Optic Nerve Cup-to-Disc Ratio Measurement Variability with Semi-Automated Digital Analysis

Jean Hess Barden
Rachel Hollenbeck
Advisor: James Paramore, O.D.

Abstract

Purpose: To evaluate the reliability of semi-automated digital analysis in assessing optic nerve cup-to-disc ratio.

Subjects: Four subjects (four eyes) were evaluated. All subjects had healthy, non-glaucomatous optic nerve heads.

Methods: The subjects' optic discs were captured, stored, and analyzed by two examiners using the Helioasis Digivid 2000 system.

Phase 1- Intra-observer intra-image variability: One image from each disc was analyzed three times consecutively by one observer in order to determine the variability of multiple measures of the same image.

Phase 2- Intra-observer inter-image variability: Three images of each disc were analyzed once by one observer in order to assess the repeatability of measurements from different images of the same disc.

Phase 3- Inter-observer intra-image variability: Three images were taken of each disc and each image was analyzed one time by each examiner in order to assess the resultant variability of two different examiners evaluating each disc.

Results:

Phase 1- Intra-observer intra-image variability: For vertical cup-to-disc (C/D) measurement the standard deviation (SD) average was 0.0177. The coefficient of variation (COV) average was 4.07%. For the horizontal C/D measurement, the SD average was 0.0263 and the average COV was 6.28%.

Phase 2:- Intra-observer inter-image variability: For vertical cup-to-disc measurement the average SD was 0.0222 and the average COV was 6.68%. For the horizontal measurement, the average SD was 0.0383 and the average COV was 8.65%.

Phase 3:- Inter-observer intra-image variability: For the vertical C/D measurement the average SD was 0.0376. The average COV was 14.68%. For the horizontal measurement, the average SD was 0.0817 and the average COV was 26.08%. After rounding the data to 0.05, the Paired Student's t-Test at 95% confidence showed that the vertical measurements were repeatable between examiners 25.22% of the time and 29.52% of the time for the horizontal measurements. However, 63% of the measurements were within 0.05 of each other and 88% were within 0.15 of each other, which showed much better repeatability.

Conclusion: Cup-to-disc ratio measurement with the Helioasis Digivid 2000 system was found to be highly reproducible and would serve as a valid and reliable way of assessing current status and progression of optic nerve cup-to-disc ratio over time.

Introduction

Glaucoma is one of the leading causes of visual impairment and blindness in the world. One of the primary diagnostic factors for this disease is the loss of optic nerve fibers. This is shown by progressive enlargement of the optic nerve cup. Since the axons of the superior and inferior optic disc are affected first, vertical elongation is typically noted first in the disease. Glaucoma also tends to affect the two eyes asymmetrically, often only showing an increase in cup-to-disc ratio of one optic nerve initially. One of the major diagnostic problems with glaucoma is that there are no symptoms of the disease until very late in the progression. This taken with the fact that the amount of optic nerve fibers decrease by 4000 a year gives one an understanding of the importance of early detection and diligent monitoring of glaucomatous signs. (1)

Analysis of the cup-to-disc (C/D) ratio has long been a source of problems for the clinician. The problem with this analysis is that it has always been a subjective estimate. An experienced clinician typically develops repeatability in his own measurements and, therefore, can feel confident whether or not a true change has occurred. Difficulty arises when a patient loses continuity of care by changing practitioners. At this point, the clinician must compare his assessment to someone else's, and this is where the subjectivity and inter-clinician variability become a problem.

Stereoscopic assessment of the optic nerve is considered the most accurate method of determining cup-to-disc ratio. When observing an optic nerve monoscopically factors such as color and blood vessel course are used to determine the amount of cupping. Within the last few years, as technology has advanced, researchers have been striving to develop more objective ways to assess C/D ratio and other determinants of nerve fiber loss. Many involve the use of stereo-photographs and computerized analysis systems. Our work involved the use of a manual computer assisted program using non-stereoscopic digital images. Although we realize the limitations of using two-dimensional photos, the purpose of this study was to show the intra-observer and inter-observer variability using a semi-automated system. Our goal in this study was to determine whether semi-automated digital analysis could be used to decrease clinician error in assessing cup-to-disc ratio.

