

CHAPTER 8

Sensitivity and specificity of the GDx; clinical judgement of standard print-outs vs the Number

Colen TP and Lemij HG
J Glaucoma, in press

Abstract

Purpose The Number is a standard parameter of the GDx that reportedly separates well between normal and glaucomatous eyes. We evaluated the sensitivity and specificity of the Number and examined whether expert clinical judgement of GDx printouts might lead to a better separation.

Methods Two experienced observers judged 800 GDx scans on 400 randomly presented printouts from 200 glaucoma patients and 200 age-matched normal subjects. Their sensitivity was assessed for all glaucoma patients together, and also for mild, moderate and severe glaucoma separately. Their specificity was determined in the normal subjects. The same was also done for the Number, at various critical values.

Results Both observers discriminated better than the Number. At a critical value of 23, the specificity of the Number was 81.5%, which matched the lowest specificity of the 2 observers: 82.5% and 92.0% for observers #1 and #2, respectively. At these specificities, the sensitivity of the two observers and of the Number were 92.0%, 89.5% and 85.5%, respectively. The sensitivity increased with the severity of glaucoma. The Kappa values for intra observer agreement were 0.80 and 1.0.

Conclusions The Number yielded acceptable sensitivity and specificity values at a critical value of 23 in our test population. However, the clinical judgements of the printouts by both expert observers resulted in a better separation between normal and glaucomatous eyes.

Introduction

The GDx (Laser Diagnostic Technologies, San Diego, CA) is a scanning laser polarimeter¹ that assesses nerve fiber layer (NFL) thickness in the peripapillary retina. The working principle of the GDx is based on the phase shift of polarized laser light as it passes through a birefringent medium such as the retinal NFL. This phase shift is called retardation and is thought to be linearly correlated with NFL thickness, as has been shown in a monkey model.² The GDx has been shown to discriminate well between normal and glaucomatous eyes.³⁻¹⁸

Today, there are no standard procedures for interpreting GDx data. The software compares 13 parameters to a normative database and flags those that are outside normal limits. A 14th parameter is called the Number, which is a probability score ranging from 1 (low probability of glaucoma) to 100 (high probability of glaucoma). The Number is derived from other parameters by a proprietary algorithm that has been established by a neural network. The sensitivity and specificity of each individual parameter, as well as of fixed combinations in an advanced statistical model, have been examined.^{7,12,13,16-19} Of these, the Number is generally the single best parameter, at cut-off levels ranging between 17 and 39.^{7,12,16,17}

Another way of assessing GDx data would be by examining the so-called symmetry analysis printout. In such a printout, each of both eyes is represented by a reflectance

image, a color-coded retardation map, the 14 parameters and a so-called TSNIT graph. This graph represents the circumferential cross-sectional retardation at a specified distance around the optic disc. We would argue that an expert clinical judgement and weighing of all available information on the printouts might yield a better discrimination between normal and glaucomatous cases than any single parameter would. This approach was tentatively explored by Choplin and Lundy.¹⁸ The expert observer might readily recognize artifacts, e.g. caused by motion or large areas of peripapillary atrophy,²⁰⁻²² and therefore judge the printouts differently from a fairly simple software algorithm. In addition, experience with the large variation across normal eyes, both in general NFL thickness, and in typical distributions such as split bundles,²³ might also add to the separating power of expert subjective judgement. The same would apply to localized, wedge-shaped NFL defects that are, in our experience, typically missed by the standard parameters. Finally, examining both eyes on a single printout would allow appreciating any marked asymmetries between the two eyes, and might further improve the detection of glaucoma.

In this study, we assessed the sensitivity and specificity of expert clinical judgement of 'symmetry analysis' printouts of 200 glaucoma patients and 200 age-matched normal subjects. Their scores were compared to the sensitivity and specificity of the Number at various cut-off values.

Subjects We used the so-called symmetry analysis printout that provides information of both eyes simultaneously, because this resembled a clinical setting. The tenet of our approach was that we looked at paired eyes per person, instead of at individual eyes. **Methods**

We recruited 255 consecutive glaucoma patients from our glaucoma clinic. Inclusion criteria were: Caucasian ethnic origin, age between 20-80 years and a diagnosis of glaucoma (on the basis of reliable and repeated glaucomatous visual field abnormalities with matching optic disc abnormalities as established by one of our three glaucoma specialists). A visual field exam was classified as reliable when it met the reliability criteria described by Anderson and Patella.²⁴ Exclusion criteria were diabetes, hypertension requiring medical treatment, any history or ocular disease or surgery. Of these, 14 patients (5.5 %) were unsuitable for GDx imaging due to a very large zone of peripapillary atrophy (13 patients) or inability to fix (1 patient). In the end, reliable, high quality images could be obtained in 241 (94.5 %) glaucoma patients.

