THE ADVANCED GLAUCOMA INTERVENTION STUDY (AGIS): 10. VARIABILITY AMONG ACADEMIC GLAUCOMA SUBSPECIALISTS IN ASSESSING OPTIC DISC NOTCHING*

BY *Douglas E. Gaasterland, MD, Beth Blackwell, ScD (*BY INVITATION*), Leonard G. Dally, MSc (*BY INVITATION*), Joseph Caprioli, MD, L. Jay Katz, MD (*BY INVITATION*), Fred Ederer, MA, FACE (*BY INVITATION*)* AND *The AGIS Investigators (*SOME BY INVITATION*)*

ABSTRACT

Purpose: An analysis of data from the Advanced Glaucoma Intervention Study (AGIS) has found eyes reported to have partial optic disc rim notching (not to the edge) at baseline to have less risk of subsequent visual field loss than eyes with no notching. Because this is counterintuitive and because classification of notching had not been defined in the AGIS protocol, we have assessed AGIS ophthalmologists interobserver and intraobserver agreement on notching.

Methods: Fourteen glaucoma subspecialists classified notching in 26 pairs of stereoscopic disc photographs of eyes with mild to severe glaucomatous optic neuropathy. They classified images as showing either no notching, notching not to the edge, or notching to the edge. Several hours later, 10 of them classified the same images a second time.

Results: In an analysis of interobserver agreement, of 26 stereoscopic images, a plurality of ophthalmologists classified notching as absent in 9 (35%), as present but not to the edge in 7 (27%), and as present and to the edge in 10 (38%). All 14 ophthalmologists (100%) agreed on the classification of 7 (27%) of the images, and 13 of the 14 ophthalmologists (93%) agreed on the classification of 4 additional images (15%). Of these 11 images with at least 93% agreement, notching was reported as absent in 3 (27%) and to the edge in 8 (73%). In the remaining 15 images, there was substantial disagreement about whether notching was present and, if so, whether it was to the edge. In an analysis of intraobserver agreement, none of the 10 ophthalmologists who completed the viewing a second time classified all eyes exactly the same as the first time, though 5 ophthalmologists made 4 or fewer reclassifications. Overall, 80% of the original classifications were reproduced on second reading. Of the initial classifications that were not reproduced, slightly more than half were first classified as having notching not to the edge.

Conclusion: Without definitions or examples of optic disc rim notching, the glaucoma subspecialists had relatively high intraobserver agreement but were likely to disagree with each other in characterizing the degree of disc rim notching. We recommend development of a standard photographic classification of disc rim notching. The classification should be tested for inter- and intra-observer agreement.

Tr Am Ophth Soc 2001;99:177-185

INTRODUCTION

When protocol development and planning for the Advanced Glaucoma Intervention Study (AGIS) started in 1986, the investigators recognized the importance of optic disc analysis in glaucoma diagnosis and monitoring. In order to reduce study cost and complexity, however, they accepted a recommendation to eliminate stereoscopic optic disc photography and an optic disc photograph reading center from the study plan. Instead, the investigators decided to use clinical stereoscopic slit-lamp biomicroscopic examinations to evaluate optic disc rim characteristics, including a determination as to whether there was notching of the neural rim of the optic disc and, if so, whether it extended only partially or completely to the edge of the disc.

In an ongoing data analysis to determine which baseline characteristics are predictive of subsequent deterioration of visual function, we find that eyes in one, but not the other, randomly assigned surgical treatment sequence and reported to have partial disc rim notching (not to the

*From the Department of Ophthalmology, Georgetown University School of Medicine, Washington, DC (Dr Gaasterland); the EMMES Corporation, Potomac, Maryland (Dr Blackwell, Mr Dally, and Mr Ederer); Jules Stein Eye Institute, University of California, Los Angeles (Dr Caprioli); and Wills Eye Hospital, Philadelphia, Pennsylvania (Dr Katz). A complete list of the AGIS investigators and their participating institutions appears at the end of this manuscript. Supported by grants 2 U10 EY06824 through 2 U10 EY06827, 2 U10 EY06830 through 2 U10 EY06835, 2 U10 EY07057, and 7 U10 EY09640 from the National Eye Institute and the Office of Research on Minority Health, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland.

edge) at baseline have a significantly lower rate of subsequent visual field loss than eyes reported to have no notching. This counterintuitive result has prompted us to question the consistency of and agreement among AGIS glaucoma subspecialists in classifying this optic disc characteristic.

There have been several interesting studies of the reproducibility of determining optic disc cupping.[1-6] We have found no reports of studies of reproducibility of determining optic disc rim notching.

METHODS

OVERALL AGIS METHODS

The AGIS protocol and the baseline characteristics of the 591 subjects enrolled in AGIS from 11 (later 12) participating clinical centers are described in detail elsewhere.[7] We briefly summarize the AGIS methods here. Appropriate institutional review boards approved the AGIS protocol, and all enrolled patients provided informed consent.

