Reliability of the Disk Damage Likelihood Scale

JEFFREY D. HENDERER, MD, CONNIE LIU, BA, MUGE KESEN, MD, UNDRAA ALTANGEREL, MD, ATILLA BAYER, MD, WILLIAM C. STEINMANN, MD, AND GEORGE L. SPAETH, MD

• PURPOSE: **To report the reliability of the glaucoma disk damage likelihood scale (DDLS) in comparison to the Armaly cup/disk ratio by determining the interobserver and intraobserver agreement for optic disk stereo photographs and the interobserver agreement for in vivo patient measurements of the optic disk.**
• DESIGN: **Observational case series.**
• METHODS: **Optic disk photographs: 48 stereo pairs of optic nerve photographs were selected from patients with a spectrum of glaucomatous visual field loss. Two masked observers graded the optic disk photographs three times according to the DDLS and Armaly cup/disk ratio. Interobserver and intraobserver agreements were calculated using the test–retest method. Patient measurements: three observers performed in vivo patient measurements on 34 eyes of glaucoma clinic patients and made a single determination of the DDLS stage and Armaly cup/disk ratio, based on the indirect biomicroscopic examination. Level of interobserver agreement was tabulated.**
• RESULTS: **Optic disk photographs: interobserver and intraobserver agreement for the vertical DDLS measurement was greater than for two determinations (clinical impression and measured) of the vertical Armaly cup/disk ratio (interobserver: 85% vs 68% and 74%, respectively; intraobserver grader 1: 97% vs 89% and 80%, grader 2: 99% vs 95% and 89%, respectively). In vivo patient measurements: the interobserver agreement for the DDLS and Armaly cup/disk ratio was similar (70.1% vs 67.6%, respectively).**
• CONCLUSIONS: **For the stereo optic disk photographs, the inter- and intra-observer agreement for the DDLS is greater than the Armaly cup/disk ratio. For the in vivo patient measurements, the level of agreement for the DDLS and the Armaly cup/disk ratio is similar. (Am J Ophthalmol 2003;135:44–48. © 2003 by Elsevier Science Inc. All rights reserved.)**

Accepted for publication Aug 20, 2002.
From the William and Anna Goldberg Glaucoma Service, Wills Eye Hospital, Thomas Jefferson University School of Medicine (J.D.H., M.K., U.A., A.B., W.C.S., G.L.S.), Philadelphia, Pennsylvania; Temple University School of Medicine (C.L.), Philadelphia, Pennsylvania; The Tulane Center for Clinical Effectiveness and Prevention, Tulane University Health Sciences Center, Tulane University (W.C.S.), New Orleans, Louisiana.
This work is supported in part by the Glaucoma Service Foundation, Philadelphia, Pennsylvania.
Inquiries to Jeffrey D. Henderer, MD, Wills Eye Hospital, Glaucoma Service, 840 Walnut Street, Suite 1120, Philadelphia, PA 19107; fax: (215) 928-3903; e-mail: henderer@willsglaucoma.org

CURRENT METHODS OF OPTIC NERVE CLASSIFICAtion focus on the cup, more specifically the cup/disk ratio, as a measure of severity of damage. Such a system was first promoted by Armaly[1,2] in the late 1960s and has been the predominant way of describing the optic nerve ever since. This system offers the advantages of a standardized way of readily evaluating the optic nerve and communicating the results. Furthermore, the cup/disk system does correlate with visual field damage,[3,4] but it suffers from three principal shortcomings. First, it does not describe directly the actual change occurring in glaucoma, namely, the loss of neuroretinal rim tissue. It is this loss that is manifest in an increasing cup/disk ratio. Considering the rim directly rather than the cup may seem of little practical concern, but a problem arises when nerves have oddly shaped cups or notches in the rim. Such changes can indicate severe loss of tissue, but if focal enough, may not increase the overall cup/disk ratio to indicate that there is severe damage present. Refining the cup/disk ratio by considering the vertical ratio[5,6] may be better at identifying glaucoma, but this method suffers from the same problems. Second, the cup/disk ratio system does not possess high interobserver reliability.[7–10] Attempts to overcome this have relied on devices to assist in photographic interpretation[11–13] and on the use of semi-automated measurements of the optic nerve, typically using a imaging system to generate a topographic map.[14–17] Such automated measurements can be prone to error when measuring cup/disk ratios,[18] but they can be very useful for detecting subtle changes that can indicate glaucomatous progression.[19,20] Third, the cup/disk ratio does not correct for the effects of optic nerve size. Large optic nerves have large optic cups and often elevated cup/disk ratios.[21]