As control subjects, 272 healthy volunteers were recruited from the hospital staff, their friends and relatives, and spouses of patients. They met the inclusion criteria of: Caucasian ethnic origin, age between 20-80 years, intra ocular pressure ≤ 21 mm Hg, a normal appearance of the optic nerve head and both normal and reliable visual fields (Humphrey Field Analyzer, 24-2 full threshold program, Dublin, CA). Exclusion criteria were: diabetes, hypertension requiring medical treatment, any history of eye disease or surgery, a vertical cup/disc ratio of 0.6 or more, and an asymmetry in cup/disc ratio greater than 0.2 between the two eyes. Nine of them (3.3 %) were unsuitable for GDx imaging due to either a very large zone of peripapillary atrophy, or a tilted disc yielding an unreliable GDx scan. In the end, reliable, high quality images could be obtained in 263 subjects (96.7%).

IRB/ Ethics Committee approval was obtained for this study and written informed consent was obtained from all participants after all procedures had been fully explained.

After all participants were imaged, we found that the glaucoma group was, on average, older than the normal group. Since age is shown on the GDx printout, this could potentially bias the results. Therefore, we used a computer algorithm to select 200 normal subjects and 200 glaucoma subjects so that the mean age for both groups was comparable (57.8 years versus 59.7 years; $p=0.061$). Demographic data of the normal subjects and glaucoma patients have been summarized in table 8-1.

Measurement procedures All subjects were imaged with the GDx (Laser Diagnostic Technologies, San Diego, CA, software version 2.0.09) by two experienced operators. An imaging session for one eye consisted of obtaining 3 single images of high quality (i.e. good focus, centered optic disc, equal image illumination throughout the image and no motion artifacts). These 3 images were then aligned and converted by the software into one 'mean image'. The results for both eyes were printed on a single sheet of paper (the so-called symmetry analysis printout). During all measurements, we saw to it that patients had their heads as upright as possible. Pupils were undilated and ambient lights were left on.

Interpreting the printouts A total of 400 GDx printouts were presented in random order to the two authors. This was done in two separate sessions on two separate days. The optic nerve head on the reflectance image was masked by a third person with the use of a non-transparent marker. The observers were also masked to the name of the patient and operator initials. The observers had to identify a scan either as glaucomatous (when one or both eyes were glaucomatous) or as normal (when both eyes were normal).

Judging the printouts consisted of these steps: assessing image quality, assessment of the retardation image (overall impression, symmetry between superior and inferior bundle, presence of split bundles, presence of wedge-shaped defects), assessment of the TSNIT plot (position of the plot relative to the normal distribution, symmetry between superior and inferior bundle, shape and maximum height of superior and inferior humps relative to the nasal area), assessment of symmetry between the two TSNIT plots, and finally, evaluation of the Number. The observers were also allowed to look at the other parameters.

Statistical Methods In addition to determining the sensitivity for the entire glaucoma group ($n=200$), we also calculated the sensitivity separately for patients with mild glaucoma (mean deviation >-6 dB), moderate glaucoma (mean deviation <-6 dB but >-15 dB) and severe glaucoma (mean deviation <-15 dB). The specificity was the percentage of the normal subjects that were scored as normal.

Of all 400 scans, 20 were randomly selected and presented again to both observers, in order to calculate the kappa value for intra-observer reproducibility.

The sensitivity and specificity of the Number was first calculated at a critical value of 30, since this value is suggested by the manufacturer and used by many clinicians. We repeated the calculations at a critical value of 23. We selected this value because it lowered the specificity to a level that best matched the lowest specificity of the 2 observers, to facilitate the comparison of sensitivities.

Observer #1 achieved an overall sensitivity and specificity of 92.0% and 82.5%, respectively (table 8-2). For observer #2, these values were 89.5% and 92.0%, respectively. The sensitivities and specificities of the Number have also been given in the table at critical values of 30 and 23. Both observers performed better than the Number at a cut-off value of 23, especially in the group with mild glaucoma.