Patients and Methods

Optic nerve head images were captured, stored, and analyzed with the Helioasis Digivid 2000 system. This system utilizes a non-mydriatic retinal camera. One feature of the computer program allows determination of the C/D ratio when the operator clicks the cursor on the outer and inner margins of the optic nerve rim, thus making four determinations for placement of the cursor on each axis assessed. In this study only the vertical and horizontal axes were assessed, since these are the most commonly reported measurements. Since the images were not in stereo, two-dimensional clues, as discussed above, were used in determining the optic nerve margins. It was decided that no criteria on placement of the cursor would be established between observers prior to assessing the images in order to make a more realistic comparison of results. In a routine examination setting, a practitioner typically would not have another practitioner's measurement of the C/D ratio, and if he did, not the rationale he used to make the assessment.

Four subjects were included in the study. One eye from each subject was used for analysis. At the time of the study, none of the patients were known to have glaucoma or any other optic nerve disease or degeneration. All patients were between the ages of 20 and 26. Healthy eyes were used because again the intent was not to diagnose glaucoma from use of this system but only to assess the variability using semi-automated analysis. Two operators captured the retinal images and used the Digivid system to assess the C/D ratios.

The variability of the cup-to-disc ratio was studied using three different procedures:

1) Intra-observer intra-image variability: One image from one eye of each subject was analyzed three times consecutively. One observer analyzed the images from two subjects and the other observer analyzed the other two images. The purpose of this study was to determine the variability that exists in repeatedly measuring the horizontal and vertical C/D ratio on the same image. (Figure 1)
2) Intra-observer inter-image variability: Three images of each eye were taken and analyzed by each observer. The subject was asked to sit back from the camera for one minute and move his/her head around before recapturing the image. The goal of this study was to assess the consistency of taking horizontal and vertical C/D measurements from different images of the same disc. (Figure 2)
3) Inter-observer intra-image variability: Three images were taken of each eye and each image was analyzed once by each observer. Each operator had no knowledge of the other operator's data. As mentioned earlier, the observers criteria for determination of the rim margins had not been predetermined at the onset of the study so each had to use her own clinical judgement. This study assessed the variability that results from different operators determining the position of the rim margins in order to evaluate C/D ratio both horizontally and vertically. (Figure 3)

The data from these three studies was analyzed by calculating the mean, standard deviation, and coefficient of variation. A Paired Student's t-Test was used to make a comparison between the data of the two observers in study three. The t-Test deals with the problems associated with inference based on small sample sizes. (2)

Results

1) Intra-observer intra-image study: The standard deviation (SD) for the vertical measurements ranged from 0.0003 to 0.0379 with an average of 0.0177. The coefficient of variation (COV) ranged from 0.08% to 6.77% with an average of 4.07%. The SD for the horizontal measurements ranged from 0.0145 to 0.0388 with an average of 0.0263. The COV ranged from 3.96% to 9.75% with an average of 6.28%. Since one operator took measurements for half the subjects and another operator took measurements for the other half we compared their accuracy. The results showed a consistent degree of variability with a COV of 4.35% for Observer 1 and a COV of 4.65% for Observer 2. (Table 1)

2) Intra-observer inter-image study: The SD for the vertical measurements ranged from 0.0152 to 0.0353 with an average of 0.0222. The COV ranged from 2.82% to 10.30% with an average of 6.68%. The SD for the horizontal measurements ranged from 0.0102 to 0.0919 with an average of 0.0383. The COV ranged from 3.50% to 16.09% with an average of 8.65%. We again had two operators, with each measuring half of the subjects. The consistency between the observers was even closer in this study with a COV of 8.55% for Observer 1 and a COV of 8.54% for Observer 2. (Table 2)

3) Inter-observer intra-image study: The SD for the vertical measurements ranged from 0.0042 to 0.1091 with an average of 0.0376. The COV ranged from 1.15% to 46.66% with an average of 14.68%. The SD for the horizontal measurements ranged from 0.0204 to 0.1881 with an average of 0.0817. The COV ranged from 6.81% to 73.56% with an average of 26.08%. In addition, by comparing the means of the vertical measurements by the two observers with the t-Test it was determined with 95% confidence that 22.33% of the time their results would be the same or that roughly 78% of the time they would be different. For the horizontal measurements, the t-Test found with 95% confidence that the results would be the same 25.46% of the time and different about 75% of the time. These statistics refer to achieving exactly the same C/D ratio to four decimal places, which would be impossible clinically. After rounding the data to 0.05 we compared the data again with the t-Test. This raised the probability of achieving the same results to 25.22% for the vertical measurements and 29.52% for the horizontal measurements. Our last assessment took into account how different the two sets of data actually were. We found that 63% or 5/8 of the measurements were within 0.05 of each other and 88% (7/8) were within 0.15 of each other. (Table 3)

