowest both for eyes with entry horizontal cup/disk ratio of 0.0 DD and for eyes with entry horizontal cup/disk ratio greater than 0.6 DD. The horizontal cup/disk ratio of eyes graded 0.0 DD at entry tend to be graded higher at year 1 and eyes graded 0.6 DD or more at entry tend to be graded lower at year 1, suggesting a possible "regression to the mean effect."[16] Graphs of years 2 and 3 data (not shown) are similar.

• **COMPARISON OF SEQUENTIAL FULL-FRAME AND SIMULTANEOUS SPLIT-FRAME PHOTOGRAPHY:** We compared reproducibility of sequential full-frame photographs (n = 2,308–2,418 stereo pairs each year) and simultaneous split-frame photographs (n = 527–535 stereo pairs each year). The mean difference between repeat gradings in year 1 was 0.004 (SD = 0.081) for full-frame photographs and −0.005 DD (SD = 0.077) for split-frame photographs. Differences between repeat gradings for full-frame and split-frame photographs for years 2 and 3 were similar in magnitude to year 1. The percentage of regrades that differed by 0.2 DD or more for years 1, 2, and 3 are 4%, 6%, and 7%, respectively, for full-frame and 4%, 4%, and 5%, respectively, for split-frame photographs. The percentage of regrades that differed by more than 0.2 DD is less than 1% for both full-frame and split-frame photographs for years 1, 2, and 3. Both cameras yielded similar ranges and distributions of horizontal cup/disk ratios for each year of the analysis (Pearson χ^2 test $>$ 0.10 all years).[17] In this study, estimates of horizontal cup /disk ratio from sequential full-frame photographs and simultaneous split-frame photographs appeared comparable and equally reproducible.

• **COMPARISON OF TECHNICIAN READERS AND GLAUCOMA SPECIALISTS:** We compared horizontal cup/disk ratio of the same stereophotographs graded by two glaucoma specialists and regraded by two technician readers. This sample is limited to stereophotographs that did not require senior reader adjudication. These data, presented in Table 3, show good agreement between technician readers and glaucoma specialist readers. Differences, when they are present, are too small to be clinically significant. Differences of 0.2 DD or more at years 1, 2, and 3 occurred 5%, 7%, and 8%, respectively; the intraclass correlation coefficients are 0.92, 0.89, and 0.90, respectively.

• **AGREEMENT BETWEEN PRIMARY READERS:** The high agreement between estimates of horizontal cup/disk ratio could possibly reflect the consensus/adjudication process and could mask poor initial agreement between primary readers.

To evaluate this hypothesis, we selected a random sample of 100 independent pre-consensus/adjudication readings of baseline photographs graded at entry. [Of this sample, 53 photographs were read by the two technician readers, 19 by two glaucoma specialists, and 28 by a technician reader and a glaucoma specialist.] In this sample of pre-consensus readings, 7% differed by 0.2 DD or more. Only 1 out of the 100 independent readings between the primary readers differed by more than 0.2 DD. Table 4 provides the distribution of the absolute value of the differences between primary readers. The intraclass correlation coefficient of agreement between primary readers before consensus/adjudication is 0.89.

DISCUSSION

THE TASK OF ESTIMATING OPTIC DISK PARAMETERS RELIABLY is complex and challenging. In the OHTS, optic disk stereophotographs are taken at entry and annually thereafter. Determination of progressive optic disk damage is performed by a masked, side-by-side comparison of baseline and follow-up slide sets to ascertain thinning of the neuroretinal rim in any meridian. The side-by-side comparison yields a "yes" or "no" answer. Horizontal cup/disk ratio, which is estimated for both baseline and follow-up slide sets, is graded for descriptive and analytic purposes. Agreement between regradings of horizontal cup/disk ratio in the OHTS is among the highest reported to date; the intraclass correlation coefficient is 0.92 or higher for all 3 years of repeated gradings. The simple percent of gradings that differ by 0.2 DD or more is 4%, 6%, and 7% for the 3 years, respectively. The percent differing by more than 0.2 DD was 1% or less at all years. This high rate of agreement is consistent for each year of study, for both