To be eligible for AGIS, patients had to be between 35 and 80 years old and have open-angle glaucoma that could no longer be controlled by medications alone. Additionally, the eye, while on maximum tolerated and effective medical treatment, had to meet one of several combinations of intraocular pressure (IOP) and visual field defect score criteria. One of the combinations that established eligibility of an eye was a glaucomatous visual field defect and consistently elevated IOP of 18 mm Hg or greater and deterioration of the optic disc neural rim. Disc rim deterioration was defined as occurrence of one or more of the following: development of disc rim hemorrhage, decrease in rim width of at least 50% in any one location, development of a notch to the edge of the disc, or increase of horizontal or vertical cup-disc ratio of 0.2 or greater. At the baseline slit-lamp biomicroscopic fundus examination, the AGIS ophthalmologist determined horizontal and vertical cup-disc ratio (from the neural rim widths in the horizontal and vertical meridians), location of the thinnest rim, hemorrhage(s) on rim tissue, and notching. If there was notching, the examiner reported whether it was to the edge of the disc or not. No definitions or guidelines for the three-category classification of notching were provided.

METHODS OF THE PRESENT STUDY

Expert academic glaucoma subspecialist ophthalmologists at 2 centers provided high-quality stereoscopic photographs of optic discs of glaucoma patients. One set of images, from the Jules Stein Eye Institute, Los Angeles, California, consisted of 13 pairs of sequential 35 mm full-frame transparencies; the other set, from the Wills Eye Hospital, Philadelphia, Pennsylvania, consisted of 13 pairs of simultaneous 35 mm split-frame photographs. The photographs were of eyes judged by the providers to have mild to severe glaucomatous optic neuropathy and no notching, notching to the edge of the disc, or notching not to the edge. Based on the providers' appraisal of notching, one of us (D.E.G.) arrayed the photographs in each set in a disarranged order of notching.

During a 1-day meeting of the full group of AGIS investigators in September 2000, 14 academic glaucoma subspecialists each viewed and classified the 26 stereoscopic pairs of photographs. In the morning, without provision of notching definitions or discussion of definitions, each ophthalmologist viewed the 26 paired stereoscopic disc images in the disarranged order of notching and recorded his or her assessment of each optic disc as having no notching, notching not to the edge, or notching to the edge. In the afternoon of the same day, 10 of the observers, masked as to their previous responses, viewed the 26 paired images in the same order and recorded their assessments a second time.

RESULTS

INTEROBSERVER AGREEMENT

Agreement among all 14 ophthalmologists occurred for only 7 of the 26 paired stereoscopic disc images (27%) (Table I); all agreed that 3 images showed no notching (an example is shown in Fig 1) and that 4 showed notching to the edge (an example is shown in Fig 2). Agreement among 13 or 14 observers (93% to 100% concordance)

TABLE I: INTEROBSERVER AGREEMENT: ASSESSMENTS FROM 14 GLAUCOMA SUBSPECIALIST OPHTHALMOLOGISTS OF OPTIC DISC NOTCHING IN STEREOSCOPIC DISC PHOTOGRAPHIC IMAGES OF 26 GLAUCOMATOUS EYES

NO. OF DISC IMAGES	DEGREE OF NOTCHING REPORTED		
	NONE	NOT TO EDGE	TO EDGE
3	14	0	0
1	12	2	0
1	10	4	0
1	10	3	1
2	9	5	0
1	9	3	2
1	4	8	2
1	3	8	3
1	3	7	4
1	3	3	8
1	2	10	2
1	1	3	10
1	1	0	13
2	0	9	5
1	0	8	6
3	0	1	13
4	0	0	14
Total **26**	**118**	**90**	**156**

occurred for 11 images, with the observers agreeing that 3 images showed no notching and 8 showed notching to the edge of the disc.

In their classifications of the 26 stereoscopic images, a majority of the ophthalmologists agreed for 25 images. The size of the majority was less than two thirds for 9 of the images. The majority reported no notching in 9 images, notching not to the edge in 6, and notching to the edge of the disc in 10 (Table I). For 1 image, half the observers reported that notching was not to the edge, 3 observers reported no notching, and 4 reported notching to the edge (Fig 3).

Substantial numbers of dissimilar responses occurred both in images that the majority classified as having no notching and in images the majority classified as having notching to varying extent. In fact, for 4 of the images, the size of the majority was only 57% (8/14) of the ophthalmologists. For example, for 1 image, 8 observers reported notching not to the edge while 3 reported no notching and 3 reported notching to the edge (Fig 4).

INTRAOBSERVER AGREEMENT

Although none of the 10 ophthalmologists who completed a second evaluation of the 26 stereoscopic images demonstrated perfect agreement with his or her first classification, 5 classified at least 22 (85%) of the 26 images the same each time. Of the 5, 1 ophthalmologist classified only 1 image differently at the second viewing and 4 classified 2 to 4 images differently.

Table II shows the results of the first classifications for each of the 10 ophthalmologists and the changes in classification on the second reading. Out of 260 chances (10 ophthalmologists and 26 images) for intraobserver agreement, there is agreement in 209 (80%). Of the 51 paired observations that differed, 27 (53%) were in the images classified at first assessment as having notching not to the edge, while the remainder were divided almost equally between those classified during first assessment as having no notching and those as having notching to the edge. There was 87% (72/83) intraobserver agreement on images classified as having no notching on first assessment, 56% (34/61) agreement on having notching not to the edge on first assessment, and 89% (103/116) agreement on having notching to the edge on first assessment (Table II).