The sensitivity increased with the severity of glaucoma. Intra-observer reproducibility was very good for observer #1 ($\text{Kappa} = 1.0$) and good for observer #2 ($\text{Kappa} = 0.80$).

In this study, we explored the sensitivity and specificity of both the Number, and of two expert observers, in a large group of glaucoma patients and age-matched controls. We found that both observers discriminated better than the Number which confirms a similar study by Sanchez-Galeana *et al.*²⁵ The difference between expert judgement and the Number in our study was most marked in the group of early glaucoma. In another study (Colen *et al.*, unpublished data) we found that none of the standard parameters discriminated better than the Number.

Choplin and Lundy¹⁸ first explored the power of expert clinical judgement of GDx printouts for detecting glaucoma. This resulted in an average sensitivity and specificity of 80% and 91%, respectively. Nicolela *et al.*,²⁶ also compared the performance of expert clinical judgement of printouts with the performance of parameters. They concluded that clinical judgement of printouts was inferior to the performance of automatically generated parameters (both the Number and a newly devised logistic regression model), which contrasts with our findings and those of Sanchez-Galeana *et al.*²⁵ Differences in study design may account for this contrast: the glaucoma patients in our study had been consecutively recruited from the glaucoma clinic, whereas Nicolela *et al.*,²⁶ selected their patients on the basis of typical visual field defects. Moreover, they judged images of only one eye per patient. In our study, the observers judged images of both eyes of each patient, allowing them to look for any asymmetries. Sanchez-Galeana *et al.*,²⁵ also used the single eye printouts and their reported sensitivity and specificity values were slightly lower than in our study.

It is often unclear how subjective assessments are made. To our knowledge, we are the first to present a systematic description of this process. Our methods have recently become available as an interactive CD-ROM tutorial on the interpretation of GDx data.²⁷ We have also described four image quality criteria. The maximum diagnostic power of the GDx may be limited by a poor quality of the images. Especially motion artifacts have a substantial effect on retardation values, as we have recently described.²² This was also illustrated in the study by Nicolela *et al.*,²⁶ where

Results

Discussion

the sensitivity of two observers increased after 6 images with motion artifacts were excluded from the study group.

Sensitivity and specificity values in the lower eighties to lower nineties, as found in our study, suggest that the GDx may be useful in an ophthalmic clinic. Its role for follow-up for glaucoma may be more important, but still needs to be assessed. Some investigators state that the GDx may not be suited for a screening setting,²⁵ because the sensitivity and specificity values will be lower in a population with less advanced glaucoma. Indeed, we found that the sensitivity increased from early to moderate to advanced glaucoma, both for the observers and for the Number. This was also reported by others.^{7,12,26} However, we argue that the GDx may also play a role in a screening setting. Those who are diagnosed as having glaucoma in a screening project are not exclusively patients with early glaucoma. In fact, in a recent, publicly advertised glaucoma screening program, 22 of 197 screened subjects were found to have glaucoma.¹⁰ Their average mean defect (MD) was -8.2 dB, which may be classified as worse than early glaucoma. In general, the earlier glaucoma is detected, the better, as long as the specificity maintains an acceptable level. To keep the specificity acceptable, one may have to accept that one misses the early glaucoma cases. One has to assess whether the benefits of screening, i.e. detecting the (more advanced) cases of glaucoma, outweigh the costs, which include the missed (earlier) cases. These missed cases may perhaps be detected later with more advanced disease.

The sensitivity and specificity values of the Number (at a critical value of 23) were 86% and 82%, respectively. Sensitivity and specificity values reported by other investigators vary widely.^{7,10,12,13,16,17,25,26,28-30} Several factors may have contributed to these differences of which the most important probably are: differences in population and sample size, the severity of glaucoma, differences in image quality, and the cut-off point of the Number. Therefore, a direct comparison of the various reports probably has only limited meaning.

The GDx compensates for retardation arising from the cornea assuming corneal birefringence of 60 nanometer with a slow axis of 15 degrees nasally downward. In eyes with a different axis, this compensation may be inadequate.³¹ With the aid of a macular scan and a hardware modification of the instrument, the corneal retardation can now more effectively be removed.³² Preliminary data from our own group has shown that such a better corneal compensation increases the sensitivity and specificity further, although the exact extent of this effect remains to be investigated (Reus *et al.*, unpublished data).