TABLE 3. Comparison of Horizontal Cup/Disk Ratio (Mean ± SD) of Technician Readers and Glaucoma Specialists Grading the Same Baseline Stereoscopic Optic Disk Photographs

Time of Regrading	N	Glaucoma Specialists (Graded at Baseline) Mean ± SD	Trained Technician (Regraded at Follow Up) Mean ± SD	Difference* Mean ± SD	95% Confidence Interval of Difference	*P* Value†
Year 1	95	0.37 ± 0.19	0.38 ± 0.18	0.02 ± 0.08	0.001 to 0.03	.048
Year 2	368	0.38 ± 0.20	0.38 ± 0.18	0.02 ± 0.09	−.007 to .01	.65
Year 3	379	0.38 ± 0.21	0.40 ± 0.20	0.01 ± 0.10	.003 to .02	.011

*Difference = (technician estimate minus glaucoma specialist estimate).
†Paired *t* test.

TABLE 4. Distribution of Differences (Percent) in Horizontal Cup/Disk Ratio Between Primary Readers Prior to Consensus Adjudication in a Random Sample of 100 Stereoscopic Optic Disk Photographs

Absolute Value of Difference Between C/D Ratio Readings	Percent n = 100
0.0	73%
0.1	20%
0.2	6%
0.3	1%
	100%

sequential and split-frame cameras, for technician readers and glaucoma specialists, and for independent gradings of two primary readers before consensus/adjudication.

The OHTS protocol for evaluation of horizontal cup/disk ratio appears cost-effective. The average time to review a set of photographs is 2–3 minutes/eye for the majority of stereoscopic photographs and 4–5 minutes/eye for photographs of poor quality. Processing of stereophotographs from the time of receipt to completion of review is one day.

Previous studies provide testimony to the fact that high interobserver or intraobserver agreement is not easily achieved. Table 5[2–5,11,18–22] is a list, not necessarily complete, of published studies on the reproducibility of cup/disk ratio estimates from optic disk stereophotographs. It is important to emphasize that all studies in Table 5 describe agreement between repeat gradings of the same optic disk stereophotographs and are not studies of the reproducibility of the data acquisition process. Since most clinicians would consider a change of ≥ 0.2 DD in cup/disk ratio to be clinically significant, it is troubling that some studies report that 20% or more of the gradings between observers disagree by this amount.[5,11,18]

Despite some discouraging results in Table 5, these studies should not be interpreted as a sweeping indictment of optic disk evaluation from stereophotography. Though most of these studies utilized a formal written protocol for grading the optic disk, only OHTS and Klein[19] report if observers were required to demonstrate mastery of the standardized protocol and meet minimum performance criteria before evaluation of study data. The goal of some of these studies was to explore the complexity of grading stereoscopic optic disk photographs among glaucoma specialists.[1–5]

The high agreement observed between repeat gradings in the OHTS reflects the entire quality assurance protocol of the OHTS Optic Disc Reading Center, starting with data acquisition. The optic disk photography protocol includes certification of photographers, standardized photography at 22 clinical centers, and ongoing monitoring of protocol adherence by the Optic Disc Reading Center. Grading of optic disk photographs at the OHTS Optic Disc Reading Center begins with assessment of the technical quality of the stereophotographs. Two to three percent of the stereoscopic photographs are not graded due to poor technical quality and these eyes are rephotographed. A multistage protocol protects against reader variability. These safeguards include the completion of independent grading by two primary readers and a consensus grading when the two readers differ by more than 0.2 DD. Unresolved differences are referred to senior readers. Thus, high agreement between repeated gradings reflects the entire flow of data acquisition and processing not only the performance of readers at the Optic Disc Reading Center. The OHTS optic disk photography protocol was designed for research purposes and these results cannot be generalized to the reproducibility of cup/disk ratio estimates by ophthalmoscopy or stereoscopic optic disk photography in routine clinical practice. The Optic Disc Reading Center protocol is similar to those used by the Wisconsin Fundus Photography Reading Center[7] and other ophthalmologic photography reading centers.[8–10]