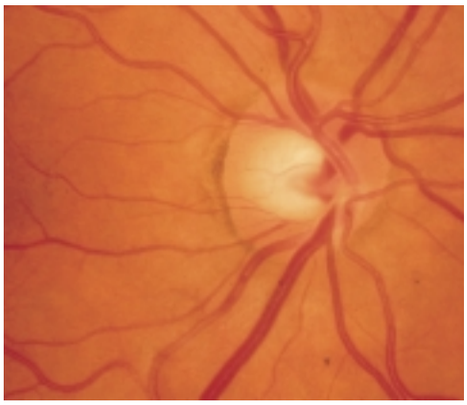
DISCUSSION

In this study of agreement among AGIS glaucoma subspecialist ophthalmologists in identifying optic disc rim notching in stereoscopic photographic images of optic discs, we find that for 11 of the 26 images presented there is substantial agreement between observers about the degree of notching (Table I). For 12 images, the ophthalmologists widely disagreed as to whether there was no notching, notching not to the edge, or notching to the edge. By contrast, when the observers classified the images a second time, a large majority (80%) of the second classifications were in agreement with the first classification (Table II). This indicates that the ophthalmologists participating in this study generally have consistent personal definitions of the 3 degrees of optic disc rim notching, but that they often differ between one another about the definitions.

Agreement among ophthalmologists in a study like this depends not only on the choice of photographs but also on the clarity and detail of instructions provided to the readers prior to assessment. Interestingly, in 1 study of optic disc cupping, even when efforts were made to provide clear definitions and instructions, along with photographic examples, a high level of disagreement continued.[1] Because the AGIS protocol did not provide definitions or guidelines on how to classify optic disc rim notching during clinical stereoscopic slit-lamp biomicroscopic fundus examination, we implemented the present study of stereoscopic photographs without instruction or guidelines on how to classify notching into the three-category response requested on AGIS data forms.[8]

Several aspects of the current study design may have caused an artefact in the intraobserver reproducibility. First, one participating ophthalmologist (D.E.G.) had earlier disarranged the images in the 2 sets based on the classification provided by the source ophthalmologists. Repeating the analysis after removing the assessments of this ophthalmologist (D.E.G.) had little effect on the results (data not shown). Second, with the interval between the first and second classifications only a few hours and with no rearrangement of the order of the images in the sets, it is possible that some readers remembered some of their first assessments.

For years, most ophthalmologists have recognized disc rim notching (focal rim thinning) as a sign of glaucoma damage. For example, in a seminal study, Hitchings and Spaeth[9] included focal notching of the neuroretinal rim as 1 of 5 morphologic types of change found in glaucomatous optic discs. Yet, of 4 current glaucoma textbooks,[10-13] notching is listed in the index of only 1,[10] and a description of disc rim notching is provided in only 2.[10,11] In planning AGIS, the investigators assumed that all participating ophthalmologists shared a clear ability to recognize and characterize this sign of glaucomatous optic neuropathy. According to their examinations, at study baseline 196 of 770 enrolled eyes (25%) had optic disc rim notching to the edge and another 79 eyes (11%) had notching not extending to the edge of the disc. On the basis of the results of the present study, we question the reliability of the classifications, particularly for the 79 eyes reported as having notching not to



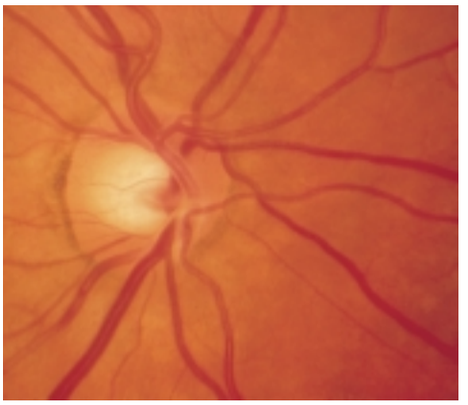


FIGURE 1A AND 1B

Stereoscopic pair of sequential full-frame photographs showing an optic disc characterized by all of 14 expert observers as having no notching of neural rim.

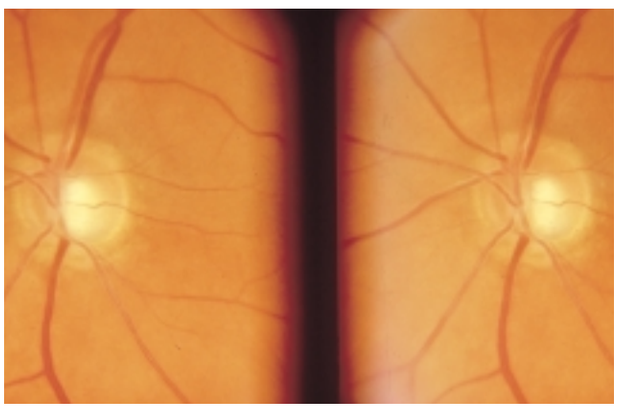


FIGURE 2

Stereoscopic pair of simultaneous split-frame photographs showing an optic disc characterized by all of 14 expert observers as having notching of neural rim to edge of disc.