Discussion

In comparing the three studies by looking at the differences in the average COV and SD it seems clear that there is a trend. In every study the COV was smaller for the vertical component and larger for the horizontal. This corresponds with results from Rolando, et al. that found vertical C/D ratio to be the most reproducible measurement with the Topcon Image-net System. (3) A reasonable explanation for this phenomenon would be that it is easier to judge rim margins vertically than it is to judge them horizontally. However, much of the data overlaps so it is difficult to make a judgement on whether it is actually more difficult to judge horizontal margins.

It was also clear from looking at the data that the average COV and SD increased from studies one to three. This would lead us to believe that one operator taking multiple measurements on the same photo gives the highest reproducibility; one operator measuring multiple images of the same disc is less reproducible, and multiple operators measuring the same image is the least reproducible. This finding also corresponds with data from the study mentioned previously by Rolando et al. where intra-observer intra-image percent coefficient of variability ranged from 1.64 to 12.8, the intra-observer inter-image variation ranged from 1.63 to 96 (1.63 to 15.99 after excluding outlying data), and the intra-image inter-observer evaluation showed a range of 1.5 to 22.5. (3) In addition, a study by Varma et al. which used the Image-net System 2000 showed intra-image intra-observer variation ranged from 1% to 7% while the intra-image inter-observer variation ranged from 1% to 55%. All our values compare well with these studies except the data from subject 2 in study 3. The percent COV for this subject was 46.66% for vertical C/D ratio and 73.56% for horizontal C/D ratio. After discussing the analysis of that particular image we could appreciate the great differences in assessment of the rim margins. If we excluded this data and did a 'Best Case Analysis,' our range of COVs for all data in that study changed from 1.15% to 73.56% to a range of 1.15% to 14.26%, which compares much more closely.

In preparing studies 1 and 2, we felt that we could assess intra-observer variability by having two observers each measure half of the subjects. When we started compiling data, we felt this technique may have decreased the accuracy of our results. However, after separating data by observer, we can see that the level of variability was very consistent among the two observers for each study. In study 1 the difference in variability was 0.30% and only 0.01% for study 2. In retrospect, although we attained a high level of accuracy with this technique, only one observer should have been utilized in each of the first two studies.

In looking at the t-Test results from study 3, it seemed as if the variability between the two operators was too great to make this technique a reliable way to track optic nerve cupping unless the same person operated the system year after year. However, as was mentioned in the results section, that particular test was too precise in it's definition of "the same." After we rounded the data to 0.05 we saw a slight increase in the repeatability. Once we broke the data down into 0.05 increment differences, we could see that the reliability increased dramatically. We felt that this high reproducibility was much greater than would be seen by ophthalmoscopy clinically. If we had discussed criteria for judging optic nerve margins before the study, we feel our variability would have been even further reduced. However, we chose not to do this for the following reason: Zangwill, et al. writes, "Although lack of training will reduce agreement, it is probably a better representation of actual variation among glaucoma experts in different institutions." (4)

Conclusions

There have been remarkable advances in the diagnosis and treatment of glaucoma. Assessing the progression of optic nerve cupping still remains a critical factor in the diagnosis of glaucoma, and this measurement is still being made primarily by clinician estimation during ophthalmoscopy. Now that computer assisted semi-automated and automated systems exist, it is questionable whether they will become an intregral part of the diagnosis of glaucoma. Semi-automated devices have been proved to detect longitudinal optic disc changes more sensitively than clinical assessment of stereoscopic photographs. (5) This study showed the high reproducibility of cup-to-disc ratio measurements that can be achieved with a semi-automated system. The fact still remains that stereoscopic viewing of the optic nerve is the most accurate method to determine C/D ratio, and that is why we did not assess the precision of measurements or the accuracy of diagnosing glaucoma with the Helioasis Digivid 2000 system. We simply wanted to show that measurements of C/D ratio with semi-automated analysis could be very reproducible. A measuring and analyzing system like the one in our study, utilized with stereoscopic images could prove to be an indispensable way to track a patients optic nerve over his/her lifetime.