We wondered if the OHTS sample, which reflects OHTS eligibility criteria for a broad range of ocular and systemic conditions in addition to those specific to optic disk status, might have resulted in a sample that disposed towards high interobserver agreement. OHTS eligibility criteria does not exclude small cup/disk ratios, highly myopic eyes, or eyes with tilted discs which have been reported to have lower interobserver agreement.[3,20,22] The

TABLE 5. Observer Agreement by Camera Type, Sample Size, Patient Diagnosis, and Observer. Agreement is Reported as Kappa Statistic Unless Otherwise Indicated

Author	Camera	N of Eyes Diagnosis	Observers	Cup/Disc Ratio Agreement V = Vertical H = Horizontal: Interobserver	Intraobserver	Percent Difference ≥ 0.2 DD: Interobserver	Intraobserver
Hitchings and associates[22]	Full frame	60 OHT	3 ophthalmologists using grid	Not reported	Not reported	8%–20%	>0.2 DD 7%–28%
Klein and associates[11]	Full frame	200 control 186 diabetic	2 technicians using circles	0.77–0.88*	40 eyes 0.92–0.95*	24%–27%	Not reported
Klein and associates[19]	Full frame	408 diabetic 30 glaucoma	2 clinicians 1 technician using circles	V: 0.65–0.70* H: 0.67–0.84*	Not reported	Not reported	19 eyes 0%–5%
Tielsch and associates[2]	Full frame	666 control 265 OHT 118 glaucoma	2 glaucoma specialists	V: 0.88† H: 0.86†	214 eyes V: 0.94–0.95† H: 0.93†	V: 17% H: 19%	V: < 10% H: <15%
Varma and associates[3]	Split frame	8 normal 16 suspects 11 glaucoma	2 glaucoma specialists	V: 0.84 H: 0.78	V: 0.81–0.89 H: 0.77–0.89	V: 13% H: 6%	V: 3%–6% H: 2%–8%
Varma and associates[4]	Split frame	31 normal 29 glaucoma 15 unknown	6 glaucoma specialists	V: 0.67	V: 0.71–0.96	V: 19%	V: 0%–12%
Abrams and associates[18]	Split frame	29 glaucoma 31 control 15 unknown	6 optometrists 6 residents 6 ophthalmologists	V: 0.56 V: 0.56 V: 0.68	V: 0.69 V: 0.78 V: 0.79	V: 29% V: 28% V: 20%	V: 14% V: 9% V: 5%
Zangwill and associates[5]	Split frame	15 normal 15 glaucoma	3 glaucoma specialists	V: 0.46–0.79‡ H: 0.55–0.64‡	Not reported	V: 19% H: 24%	4 eyes V: 0% H: 8%
Shuttleworth and associates[20]	Digital full frame	98 normal 49 suspect 49 glaucoma or optic neuropathy	2 ophthalmologists	V: 0.90† H: 0.89†	V: 0.92† H: 0.94†	V: 3% H: 2%	V: 2% H: 0%
Harper and associates[21]	Split frame	15 glaucoma 22 suspicious 11 control	3 optometrists 2 ophthalmologists	V: 0.23–0.64 H: 0.20–0.64	V: 0.71–0.86 H: 0.70–0.84	Not reported	Not reported
Feuer and associates	Full frame and split frame	>2800 OHT substudy of 100 eyes for between observer agreement	2 technicians 3 glaucoma specialists§ 2 technicians 3 glaucoma specialists	H: 0.93† Yr1 H: 0.92† Yr2 H: 0.92† Yr3 H: 0.89†		H: 5% Yr1 H: 6% Yr2 H: 7% Yr3 H: 7%	

*Pearson correlation coefficient.
†Intraclass correlation coefficient.
‡Kappa weighted as described by Tielsch and associates[2] and Zangwill.[5]
§OHTS agreement is neither truly interobserver or intraobserver, since the estimation of C/D ratio is after consensus/adjudication as necessary.