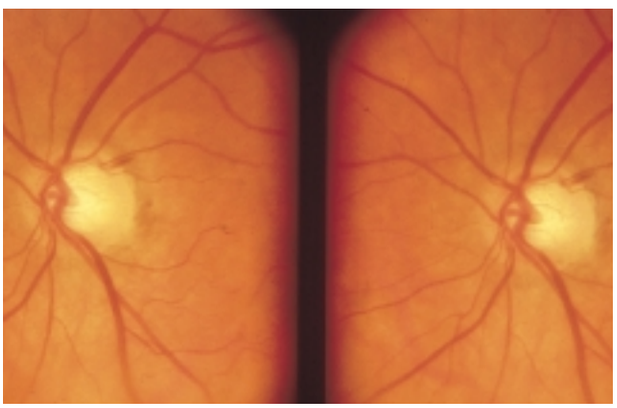


FIGURE 3

Stereoscopic pair of simultaneous split-frame photographs showing an optic disc characterized by 7 of 14 expert observers as having notching not extending to edge of disc, by 3 others as having no notching, and by another 4 as having notching to edge of disc. There is a disc rim hemorrhage at the 1:30-o'clock meridian.

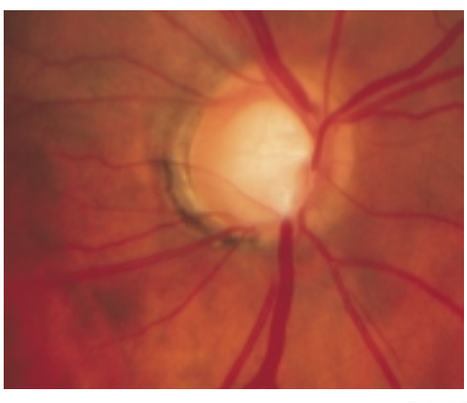


FIGURE 4A AND 4B

Stereoscopic pair of sequential full-frame photographs showing an optic disc characterized by 8 of 14 expert observers as having notching not extending to edge of disc, by 3 others as having no notching, and by another 3 as having notching of neural rim to edge of disc. A wisp of vitreous condensation extends off nasal rim in this eye. The first classifications by 10 expert observers of notching in this image often differed from the second.

TABLE II: INTRAOBSERVER AGREEMENT: CLASSIFICATIONS OF OPTIC DISC NOTCHING AND CHANGES IN CLASSIFICATIONS DURING SECOND ASSESSMENT OF STEREOSCOPIC DISC PHOTOGRAPHS OF 26 GLAUCOMATOUS EYES BY 10 GLAUCOMA SUBSPECIALIST OPHTHALMOLOGISTS*

		FIRST ASSESSMENT OF NOTCHING								
		NONE			NOT TO EDGE			TO EDGE		
		SECOND ASSESSMENT OF NOTCHING								
OBSERVER	NO. OF CHANGES	NONE	NOT TO EDGE	TO EDGE	NONE	NOT TO EDGE	TO EDGE	NONE	NOT TO EDGE	TO EDGE
1	1	7	1			5				13
2	2	11	1			5			1	8
3	3	10				2	1		2	11
4	4	11	1	1		2		2		9
5	4	4	1		1	6	2			12
6	5	8	2			3	3			10
7	7	8	2		1	4	2	1	1	7
8	7	5	1		1	4	4		1	10
9	7	4			4	3	1		2	12
10	11	4		1	5		2	3		11
Changes:	51		9	2	12		15	6	7	
No changes:		72				34				103

*Several hours after their first assessment, ophthalmologists, masked as to first responses, again viewed and classified same image sets with images in same order as during first assessment.
Note: Shaded columns have no change in second assessment compared with first assessment.

the edge of the disc. The 7 images with the largest number of interobserver and intraobserver disagreements in the present study are the 7 classified by at least half of the ophthalmologists as having notching not to the edge.

We recognize that in this study some agreement could have occurred by chance. The kappa statistic, which has been used to assess the degree of concordance in multi-observer studies, measures the amount of agreement beyond what is expected to occur by chance. We calculated kappa (κ) for the present study, finding κ = 0.51 (P <.001) for inter-observer agreement among the 14 ophthalmologists who classified the 26 images. For intraobserver agreement, we found κ to range from 0.36 to 0.94 (mean, 0.69). In recognition that the amount of agreement in the present study may be affected by the composition of the photographic sets, and because, as Siegel and coauthors stated,[14]. "…by itself, kappa is not informative enough to evaluate the appropriateness of a grading scheme for comparative studies," we decided not to emphasize the kappa statistic in this report.

Similar to the previous findings on optic disc cupping by Kahn and associates,[1] Lichter,[2] and other investigators,[3-6] we have found substantial variation among expert observers in classifying notching of the neural rim of the optic disc. It is reassuring that individual ophthalmologists in this study tended to be consistent in classifying the images.

This leads to some recommendations for future research in optic disc rim notching. A critical step is to develop a photographic classification of notching along the lines of the Airlie House Classification of diabetic retinopathy[15] or the LOCS II classification of cataract.[16] The classification, to be based on standard photographs that illustrate or set limits on various degrees of notching, should be tested for intraobserver and interobserver agreement and then modified as needed. The standard photographs should be stereoscopic, taken by either the sequential full-frame or simultaneous split-frame technique. If good reproducibility is achieved, the classification can be applied in single center or multicenter studies in which notching of individual optic discs is assessed either at a disc photograph reading center or by ophthalmologists sitting at slit-lamp biomicroscopes. A good classification could become widely adopted and might provide a basis for comparisons with results from retinal nerve fiber layer analyzers.