Figure 1. Intra-observer intra-image variability: 1 image of 4 different optic discs was analyzed 3 times each.

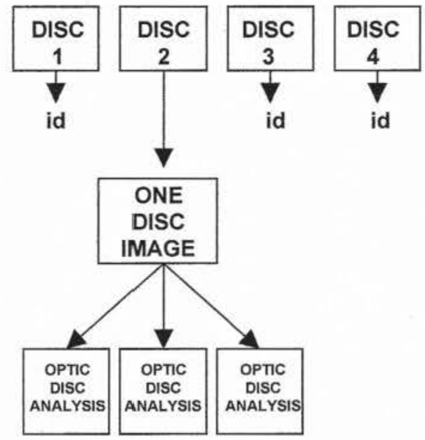


Figure 2. Intra-observer inter-image variability: 3 images of 4 different optic discs were analyzed 1 time each.

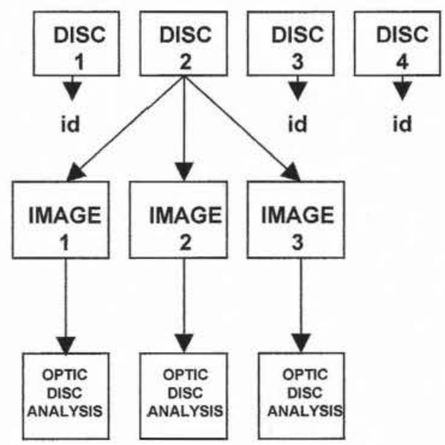
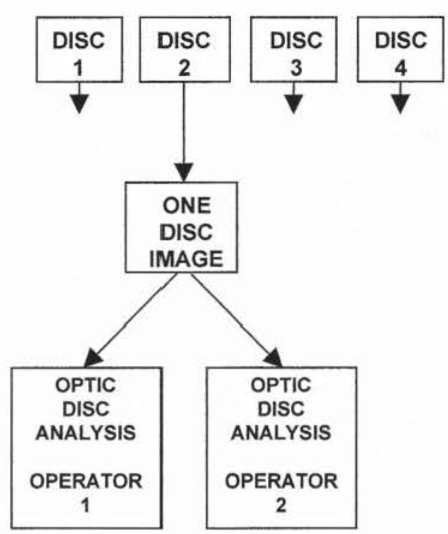


Figure 3. Inter-observer intra-image variability: 1 image of 4 different optic discs was analyzed once by two different observers.



adapted from a diagram by Iester, et al. (6)

Table 1. Intra-observer intra-image variability

	Measurement 1	Measurement 2	Measurement 3	Mean	SD	COV
Subject 1 - V C/D	0.5217	0.5974	0.5587	0.5593	0.0379	6.77%
Subject 2 - V C/D	0.3110	0.3510	0.3385	0.3335	0.0205	6.14%
Subject 3 - V C/D	0.2968	0.2965	0.2963	0.2965	0.0003	0.08%
Subject 4 - V C/D	0.3651	0.3572	0.3809	0.3677	0.0121	3.28%
Total Vertical C/D					0.0177	4.07%
Subject 1 - H C/D	0.6769	0.6064	0.6136	0.6323	0.0388	6.14%
Subject 2 - H C/D	0.3887	0.3596	0.3698	0.3727	0.0148	3.96%
Subject 3 - H C/D	0.2856	0.2592	0.2829	0.2759	0.0145	5.26%
Subject 4 - H C/D	0.3380	0.4082	0.3937	0.3800	0.0371	9.75%
Total Horizontal C/D					0.0263	6.28%
Observer 1 C/D					0.0203	4.35%
Observer 2 C/D					0.0193	4.65%