range of horizontal cup/disk ratio represented in OHTS is broad with a substantial number of eyes with horizontal cup/disk ratios at the low range which is known to be more difficult to grade reliably.[3,20,21] Eleven percent of the eyes (335 of 2953 eyes) were estimated to have horizontal cup/disk ratios of 0.0 DD to 0.1 DD at entry. In this report, agreement for eyes with small horizontal cup/disk ratios did not appear to be lower than for eyes with larger horizontal cup/disk ratios. The OHTS sample consists entirely of participants with ocular hypertension in whom reproducibility of cup/disk ratio estimates has been reported to be lower than in eyes of glaucoma patients.[2] Thus, the high agreement reported in this study seems unlikely to be attributable to ocular characteristics of the sample that favor high interobserver agreement.

Some clinicians prefer to record vertical cup/disk ratio, largest cup/disk ratios, or both horizontal and vertical cup/disk ratios in glaucoma patients. These indices are highly correlated within individuals.[23] We chose to restrict this report to horizontal cup/disk ratio because it is the most time-honored measure and to report vertical cup/disk ratio with other disk features in future publications.

This report demonstrates that well-trained technician graders can make highly reproducible visual estimates of the horizontal cup/disk ratio and that a high volume of optic disk stereophotographs can be processed efficiently. These results suggest that horizontal cup/disk ratio measurements in the OHTS are sufficiently reproducible to provide useful information about the relationship of horizontal cup/disk ratio to the prognosis of individuals with ocular hypertension.

APPENDIX: OCULAR HYPERTENSION TREATMENT STUDY GROUP

Participating clinics, committees, and resource centers in the Ocular Hypertension Treatment Study as of March 1, 2001. Investigators and coordinators and staff are listed, respectively.

Clincial Centers

Bascom Palmer Eye Institute, University of Miami, Miami, Florida: Donald L. Budenz, MD,* Francisco E. Fantes, MD, Steven J. Gedde, MD, Richard K. Parrish II, MD; Madeline L. Del Calvo, BS.

M. Angela Vela, MD, PC, Atlanta, Georgia: M. Angela Vela, MD,* Thomas S. Harbin, Jr, MD, Paul McManus, MD, Charles J. Patorgis, OD, Ron Tilford, MD; Laura Brannon, Gail Degenhardt, Montana L. Hooper, COT, Stacey S. Goldstein, COMT, June M. LaSalle, COA, Debbie L. Lee, COT, Michelle D. Mondshein, Marianne L. Perry, COT, Ramona Weeden, Julie M. Wright, COT.

Cullen Eye Institute, Baylor College of Medicine, Houston, Texas: Ronald L. Gross, MD,* Silvia Orengo-Nania, MD; Pamela M. Frady, COMT, CCRC, Benita D. Slight, COT, EMT-P.

Devers Eye Institute, Portland, Oregon: George A, (Jack) Cioffi, MD,* Elizabeth Donohue, MD, Steven Mansberger, MD, E. Michael Van Buskirk, MD; Kathryn Sherman, JoAnne M. Fraser, COT.

Emory University Eye Center, Atlanta, Georgia: Allen D. Beck, MD,* Anastasias Costarides, MD; Donna Leef, MMSc, COMT, Jatinder Bansal, COT, David Jones, COT.

Henry Ford Medical Center, Troy, Michigan: G. Robert Lesser, MD,* Deborah Darnley-Fisch, MD, Monica Gibson, MD, Nauman R. Imami, MD, James Klein, MD, Talya Kupin, MD, Rheft Schiffman, MD; Melanie Gutkowski, COMT, CO, Jim Bryant, COT, Amanda Cole-Brown, Ingrid Crystal Fugmann, COMT, Wendy Gilroy, COMT, Norma Hollins, COT, RN, Sue Loomis, Melina Mazurk, COT, Colleen Wojtala.