APPENDIX

AGIS CENTERS AND INVESTIGATORS: PARTICIPATING INSTITUTIONS, CURRENT INVESTIGATORS, AND FORMER INVESTIGATORS WHO PARTICIPATED FOR 2 OR MORE YEARS

STUDY CO-CHAIRMEN

Douglas E. Gaasterland, MD; Fred Ederer, MA, FACE

CLINICAL CENTERS

Abbreviations: CD, Clinic Director; CI, Co-Investigator; CC, Clinic Coordinator; CM, Clinic Monitor; SD, Satellite Director; SC, Satellite Coordinator; T, Technician.

Emory University, Atlanta: Allen Beck, MD (CD); Anastasios Costarides, MD (CI); Donna Leef, MMSc, CO, COMT (CC)(CM); John Closek, COT (T); Juanita Banks (T); Sheena Jackson (T); Kathy Moore (T). Past participating personnel: Angela Vela, MD (CD); Reay H. Brown, MD (CD); Mary Lynch, MD (CI); Johnny Gunsby, COT; Kathy Lober, COA; Twyla Marsh, COA; Candace Stepka

Georgetown University, Washington, DC: Douglas E. Gaasterland, MD (CD); Robin Montgomery, COA (CC)(CM); Donna Clagett, COA (T). Satellite facility: Frank Ashburn, MD (SD); Karen Schacht, COT (SC). Past participating personnel: Ellen Coyle, COMT (CC); Melissa Kellogg Garland, COA (CC); Susan Lauber, MA (CC); Karl Michelitsch, COMT (CC) (deceased); Suzanne Plavnieks, COT (CC); Lynn Vayer, COT (SC); Elizabeth Burt, COT; Mary Hundley, COT; Anne Rae, COT

Medical College of Virginia, Richmond: Robert C. Allen, MD (CD); Eydie Miller, MD (CI); Amy Sporn, OD (CC)(CM). Past participating personnel: C. Kay Fendley, COT (CC); L. Sharon Hoyle, COMT

Ohio State University, Columbus: Paul A. Weber, MD (CD); Kathryne McKinney, COMT (CC)(CM); Diane Moore, COA (T). Satellite facility: N. Douglas Baker, MD (SD); Fred Kapetansky, MD (CI); David Lehmann, MD (CI); Tammy Lauderbaugh (T). Past participating personnel: Robert Derick, MD (CI); Becky Gloeckner, COT (SC); Lori Black, COA (SC); Kris Coleman, COT; Mary Cassady, COA; Lisa J. Sharf, COA; Billi Romans; Yvonne Satterwhite; Lori Simmons

Piedmont Hospital, Atlanta: M. Angela Vela, MD (CD); Thomas S. Harbin Jr., MD (CI); Laura Brannon, COMT (CC)(CM); June LaSalle, COA (T); Gail Degenhardt (T); Stephanie Ann Bridgman (T). Past participating personnel: Randall R. Ozment, MD (CI); Johnny Gunsby, COT (T); Montana Hooper, COT (CC); Julie Wright, COT (T); Stacy Goldstein, COMT (SC); Linda Butler, COT; Marianne Perry, COT; Anne Eckel, COA; Anja Martin, COA; Celeste Session, COA; Dana Nummerdor; Lisa Wille

Oakland University/Detroit, Southfield, Michigan: Marshall N. Cyrlin, MD (CD); Holly Dubay (CC)(CM). Past participating personnel: Roselyn Fazio, BS, COT (CC); Patricia S. Corbin (CC)

University of Illinois, Chicago: Jacob T. Wilensky, MD (CD); Kim Lindenmuth, MD (CI); David Hillman, MD (CI); Catherine A. Carroll, COMT (CC)(CM); Jennifer Hatton, COT (T). Satellite facility: Sriram Sonty, MD (SD); Catherine A. Carroll, COMT (SC). Past participating personnel: Eve J. Higginbotham, MD (CD); Gary Scholes, MD (CI); Rosanna Uva, COT (CC); Julie Fiene, COT; Diane Frohlichstein, COT; Valeria Gates, COT; Loreen Pappas, COT; Donna Rathbone, COT; Marlem Tadelman, COT, Gloria Hopkins, LPN

University of Michigan, Ann Arbor: Paul R. Lichter, MD (CD); Terry J. Bergstrom, MD (CI); Sayoko E. Moroi, MD, PhD (CI); Carol J. Pollack-Rundle, COMT (Co-CC)(T); Carol Standardi, RN, CRNO (Co-CC)(CM); Lynette Abt, COT (T); Terri Van Heck, COT (T). Past participating personnel: Gregory L. Skuta, MD (CD); Eve J. Higginbotham, MD (CD); Robert M. Schertzer, MD (CI); Donna Wicker, OD; Barbara Michael, COT; Desiree Aaron, COA; Judith Birk, COA; Rebecca S. Brown, COA; Joyce Dederian, COA; Linda Kruscke, COA; Jennifer Ziehm-Scott, COA; Renee Papierniak-Dubiel

University of Virginia, Charlottesville: Bruce E. Prum Jr, MD (CD); Steven A. Newman, MD (CI); Lee Powell, COT (CC)(CM); Christine Evans, COMT (T); Nancy Barbour, COA (T); Lil Shoffstall-Tyler, COA (T); Terri Voight, COA (T). Past participating personnel: John R. Nordlund, MD, PhD (CD); Louis J. Schott, MD (CI); Robert Fornili, OD (CC); James Chisholm, COA; Carolyn Harrell, COA; Christi Harris, COA; Ellen Murphy, COA