  0002-9394/03/$30.00
PII S0002-9394(02)01833-0

Digital image analysis can determine the area of the neuroretinal rim, which has been demonstrated to be a constant despite variations in disk area between individuals and racial groups.[21–23] "Large cups" are not always pathologic, and such "physiologic cupping" can easily be confused for glaucoma.

The glaucoma disk damage likelihood scale (DDLS) is a novel method for classifying glaucomatous optic nerve damage. As in any test, the design must be accurate, reproducible, reliable, and easy to use. Each of these aspects must be investigated individually. To assess accuracy, the degree of optic nerve damage, as defined by the scale, should match other indicators of damage (that is, visual field loss). To investigate the other three aspects of the DDLS performance as a test, it must be judged against the existing methods of optic nerve classification to determine how it compares in reproducibility, reliability, and ease of use. The current study is designed to determine the interobserver and intraobserver agreement of the DDLS optic nerve staging scale and to compare it to the Armaly cup/disk scale.

DESIGN

THIS STUDY WAS DONE AS AN OBSERVATIONAL CASE SERIES.

METHODS

A RETROSPECTIVE REVIEW OF THE RECORDS OF THE GLAUCOMA Service Diagnostic Laboratory at the Wills Eye Hospital was performed to identify patients who had 35 mm stereo optic disk photographs and a Swedish Interactive Thresholding Algorithm (SITA) Standard or FASTPAC visual field examination performed within 6 months of each other. Patients were placed into one of four levels of visual field damage: no damage, mild, moderate, and severe. The criteria were as follows (adapted from Hodapp and associates[24]):

1. No damage (all items are required)
   a. Normal Glaucoma Hemifield Test (GHT)
   b. Pattern standard deviation (PSD) (or corrected pattern standard deviation [CPSD] in the case of a FASTPAC strategy) with a *P* value greater than 5%
   c. No point in the central 5 degrees has a value of less than 15 decibels
   d. A pattern deviation that has no points depressed below the 1% level unless it is located at the extreme top or bottom row of the plot
2. Mild damage (at least one item required)
   a. A GHT result outside of normal limits
   b. A PSD with a *P* value less than 5%
   c. A pattern deviation that has less than one quadrant depressed at the 1% level or worse that is not limited to the extreme top or bottom row of the plot
   d. The central 5 degrees may not have any values less than 15 decibels
3. Moderate damage
   a. Must meet the criteria for mild damage *and*
   b. A pattern deviation plot that has greater than one quadrant, but less than two quadrants depressed at the 1% level or worse
   c. All points in the central 5 degrees must be greater than 0 decibels
4. Severe damage
   a. Must meet the criteria for moderate damage *and*
   b. A pattern deviation plot that has more than half of the plot depressed at the 1% level or worse *or*
   c. Alternatively, if any point in the central 5 degrees has a value of ≤0 decibels, the field will be classified as severe damage, despite the results of the pattern deviations and other aforementioned criteria

Records were excluded if stereo photographs of the optic disk were not available or the photographs were too blurred to permit interpretation.