Johns Hopkins University, School of Medicine, Baltimore, Maryland: Donald J. Zack, MD,* PhD, Donald A. Abrams, MD, Nathan G. Congdon, MD, Robert A. Copeland, MD, David S. Friedman, MD, Ramzi Hemady, MD, Eve J. Higginbotham, MD, Henry D. Jampel, MD, MHS, Omofolasade B. Kosoko, MD, Scott LaBorwit, MD, Stuart J. McKinnon, MD, PhD, Irvin P. Pollack, MD, Sreedhar V. Potarazu, MD, Harry A. Quigley, MD, Alan L. Robin, MD; Rachel Scott, BS, COA, Rani Kalsi, Felicia Keel, COA, Lisa Levin, Robyn Priest-Reed, MMSc.

Charles R. Drew University, Jules Stein Eye Institute, UCLA, Los Angeles, California: Anne L. Coleman, MD,* Richard S. Baker, MD, Luca O. Brigatti, MD, Y.P. Dang, MD, Simon K. Law, MD, Robert K. Stevens, MD; Jackie R. Sanguinet, BS, COT, Bobbie Ballenberg, COMT, Salvador Murillo, Manju Sharma.

W.K. Kellogg Eye Center, Ann Arbor, Michigan: Terry J. Bergstrom, MD,* Sayoko E. Moroi, MD, PhD; Carol J. Pollack-Rundle, BS, COT, Michelle A. Tehranisa, COA.

Kresge Eye Institute, Wayne State University, Detroit, Michigan: Bret A. Hughes, MD,* Mark S. Juzych, MD, Mark L. McDermott, MD, John M. O'Grady, MD, John M. Ramocki, MD, Stephen Y. Reed, MD, Dian Shi, MD, Dong H. Shin, MD, PhD; Juan Allen, Mary B. Hall, Laura L. Schulz, CNA, Linda A. Van Conett, COT.

University of Louisville, Louisville, Kentucky: Joern Soltau, MD,* Gustava E. Gamero, MD, Judit Ambrus, MD, Robert D. Fechtner, MD, Jianming X. Ren, MD, Robb Shrader, MD, Gil Sussman, MD, Thom Zimmerman, MD, PhD; Sandy Lear, RN, Kathleen Coons, COT.

Mayo Clinic/Foundation, Rochester, Minnesota: David C. Herman, MD,* Douglas H. Johnson, MD, Paul H. Kalina, MD; Becky A. Nielsen, LPN, Nancy J. Tvedt.

New York Eye & Ear Infirmary, New York, New York: Jeffrey M. Liebmann, MD,* Robert Ritch, MD, Robert F. Rothman, MD, Celso Tello, MD; Jean L. Walker, COA, Deborah L. Simon, COA.

Ohio State University, Columbus, Ohio: Robert J. Derick, MD,* N. Douglas Baker, MD, David Lehmann, MD, Omar Mobin-Uddin, MD, Paul Weber, MD; Kathyrne McKinney, COMT, Lori Black, Tammy Lauderbaugh, Diane Moore, COA.

Pennsylvania College of Optometry/Allegheny University of the Health Sciences, Philadelphia, Pennsylvania: G. Richard Bennett, MS, OD,* Elliot Werner, MD, Myron Yanoff, MD; Lindsay C. Bennett, BA, Mary Jameson, Opt, TR, Maria Massini.

Scheie Eye Institute, University of Pennsylvania, Philadelphia, Pennsylvania: Jody R. Piltz-Seymour, MD,* Debbie D. Curry, MD, Anna Purna Singh, MD; Jane L. Anderson, MS, Cheryl McGill, Janice T. Petner, COA.

University of California-Davis, Sacramento, California: James D. Brandt, MD,* Craig Bindi, MD, Jeffrey J. Casper, MD, John T. Dragievich, MD, Janet Han, MD, Denise Kayser, MD, Sooyung Kim, MD, Michelle C. Lim, MD, Michael B. Mizoguchi, MD, Alan M. Roth, MD, Ivan R.

Schwab, MD; Ingrid J. Clark, COA, Vachiraporn X. Jaicheun, COA, Denise M. Owensby, BS, COA, Marilyn A. Sponzo.

University of California-San Diego, La Jolla, California: Robert N. Weinreb, MD,* J. Rigby Slight, MD; Eva Kroneker, Barbara Brunet.