Washington Hospital Center, Chevy Chase, Maryland: Arthur L. Schwartz, MD (CD); Howard Weiss, MD (CI); Anne Boeckl, MS (CC)(CM); Christine Tillman, (T). Satellite facility: Anne Boeckl, MS (SC); Lois Maloney, COT (T). Past participating personnel: Maria Cirone, MD (CI); John Gurley, MD (CI); Mark Morris, MD (CI); Maureen O'Dea, MD (CI); Stephen Pappas, MD (CI); Scott Wehrly, MD (CI); Cathy Reed, COMT (CC); Cindy V. Witol, CO (CC); Janet Browning, COT; Karen Carmody, COT; Teresa Driskell; Elaine Harris; Patrick Lopez, COT; Richard Mercer; Victoria Monks, COA; Kathy Vawter, COA; Jing Zhao (T)

Wills Eye Hospital, Philadelphia: L. Jay Katz, MD (CD); George L. Spaeth, MD (CI); Richard P. Wilson, MD (CI); Jonathan Myers, MD (CI); Fillis Samuel, COT (CC)(CM); Alaine Meli (T). Past participating personnel: Sue Kao, MD (CI); Annette Terebuh, MD (CI); Coleen C. Beckershoff

Yale University, New Haven: M. Bruce Shields, MD (CD); George Shafranov, MD (CI); Ann Leone, COT (CC)(CM); Gail Grottole, COA (T). Past participating personnel: Eydie Miller, MD (CD); Joseph Caprioli, MD (CD); Charles Tressler, MD (CI); Maureen Roche-Manna, COMT (CC)

COORDINATING CENTER

The EMMES Corporation, Rockville, Maryland: Paul C. VanVeldhuisen, MS (Director, Coordinating Center); Fred Ederer, MA, FACE (Epidemiologist); Len G. Dally, MSc (Statistician); Beth Blackwell, ScD (Statistician); Pam Inman (Data Manager); Susan Raitt, MA (Administrative Coordinator). Past participating personnel: Anne S. Lindblad, PhD (Deputy Director); James D. Knoke, PhD (Deputy Director), E. Kenneth Sullivan, PhD, (Director); Lie-Ling Wu, MS (Statistician); Gary Entler, COT (Protocol Monitor); Pamela Phillips, MHS (Protocol Monitor); Carol Smith, MPH (Protocol Monitor), Marline Bradford (Data Manager), Marsha Denekas, MLT (Data Coordinator); Elizabeth L. Wagner, MPH (Data Coordinator); Elaine Stine (Alternate Interviewer); Katherine L. Tomlin, MA (Administrative Coordinator); Tamara Voss, BA (Administrative Coordinator)

SPONSOR

National Eye Institute and The Office of Research on Minority Health, National Institutes of Health, Bethesda, Maryland: Mary Frances Cotch, PhD (NEI Representative); Lois J. Eggers (Grants Management Specialist). Past participating personnel: Richard L. Mowery, PhD (NEI Representative); Frances Goff (Grants Management Specialist); Carolyn Grimes (Grants Management Specialist); Gaye Lynch (Grants Management Specialist)

STUDY GROUPS

Policy and Treatment Effects Monitoring Board (PATEMB): Curt D. Furberg, MD, PhD (Chairman); John E. Connett, PhD; Matthew D. Davis, MD; David K. Dueker, MD; Sylvan B. Green, MD; Paul F. Palmberg, MD, PhD; Ex officio members: Fred Ederer, MA, FACE; Douglas E. Gaasterland, MD; Mary Frances Cotch, PhD. Past participating personnel: Sanford Leikin, MD; Marvin Schneiderman, PhD (deceased); The Rev Canon Michael P. Hamilton; Richard L. Mowery, PhD (Ex officio)

Operations Committee: Fred Ederer, MA, FACE; Douglas E. Gaasterland, MD; Paul C. VanVeldhuisen, MS; Arthur L. Schwartz, MD (consultant). Past participating personnel: Anne S. Lindblad, PhD; James D. Knoke, PhD; E. Kenneth Sullivan, PhD

Steering Committee: Permanent members: Fred Ederer, MA, FACE; Douglas E. Gaasterland, MD; Arthur L. Schwartz, MD; Aaron Kassoff, MD (consultant); Mary Frances Cotch, PhD. Elected investigator members: Marshall N. Cyrlin, MD (1988-1990); M. Angela Vela, MD (1988-1989); Joseph Caprioli, MD (1988-1990); Eve J. Higginbotham, MD (1989-1991); Gregory L. Skuta, MD (1990-1992); L. Jay Katz, MD (1991-1993); Paul A. Weber, MD (1992-1994); Robert C. Allen, MD (1993-1995); Jacob T. Wilensky, MD (1994-1996); Eydie Miller, MD (1995-1997); Reay Brown, MD (1996-1998); Bruce Prum, MD (1997-1999); Marshall Cyrlin (1998-2000); Allen Beck (1999-2001); Paul Lichter, MD (2000-2002). Elected Clinic Coordinator members: Donna Leef, MMSc, COMT (1988); Kathryne McKinney, COMT (1989); Cathy Reed, COMT (1990); Rosanna Uva, COT (1991); Fillis Samuel, COT (1992); Maureen Roche-Manna, COMT (1993); Carol Standardi, RN, CRNO (1994); Ellen Coyle, COMT (1995); Anne Boeckl, MS (1996); Carol J. Pollack-Rundle, COT (1997); Catherine A. Carroll, COMT (SC) (1998); Amy Sporn (1999); Ann Leone (2000). Past participating member: Richard L. Mowery, PhD