Using a stand-mounted stereo viewer (Luminos Photo Corp., Yonkers, New York, USA) and light box, two glaucoma specialists staged the optic nerves according to the DDLS (Table 1) and Armaly cup/disk ratio. Forty-eight stereo pairs of optic nerve photos were selected according to the inclusion criteria and were presented to each masked investigator in a random fashion. A 1-cm crosshair, with a 1-mm scale, was taped to the light box and used to measure the size of the cup and disk by superimposing the left photo of the stereo pair. The quotient was used to determine a measured cup/disk ratio. For purposes of the DDLS, a 1.75-mm nerve diameter was assumed for all nerves (Table 1).[25] Graders were asked to note the vertical, horizontal, and overall cup/disk ratio, the vertical, horizontal and overall DDLS stage, and the vertical and horizontal cup/disk ratio using the ruler. Each photograph was examined three times by each grader.

Statistical analysis was performed by SPSS for Windows version 10.1 (SPSS Inc., Chicago, Illinois, USA). Intraobserver and interobserver agreement was determined by the test–retest method for the DDLS, the Armaly cup/disk ratio, and the measured cup/disk ratio. Statistical significance was determined by the Fisher exact test.

Three observers examined 34 eyes of 24 consecutive glaucoma clinic patients during one morning of clinical practice. All patients seen during that morning were included. Eyes were excluded if there was no view of the optic nerve. Each observer made a single determination of the cup/disk ratio and DDLS stage using a Haag–Streit 900 slit lamp and a Volk 66-diopter lens according to the technique of Lim and associates.[26] Each observer was masked to the readings of the fellow observers. The number of interobserver agreements was tabulated. Ob-

TABLE 1. The Disk Damage Likelihood Scale (DDLS)

		True Vertical Disk Diameter (mm)		
		<1.5	1.5–2.0	>2.0
Stage	Significance	Narrowest width of rim		
0a	No damage	0.5	>0.39	>0.29
0b	Probably no damage	0.4–0.49	0.3–0.39	0.2–0.29
1	Possible damage	0.3–0.39	0.2–0.29	0.1–0.19
2	Probable damage	0.2–0.29	0.1–0.19	0.05–0.09
3	Definite damage	0.1–0.19	0.01–0.1	0.01–0.05
4	Definite damage	0.01–0.1	No rim <45°	No rim <45°
5	Definite damage	No rim <45°	No rim 45–90°	No rim 45–90°
6	Definite damage	No rim 45–90°	No rim 91–180°	No rim 91–180°
7	Definite damage	No rim >90°	No rim >180°	No rim >180°

TABLE 2. The Interobserver and Intraobserver Agreement (≤0.1 c/d or ≤ one Stage) for Selected Disk Measurements Using Stereo Optic Disk Photographs

Interobserver	Field Damage	Armaly Vertical c/d Ratio			Overall DDLS Stage		
		Reading 1	Reading 2	Reading 3	Reading 1	Reading 2	Reading 3
Grader 1 + 2	Normal	9/15	11/15	12/15	13/15	14/15	13/15
	Mild	9/12	9/12	8/12	10/12	10/12	10/12
	Moderate	7/10	7/10	4/10	10/10	10/10	10/10
	Severe	7/11	6/11	8/11	8/11	8/11	9/11
	Total	32/48	33/48	32/48	41/48	42/48	42/48
	P value*	0.9	0.7	0.2	0.4	0.2	0.6
Intraobserver		Reading 1–Reading 2	Reading 2–Reading 3	Reading 1–Reading 3	Reading 1–Reading 2	Reading 2–Reading 3	Reading 1–Reading 3
Grader 1	Normal	13/15	14/15	13/15	15/15	15/15	15/15
Grader 2	Normal	14/15	13/15	14/15	15/15	15/15	15/15
Grader 1	Mild	9/12	10/12	10/12	12/12	12/12	12/12
Grader 2	Mild	11/12	12/12	11/12	12/12	11/12	11/12
Grader 1	Moderate	10/10	8/10	8/10	10/10	10/10	10/10
Grader 2	Moderate	8/10	10/10	10/10	10/10	10/10	10/10
Grader 1	Severe	11/11	11/11	11/11	10/11	10/11	10/11
Grader 2	Severe	9/11	10/11	11/11	11/11	11/11	11/11
Grader 1	Total	43/48	43/48	42/48	47/48	47/48	47/48
		P = .2*	*P* = .4*	*P* = .5*	*P* = .3*	*P* = .3*	*P* = .3*
Grader 2	Total	42/48	45/48	46/48	48/48	47/48	47/48
		P = .7*	*P* = .4*	*P* = .6*	*P* = N/A	*P* = .4*	*P* = .4*

c/d = cup/disk; DDLS = Disk Damage Likelihood Scale.
*Fisher exact test.