University of California-San Francisco, San Francisco, California: Michael V. Drake, MD,* Allan J. Flach, MD, Robert Stamper, MD; Fermin Ballesteros, Valerie Margol, Ilya Saltykov, Peggy Yamada, COT.

University Suburban Health Center, South Euclid, Ohio: Kathleen A. Lamping, MD,* Laurence D. Kaye, MD; Angela K. McKean, Tonya Sims, Susan Van Huss.

Washington OHTS Center, Washington, District of Columbia: Douglas E. Gaasterland, MD,* Frank S. Ashburn, MD, Arthur Schwartz, MD, Howard S. Weiss, MD; Anne M. Boeckl, MS, Robin Montgomery, Donna Claggett, Deanne Griffin, Karen D. Schacht, COT.

Washington University School of Medicine, St. Louis, Missouri: Martin B. Wax, MD,* David C. Ball, MD, Michael A. Kass, MD, Allan E. Kolker, MD, Carla J. Siegfried, MD, Jonathan Silbert, MD; Arnold D. Jones, COA, Lori A. Clark, COT, Fortunata Darmody, COT, Diana L. Moellering, COT.

Committees

Executive/Steering Committee: Douglas R. Anderson, MD, James D. Brandt, MD, Donald F. Everett, MA, Douglas Gaasterland, MD, Mae E. Gordon, PhD, Dale K. Heuer, MD, Eve J. Higginbotham, MD, Chris A. Johnson, PhD, Michael A. Kass, MD, John L. Keltner, MD, Richard K. Parrish II, MD, Arthur Shedden, MD, M. Roy Wilson, MD; Jane Anderson, MS, Patricia A. Morris, Ann K. Wilder, RN, BSN.

Data and Safety Monitoring Committee: Roy Beck, MD, PhD, John Connett, PhD, Claude Cowan, MD, Barry Davis, MD, PhD (Chair), Donald F. Everett, MA (nonvoting), Mae O. Gordon, PhD (nonvoting), Michael A. Kass, MD (nonvoting), Ronald Munson, PhD, Ingrid Adamsons, MD (nonvoting), Mark Sherwood, MD, Gregory L. Skuta, MD.

Endpoint Committee: Dale Heuer, MD, Eve Higginbotham, MD, Richard K. Parrish II, MD, Mae O. Gordon, PhD.

Resource Centers

Coordinating Center-Washington University School of Medicine, St. Louis, Missouri: Mae O. Gordon,* PhD, J. Philip Miller; Joel Achtenberg, MSW, Mary Bednarski, MAS, Julia Beiser, MS, Karen Clark, Christopher Ewing, Ellen Long, CCRA, Patricia Morris, Denise Randant, Ann K. Wilder, RN, BSN.

Chairman's Office-Washington University School of Medicine, St. Louis, Missouri: Michael A. Kass, MD;* Deborah Dunn, Carolyn Miles.

Project Office, National Eye Institute, Rockville, Maryland: Donald F. Everett, MA.

Optic Disc Reading Center, Bascom Palmer Eye Institute, University of Miami, Miami, Florida: Richard K. Parrish II,* MD, Douglas R. Anderson, MD, Donald L. Budenz, MD; Maria-Cristina Wells, MPH, William Feuer, MS, Ditte Hess, CRA, Heather Johnson, Joyce Schiffman, MS, Ruth Vandenbroucke.

Visual Field Reading Center, University of California, Davis, Sacramento, California (J.L.K.); Discoveries in Sight, Devers Eye Institute, Portland, Oregon (C.A.J.): John L. Keltner, MD,* Chris A. Johnson, PhD; Kimberly E. Cello, BS, Shannan E. Banderman, MA, Bhupinder S. Dhillon, BSc, Mary A. Edwards, BS.

Ancillary Study Reading Centers

Confocal Scanning Laser Ophthalmoscopy Reading Center, University of California-San Diego, La Jolla, California: Robert N. Weinreb, MD,* Linda Zangwill, PhD; Keri Dirkes, MPH.

Short Wave Length Automated Perimetry Reading Center, Devers Eye Institute, Legacy Portland Hospitals, Portland, Oregon: Chris A. Johnson, PhD;* Erna Hibbitts.