Training and Certification Faculty: Fred Ederer, MA, FACE; Douglas E. Gaasterland, MD; Aaron Kassoff, MD (Consultant); L. Jay Katz, MD; Donna Leef, MMSc, COMT; Gregory L. Skuta, MD; Carol Standardi, RN CRNO; Fillis Samuel, COT. Past participating personnel: Anne S. Lindblad, PhD; Cathy Reed, COMT; Coleen C. Beckershoff; Gary Entler, COT; Elizabeth L. Wagner, MPH

REFERENCES

1. Kahn HA, Leibowitz H, Ganley JP, et al. Standardizing diagnostic procedures. *Am J Ophthalmol* 1975;79:768-775.
2. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1976;74:532-572.
3. Tielsch J, Katz J, Quigley HA, et al. Intraobserver and interobserver agreement in measurement of optic disc characteristics. *Ophthalmology* 1988;95:350-356.
4. Varma R, Spaeth GL, Steinman WC, Katz LJ. Agreement between clinicians and an image analyzer in estimating cup-to-disc ratios. *Arch Ophthalmol* 1989;107:526-529.
5. Varma R, Steinman WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* 1992;99:215-221.
6. Zangwill L, Shakiba S, Caprioli J, et al. Agreement between clinicians and confocal scanning laser ophthalmoscope in estimating cup/disc ratios. *Am J Ophthalmol* 1995;119:415-421.
7. The AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials* 1994;15:299-325.
8. The Advanced Glaucoma Intervention Study (AGIS). *Manual of Operations*, 1993. Springfield, Va: National Technical Information Service; accession No. PB93-220192.
9. Hitchings RA, Spaeth GL. The optic disc in glaucoma: I. Classification. *Br J Ophthalmol* 1976;11:778-785.

10. Shields MB. *Textbook of Glaucoma*. 4th ed. Baltimore: Williams & Wilkins; 1998;87.
11. Ritch R, Shields MB, Krupin T, eds. *The Glaucomas*. 2nd ed. St Louis, Mo: CV Mosby; 1996;632.
12. Epstein DL, ed, with Allingham RR, Schuman JS. *Chandler and Grant's Glaucoma*. 4th ed. Baltimore: Williams & Wilkins; 1997.
13. Stamper RL, Lieberman MF, Drake MV, eds. *Becker-Shaffer's Diagnosis and Therapy of the Glaucomas*. 7th ed. St Louis, Mo: CV Mosby; 1999.
14. Seigel DG, Podgor MJ, Remaley NA. Acceptable values of Kappa for comparison of two groups. *Am J Epidemiol* 1992;135:571-578.
15. Goldberg MF, Fine SL, eds. *Symposium on the Treatment of Diabetic Retinopathy*. Washington, DC: US Govt Printing Office; 1969. USPHS publication No. 1890.
16. Chylack LT Jr, Leske MC, McCarthy D, et al. Lens opacities classification system II (LOCS II). *Arch Ophthalmol* 1989;107:991-997.

DISCUSSION

DR MICHAEL A. KASS. Gaasterland and co-workers reviewed intraobserver and interobserver agreement in judging the presence of notching of the optic disc rim in glaucomatous optic nerves. They developed a set of 26 pairs of stereoscopic optic disc photographs, which were reviewed by 14 glaucoma subspecialists who participated in the AGIS trial. The authors found relatively good intraobserver but only fair interobserver reliability in judging notching.

To some extent, the higher intraobserver reliability may be overstated. The glaucoma subspecialists reviewing the photographs were aware that this was a study that included a large number of optic discs with notched rims. The classification system had only 3 possible outcomes (no notching, partial notching, and notching to the rim), thereby reducing the number of choices available to the readers. The readers looked at the photographs in the morning and then again later in the day; thus, the 2 reading sessions were separated by hours rather than by days, months, or years, as would be more typical. It appears from my reading of the manuscript that 2 of the 14 glaucoma subspecialists in this study supplied the photographs and thus may have been more aware of the status of the optic discs.

Other authors have also found a lack of agreement between glaucoma specialists in judging optic disc photographs. In 1976, Paul Lichter reported to this meeting that significant variability existed between glaucoma specialists in evaluating the cup/disc ratio from stereoscopic photographs.[1] The question is whether optic disc photographs can be assessed with a higher degree of interobserver and intraobserver agreement. It is my belief that this is possible, but requires a far more detailed protocol and the use of a reading center. I propose that high degrees of reproducibility can be obtained if the following steps are taken.

1. Develop a standardized protocol for taking, labeling, and shipping photographs.
2. Institute a standardized protocol for training and certifying photographers.
3. Develop a standardized protocol for training readers.
4. Institute a protocol for dealing with conflicting results, ie, reaching consensus.
5. Monitor protocol adherence in an ongoing fashion, including retraining of photographers and readers.
6. Develop a set of standard photographs for training and retesting readers and classifying discs.