servers were considered to be in agreement if the two measurements were ≤ 1 DDLS stage and ≤ 0.1 cup/disk ratio.

RESULTS

FIFTEEN PATIENTS HAD NO VISUAL FIELD DAMAGE, 12 HAD mild damage, ten had moderate damage, and 11 had severe damage. The intraobserver and interobserver agreement for selected disk parameters is given in Table 2. A summary of mean levels of agreement among the three trials for the remaining parameters is given in Table 3. Interobserver and intraobserver agreement for the DDLS is greater than the Armaly cup/disk ratio and the measured cup/disk ratio.

Of the 34 eyes, 4 were a stage 0a, 2 were at stage 0b, 7 were at stage 1, 10 were at stage 2, 5 were at stage 3, none was at stage 4, 1 was at stage 5, 2 were at stage 6 and 3 were

TABLE 3. A Summary of the Mean Interobserver and Intraobserver Agreement (≤1 DDLS Stage or ≤0.1 c/d Ratio) for the Remaining Disk Criteria (Percent Agreement and SD) Using Stereo Optic Disk Photographs

		Intraobserver	
	Interobserver	Grader 1	Grader 2
Horizontal c/d ratio	69 (0.02)	89 (0.07)	92 (0)
Vertical c/d ratio	68 (0.02)	89 (0.01)	95 (0.05)
Overall c/d ratio	74 (0.04)	90 (0.04)	90 (0.06)
DDLS horizontal stage	84 (0.03)	97 (0.01)	99 (0.01)
DDLS vertical stage	85 (0.01)	97 (0.01)	99 (0.01)
DDLS overall stage	85 (0.02)	99 (0.01)	99 (0.01)
Measured horizontal c/d ratio	65 (0.04)	79 (0.08)	93 (0.01)
Measured vertical c/d ratio	73 (0.1)	80 (0.05)	89 (0.05)
Armaly horizontal c/d ratio vs measured horizontal c/d ratio		56 (0.14)	60 (0.02)
Armaly vertical c/d ratio vs measured vertical c/d ratio		72 (0.06)	77 (0.06)

c/d = cup/disk; DDLS = Disk Damge Likelihood Scale; SD = standard deviation.

TABLE 4. Level of Interobserver Agreement for In Vivo Patient Measurements Using the DDLS Stage and c/d Ratio (≤1 DDLS Stage and ≤0.1 c/d Ratio) for the Three Observers

DDLS Stage	Agreement of All Three Observers (%) for DDLS Stage	Agreement Between Two of Three Observers (%) for DDLS Stage	Agreement of All Three Observers (%) for Armaly c/d Ratio	Agreement Between Two of Three Observers (%) for Armaly c/d Ratio
0a	4/4 (100)	4/4 (100)	4/4 (100)	4/4 (100)
0b	1/2 (50)	2/2 (100)	2/2 (100)	2/2 (100)
1	5/7 (71)	7/7 (100)	4/6 (67)	6/6 (100)
2	8/10 (80)	10/10 (100)	7/10 (70)	10/10 (100)
3	2/5 (40)	5/5 (100)	2/5 (40)	4/5 (80)
5	1/1 (100)	1/1 (100)	1/2 (50)	2/2 (100)
6	2/2 (100)	2/2 (100)	1/2 (50)	2/2 (100)
7	2/3 (67)	3/3 (100)	2/3 (67)	3/3 (100)
Total	25/34 (74)	34/34 (100)	23/34 (68)	33/34 (97)
P value*	.5	N/A	.6	.5

c/d = cup/disk; DDLS = Disk Damage Likelihood Scale.
*Fisher exact test.

at stage 7, as determined by one of us (G.L.S.). For both the DDLS and the cup/disk ratio determination, all three observers were in agreement to a similar amount (Table 4). The number of times two of the three observers were in agreement was similar using both systems.