Corneal Endothelial Cell Density Reading Center, Mayo Clinic/Foundation, Rochester, Minnesota: William M. Bourne, MD;* Becky Nielsen, LPN, Thomas P. Link, CRA, BA, Jay A. Rostvold.

* = principal investigator.

REFERENCES

1. Lichter P. Variability of expert observers in evaluating the optic disc. Trans Am Ophthalmol Soc 1976;74:532–572.
2. Tielsch J, Katz J, Quigley HA, Miller NR, Sommer A. Intraobserver and interobserver agreement in measurement of optic disc characteristics. Ophthalmology 1988;95:350–356.
3. Varma R, Spaeth GL, Steinmann WC, Katz LJ. Agreement between clinicians and an image analyzer in estimating cup-to-disc ratios. Arch Ophthalmol 1989;107:526–529.
4. Varma R, Steinman WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology 1992;99:215–221.
5. Zangwill L, Shakiba S, Caprioli J, Weinreb RN. Agreement between clinicians and confocal scanning laser ophthalmoscope in estimating cup/disc ratios. Am J Ophthalmol 1995; 199:415–421.
6. Gordon MO, Kass MA, for the Ocular Hypertension Treatment Study Group. The ocular hypertension treatment study: design and baseline description of the participants. Arch Ophthalmol 1999;117:573–583.
7. Early Treatment Diabetic Retinopathy Study Research Group. Photocoagulation for diabetic macular edema: Early Treatment Diabetic Retinopathy Study Report Number 1. Arch Ophthalmol 1985;103:1796–1806.
8. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: One-year outcome - Structure and function. Arch Ophthalmol 1990;108:1408–1416.

9. The Glaucoma Laser Trial Research Group. The Glaucoma Laser Trial (GLT): 2. Results of Argon Laser Trabeculoplasty versus Topical Medicines. Ophthalmology 1990;97:1403–1413.
10. Central Vein Occlusion Study Group. Baseline and early natural history report: the Central Vein Occlusion Study. Arch Ophthalmol 1993;111:1087–1095.
11. Klein BEK, Magli YL, Richie KA, Moss SE, Meuer SM, Klein R. Quantitation of optic disc cupping. Ophthalmology 1985; 92:1654–1656.
12. Gordon MO, Kass MA. The Ocular Hypertension Study Group (OHTS). Manual of Procedures. Washington, DC: National Technical Information Service; 1997; Publication PB97–148308NZ.
13. MacLure M, Willet WC. Misinterpretation and misuse of the kappa statistic. Am J Epidemiol 1987;126:161–169.
14. Shrout R, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull 1979;86:420–428.
15. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13:2465–2476.
16. Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley and Sons; 1986: pp. 1–28 and 193.
17. Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York: John Wiley and Sons; 1973: p. 218.
18. Abrams LS, Scott IU, Spaeth GL, Quigley HA, Varma R. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. Ophthalmology 1994;101:1662–1667.
19. Klein BEK, Moss SE, Magli YL, Klein R, Johnson JC, Roth H. Optic disc cupping as clinically estimated from photographs. Ophthalmology 1987;94:1481–1483.
20. Shuttleworth GN, Khong CH, Diamond JP. A new digital optic disc stereo camera: intraobserver and interobserver repeatability of optic disc measurements. Br J Ophthalmol 2000; 84:403–407.
21. Harper R, Reeves B, Smith G. Observer variability in optic disc assessment: implications for glaucoma shared care. Ophthal Physiol Opt 2000; 20:265–273.
22. Hitchings RA, Genio C, Anderton S, Clark P. An optic disc grid: its evaluation in reproducibility studies on the cup/disc ratio. Br J Ophthalmol 1983;67:356–361.
23. Leibowitz HM, Krueger DE, Maunder LR, et al. The Framingham Eye Study Monograph: An ophthalmological and epidemiological study of cataract, glaucoma, diabetic retinopathy, macular degeneration, and visual acuity in a general population of 2631 adults. Surv Ophthalmol 1980;24(suppl):335–610.