In the Ocular Hypertension Treatment Study (OHTS), we have had the opportunity to utilize an optic disc reading center with all of the aforementioned features. We have looked at baseline cup/disc ratio and then re-read the same photographs 1 year, 2 years, and 3 years later. The technicians who read these photographs were not aware that they were re-reading the same set. The percent of regradings at Years 1, 2, and 3 that differed by ≥ 0.2 disc diameters from the estimate of horizontal cup/disc ratio made at entry, was 4 %, 6 %, and 7 %, respectively.[2] The OHTS did not specifically address the question of notching; however, I believe that the use of standardized protocols, standardized training, and ongoing monitoring of protocol adherence can produce highly reproducible measurements.

I thank Dr Gaasterland and the other authors for the opportunity to review this paper.

REFERENCES

1. Lichter P. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1976;74:532-572.
2. Feuer WJ, Parrish RK, Wells M-C, et al. Reproducibility of readings of optic disc cup/disc ratios in the Ocular Hypertension Treatment Study. *Invest Ophthalmol Vis Sci* 1999;40:S283.

[Editor's note] DR RICHARD P. MILLS pointed out that clinicians cannot agree on the disc notching and other measurements of glaucoma damage, and even on the definaition of glaucoma. When confronted with objective measurements of disc morphology, nerve fiber layer measurements, and quality of life, why do we insist opon agreement and reproducibility levels that are higher than we expect of our own clinical assessment? DR JAKOB WILENSKY reiterated the importance of standardization and training of the examiners to achieve reproducible results. DR FREDRICK L. FERRIS emphasized that evaluations by clinicians yield the poorest results in terms of reproducibility because they used preconceived ideas and experience rather than defined rules. DR ALBERT W. BIGLAN pointed out the problems of seeing the projected slides with true binocular vision. DR BARRETT KATZ asked

how this information about discordance among observers influenced the author's method for the evaluation of optic discs.

DR DOUGLAS E. GAASTERLAND. Dr Michael Kass commented on the short interval between readings in our study and that this may have affected reproducibility. Further, he stated , the observers expected to find notching frequently in the images. Indeed, there are alternative study designs. For example, our study might have been better done with a larger number of images and by including duplicate photographs in a single reading set to be examined by the observers during one reading session. In another design, we could have had a long interval between 2 or more reading sessions. The desirability of rigorous training of observers that he emphasized is obvious, and we emphasized it in our paper.

The investigator who supplied the photographs from one center did not participate in the reading exercise. The investigator from the other photograph source center was not the person who picked the images for the set. It is possible he had seen some of the patients or their photographs in the past. He did not participate in the repeat analysis of the images. I doubt his historical proximity to the images has influenced results of our study.

Dr Kass observed that Dr Lichter, as his thesis for the American Ophthalmological Society in 1976, presented one of the earliest studies of glaucoma subspecialist accuracy and reproducibility in interpreting photographic records of optic disc cupping. That is why I was pleased to be able to present this paper on interpretation of optic disc notching at the present American Ophthalmological Society meeting.

There is accuracy possible in optic disc photographic analysis should we follow the procedures Dr Kass outlined. These are a part of the Ocular Hypertension Treatment Study (OHTS) protocol. It think these procedures are important in clinical trials, and it is a false economy to omit them. Further, clinical trials need to keep long-term documentation of accurately attained measurements. In my mind, photographs are the best current approach for optic disc analysis and record preservation.

Dr Richard Mills commented that disagreements between experienced observers about supposedly obvious clinical details are a fact of life. Here is an example. Recently, I had my eyes read by one of the new disc analyzers operated by a person with limited experience. The analysis said I have left eye glaucoma. I said, let's repeat that. An hour later we repeated it, and I was cured. So a high degree of accuracy is possible, but it is true that in this technological age we must attain proficiency in using both old and new tools. It takes training.

Dr Jabob Wilensky seconded the need for standardization and training. He gave an example from experience with Ran Zeimer's flicker comparator for disc photographs. In one study, trained resident physician observers were more accurate with this device than untrained community glaucoma subspecialist clinicians. My friend and teacher, Dr Elmer Ballintine, regularly used the flicker comparator to analyze serial monocular disk photographs from ocular hypertension patients. He voiced great enthusiasm for the method.

I agree theoretically with Dr Rick Ferris, who emphasized a need for highly trained non-clinicians to staff reading centers–clinicians bring a host of personal experience to the assessment, and this may interfere with reproducible reading of photographic records.

Dr Albert Biglan pointed out that for binocularity we need peripheral vision for fusion. This requires picture slides without words on them, and it requires that we observe 2 projected pictures forming a stereoscopic pair by slightly crossing, not diverging, our eyes. My slides violated both principles. For this presentation, I briefly tried to reverse the images–the observer's right eye image on the left side and vice versa. However, 2 of the 4 stereoscopic pairs I showed are in simultaneous split frame format, and I couldn't easily switch the right and left images. So I too, if sitting at the back of the room, would see a peak coming toward me, instead of a cup in the disc, as I looked at the images by crossing my eyes slightly. Next time I do this I'll put the paired images in a crossed configuration on a single slide without writing on it.

Finally, in response to Dr Barrett Katz who, after emphasizing the disconcordance among the observers in our study, asked whether this has changed my clinical behavior when looking at optic discs. The answer is simple. I have confidence in my clinical ability to recognize glaucomatous optic neuropathy. I believe my interpretations and those of other glaucoma subspecialists are reproducible; yet, this study indicates that discussion, training, and using photographic examples might bring our optic disc analyses into better agreement.