We also measured the level of agreement for the in vivo patients a different way, by lumping groups together. When the stages 0a to 3 and 4 to 7 are combined to create two groups of lesser and greater damage, there was no difference in the level of agreement between the groups. For the DDLS, 100% of the time two thirds of the observers agreed for both stages 0a to 3 and stages 4 to 7, 78% of the time 3/3 observers agreed for stages 0a to 3 and 64% of the time 3/3 observers agreed for stages 4 to 7 ($P =$ .4, Fisher exact test). For the Armaly cup/disk ratio, 100% of the time two thirds of the observers agreed for stages 0a to 3 while 92% of the time two thirds agreed for stages 4 to 7 ($P =$ .4, Fisher exact test). Seventy-seven percent of the time 3/3 observers agreed for the cup/disk ratio for stages 0a to 3 and 50% of the time for stages 4 to 7 ($P =$ .1, Fisher exact test).

DISCUSSION

AN EXAMINATION OF THE OPTIC NERVE IS A CRUCIAL PART of an evaluation for glaucoma. Traditionally, physicians have used the Armaly cup/disk ratio to describe the disk.

This is a useful technique, but it is limited by issues of focal rim narrowing and physiologic variability based on disk size. The new DDLS system described here directs the attention of the clinician to the optic disk size and to the region of the optic nerve that is actually being damaged, the neuroretinal rim, by using a scale of the narrowest rim/disk ratio that simultaneously accounts for the overall size of the optic nerve. Such a scale is readily usable, communicable, and overcomes two of the problems (a narrow neuroretinal rim that may not contribute to overall cup/disk ratio measurements and the problem of physiologic cupping) inherent in the Armaly cup/disk ratio.

The DDLS has limitations. It is not easy to apply to tilted optic nerves or to those with sloped temporal rims, who may have an advanced DDLS stage yet have less severe disease as measured by visual field criteria. It does not address the case in which new narrowing of the rim develops in a second area, while already existing damage in a different area remains unchanged.

This study demonstrates that the DDLS is reliable using both disk photos and in vivo patient measurements. Additionally, there is no proof that the scale does not "overcall" optic nerves that may have physiologically narrow rims (myopic tilted nerves and normal nerves with physiologically narrow temporal rims but thick superior and inferior rims) or is useful to detect change over time.

This study is limited by the fact that we did not know the disk size of the photographs that we viewed and we assumed a 1.75 mm size. We believe this is not a problem for a study of intraobserver and interobserver agreement, although it may be a significant limitation as we seek to correlate disk and field damage using this scale.

In conclusion, when examining stereo optic disk photographs, the interobserver and intraobserver agreement for the DDLS is greater than both the Armaly cup/disk system and the optic nerve measurement system. For in vivo patient measurements, the agreement between the DDLS and the Armaly cup/disk ratio is the same.