Table 2. Intra-observer inter-image variability

	Image 1	Image 2	Image 3	Mean	SD	COV
Subject 1 - V C/D	0.5217	0.5482	0.5478	0.5392	0.0152	2.82%
Subject 2 - V C/D	0.3110	0.2816	0.3082	0.3003	0.0162	5.40%
Subject 3 - V C/D	0.2968	0.2582	0.2584	0.2711	0.0222	8.20%
Subject 4 - V C/D	0.3651	0.3021	0.3612	0.3428	0.0353	10.30%
Total Vertical C/D					0.0222	6.68%
Subject 1 - H C/D	0.6769	0.5260	0.5105	0.5711	0.0919	16.09%
Subject 2 - H C/D	0.3887	0.3210	0.3337	0.3478	0.0360	10.35%
Subject 3 - H C/D	0.2856	0.3033	0.2857	0.2915	0.0102	3.50%
Subject 4 - H C/D	0.3380	0.3118	0.3124	0.3207	0.0150	4.66%
Total Horizontal C/D					0.0383	8.65%
Observer 1 C/D					0.0342	8.55%
Observer 2 C/D					0.0333	8.54%

Table 3. Inter-observer intra-image variability

	Observer 1	Observer 2	Mean	SD	COV
Subject 1 - V	0.5217	0.5129	0.5173	0.0062	1.20%
Subject 2 - V	0.3110	0.1567	0.2339	0.1091	46.66%
Subject 3 - V	0.3406	0.2968	0.3187	0.0310	9.72%
Subject 4 - V	0.3711	0.3651	0.3681	0.0042	1.15%
Total Vertical C/D				0.0376	14.68%
Subject 1 - H	0.6769	0.5529	0.6149	0.0877	14.26%
Subject 2 - H	0.3887	0.1227	0.2557	0.1881	73.56%
Subject 3 - H	0.3145	0.2856	0.3001	0.0204	6.81%
Subject 4 - H	0.2947	0.3380	0.3164	0.0306	9.68%
Total Horizontal C/D				0.0817	26.08%

Works Cited

1. Niessen, A. G., and T.J. van den Berg. "Evaluation of a Reference Set Based Grading System for Retinal Nerve Fiber Layer Photographs in 1941 eyes." Acta Ophthalmologica Scandinavian 76 (1998): 278-282.

2. http://www.physics.csbsju.edu/stats/t-test.html. 10 July 2000.

3. Rolando, M., et al. "Measurement Variability in Digital Analysis of Optic Discs." Documenta Ophthalmologica 85 (1994): 211-222.

4. Zangwill, L., et al. "Agreement between Clinicians and a Confocal Scanning Laser Ophthalmoscope in Estimating Cup/Disc Ratios." American Journal of Ophthalmology 119 (1995): 415-421.

5. Wolfs, R. C., et al. "Cup-to-Disc Ratio: Ophthalmoscopy versus Automated Measurement in a General Population: The Rotterdam Study." Ophthalmology 106 (1999): 1597-1601.

6. Iester, M., et al. "Study of the Correlation between Average Values of Optic Disc Parameters and Their Measurement Variability using Stereovideographic Digital Analysis." Ophthalmologica 209 (1995): 177-181.

Other References

Abrams, L. S., et al. "Agreement among Optometrists, Ophthalmologists, and Residents in Evaluating the Optic Disc for Glaucoma." Ophthalmology 101 (1994): 1662-1667.

Barry, C. J., and R.H. Eikelboom. "Comparison of Stereo Optic Disc Photographs from the Nidek 3-DX and Zeiss Retinal Cameras." Australian and New Zealand Journal of Ophthalmology 23 (1995): 203-205.

Barry, C. J., et al. "A Personal Computer-based Method of Stereo Chronometry for Measuring Neuroretinal Rim Width: A Pilot Study." Australian and New Zealand Journal of Ophthalmology 26 (1998): Supplement 1: S22-25.

Haslett, R. S., et al. "Inter-observer Agreement in Clinical Optic Disc Measurement using a Modified 60D Lens." Eye 11 (1997): 692-697.

Komulainen, R., K. Tuulonen, and P.J. Airaksinen. "The Follow-up of Patients Screened for Glaucoma with Non-Mydriatic Fundus Photography." International Ophthalmology 16 (1992): 465-469.

Niessen, A.G., et al. "Design of Low Cost Glaucoma Screening." Documenta Ophthalmologica 93 (1997): 293-315.

Tuulonen, A., et al. "Screening for Glaucoma with a Non-Mydriatic Fundus Camera." Acta Ophthalmologica 68 (1990): 445-449.

Yogesan, K., et al. "Evaluation of a Portable Fundus Camera for Use in the Teleophthalmologic Diagnosis of Glaucoma." Journal of Glaucoma 8 (1999): 297-301