REFERENCES

1. Armaly MF. Genetic determination of cup/disc ratio of the optic nerve. Arch Ophthalmol 1967;78:5–43.
2. Armaly M, Sayegh R. The cup/disc ratio. Arch Ophthalmol 1969;82:191–196.
3. Hoskins H, Gelber E. Optic disc topography and visual field defects in patients with increased intraocular pressure. Am J Ophthalmol 1975;80:284–290.
4. Douglas G, Drance S, Schulzer M. A correlation of fields and discs in open angle glaucoma. Can J Ophthalmol 1974;9: 391–398.
5. Weisman R, Asseff C, Phelps C, Podos S, Becker B. Vertical elongation of the optic cup in glaucoma. Trans Am Acad Ophthalmol Otolaryngol 1973;77:157–161.
6. Kirsch R, Anderson D. Identification of the glaucomatous disc. Trans Am Acad Ophthalmol Otolaryngol 1973;77:143–156.
7. Lichter P. Variability of expert observers in evaluating the optic disc. Trans Am Ophthalmol Soc 1976;74:532–572.
8. Tielsch J, Katz J, Quigley H, Miller N, Sommer A. Intraobserver and interobserver agreement in measurement of optic disc characteristics. Ophthalmology 1988;95:350–356.
9. Varma R, Steinmann W, Scott I. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology 1992;99:215–221.
10. Wolfs RC, Ramrattan RS, Hofman A, de Jong PT. Cup-to-disc ratio: ophthalmoscopy versus automated measurement in a general population: The Rotterdam Study. Ophthalmology 1999;106:1597–1601.
11. Hitchings R, 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.
12. Klein B, Magli Y, Richie K, Moss S, Meurer S, Klein R. Quantification of optic disc cupping. Ophthalmology 1985; 92:1654–1656.
13. Klein B, Moss S, Magli Y, Klein R, Johnson J, Roth H. Optic disc cupping as clinically estimated from photographs. Ophthalmology 1987;94:1481–1483.
14. Varma R, Spaeth G, Steinmann W, Katz L. Agreement between clinicians and an image analyzer in estimating cup-to-disc ratios. Arch Ophthalmol 1989;107:526–590.
15. Mikelberg F, Douglas G, Schulzer M, Airaksinen PJ, Wijsman K, Mawson D. The correlation between cup–disk ratio, neuroretinal rim area, and optic disk area measured by the video-ophthalmograph (Rodenstock analyzer) and clinical measurement. Am J Ophthalmol 1986;101:7–12.
16. Cucevic V, Brooks A, Strang N, et al. Use of a confocal laser scanning ophthalmoscope to detect glaucomatous cupping of the optic disc. Aust N Z J Ophthalmol 1997;25:217–220.
17. Varma R, Douglas G, Steinmann W, Wijsman K, Mawson D, Spaeth G. A comparative evaluation of three methods of analyzing optic disc topography. Ophthalmic Surg 1989;20: 813–819.
18. Varma R, Steinmann W, Spaeth G, Wilson R. Variability in digital image analysis of optic disc topography. Graefes Arch Clin Exp Ophthalmol 1988;226:435–442.
19. Burgoyne CF, Varma R, Quigley HA, Vitale S, Pease ME, Lenane PL. Global and regional detection of induced optic disc change by digitized image analysis. Arch Ophthalmol 1994;112:261–268.
20. Burgoyne C, Quigley H, Varma R. Comparison of clinician judgment with digitized image analysis in the detection of induced optic disk change in monkey eyes. Am J Ophthalmol 1995;120:176–183.
21. Jonas JB, Zach FM, Gusek GC, Naumann GO. Pseudoglaucomatous physiologic large cups. Am J Ophthalmol 1989; 107:137–144.
22. Varma R, Tielsch J, Quigley H, et al. Race-, age-, gender-, and refractive error-related differences in the normal optic disc. Arch Ophthalmol 1994;112:1068–1076.
23. Tsai C, Zangwill L, Gonzalez C, et al. Ethnic differences in optic nerve head topography. J Glaucoma 1995;4:248–257.
24. Hodapp E, Parrish R, II, Anderson D. Clinical decisions in glaucoma. St. Louis: Mosby-Year Book, Inc., 1993.
25. Bayer A, Harasymowycz P, Henderer JD, Steinmann WG, Spaeth GL. Validity of a new disk grading scale for estimating glaucomatous damage: correlation with visual field damage. Am J Ophthalmol 2002;133:758–763.
26. Lim CS, O'Brien C, Bolton NM. A simple clinical method to measure the optic disc size in glaucoma. J Glaucoma 1996; 5:241–245.