Bias in the current study was reduced by masking the observers to everything on the printout that contained information regarding the diagnosis of the patient except for polarization data. They were, however, allowed to evaluate the reflectance image outside the optic disc to check the location of the blood vessels. Any information about the optic disc on the reflectance image was removed with a black marker, to keep polarimetry data as pure as possible. In addition, because nerve fiber layer thickness decreases with age, the groups of normal subjects and glaucoma patients were age-matched. As it stands, scanning laser polarimetry yields reasonable sensitivity and specificity values when only the Number is considered. Better values are achieved when clini-

cians judge the printouts, but this requires experience. A more detailed description of how the authors interpret GDx data is now available on CD-ROM. Future research will focus on how to translate the way the observers interpret a printout into a mathematical model that can be used by the software, thus facilitating an automated diagnostic procedure.

	n	Age (years)	Mean deviation (dB)
Normal subjects	200	57.8 (11.6)	-0.13 (0.91)
All glaucoma patients	200	59.7 (9.1)	-10.5 (7.5)
• Mild	74	60.6 (8.1)	-3.5 (1.8)
• Moderate	78	59.3 (10.1)	-10.3 (2.7)
• Severe	48	58.9 (9.1)	-21.2 (4.8)

Table 8-1.
Demographic data of the study population

Presented are the number of subjects (n), their mean age and their mean deviation (MD) in dB on visual field testing. Standard deviations are presented in parentheses. For the glaucoma patients, the data is given again for the three subgroups separately.

	Observer #1		Observer #2		The Number @ 30		The Number @ 23	
	sens	spec	sens	spec	sens	spec	sens	spec
All	92.0	82.5	89.5	92.0	75.0	90.5	85.5	81.5
Mild	85.1		82.4		63.5		75.7	
Moderate	94.8		92.2		73.1		85.9	
Severe	98.0		95.9		95.8		100	

Table 8-2.
Sensitivity and Specificity as obtained by two different observers, and The Number

Sensitivity (sens) and specificity (spec) values are given in percentages (%) for observer #1, for observer #2, for the Number at a cut-off value of 30 and at a value of 23. The values are given for the glaucoma group as a whole, as well as for the three subgroups.

References

1. Dreher AW, Reiter K. Retinal laser ellipsometry: a new method for measuring the retinal nerve fiber layer thickness distribution. *Clin Vis Sci* 1992;7:481-8.
2. Weinreb RN, Dreher AW, Coleman A, et al. Histopathologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness. *Arch Ophthalmol* 1990;108:557-60.
3. Weinreb RN, Shakiba S, Zangwill L. Scanning laser polarimetry to measure the nerve fiber layer of normal and glaucomatous eyes. *Am J Ophthalmol* 1995;119:627-36.
4. Holló G, Suveges I, Nagymihaly A, Varga P. Scanning laser polarimetry of the retinal nerve fibre layer in primary open angle and capsular glaucoma. *Br J Ophthalmol* 1997;81:857-61.
5. Xu L, Chen PP, Chen YY, et al. Quantitative nerve fiber layer measurements using scanning laser polarimetry and modulation parameters in the detection of glaucoma. *J Glaucoma* 1998;7:270-7.
6. Waldock A, Potts MJ, Sparrow JM, Karwatowski WS. Clinical evaluation of scanning laser polarimetry: II. Polar profile shape analysis. *Br J Ophthalmol* 1998;82:260-6.
7. Weinreb RN, Zangwill L, Berry CC, et al. Detection of glaucoma with scanning laser polarimetry. *Arch Ophthalmol* 1998;116:1583-9.
8. Choplin NT, Lundy DC, Dreher AW. Differentiating patients with glaucoma from glaucoma suspects and normal subjects by nerve fiber layer assessment with scanning laser polarimetry. *Ophthalmology* 1998;105:2068-76.
9. Reyes RD, Tomita G, Kitazawa Y. Retinal nerve fiber layer thickness within the area of apparently normal visual field in normal-tension glaucoma with hemifield defect. *J Glaucoma* 1998;7:329-35.
10. Yamada N, Chen PP, Mills RP, et al. Glaucoma screening using the scanning laser polarimeter. *J Glaucoma* 2000;9:254-61.
11. Lee VW, Mok KH. Retinal nerve fiber layer measurement by nerve fiber analyzer in normal subjects and patients with glaucoma. *Ophthalmology* 1999;106:1006-8.
12. Trible JR, Schultz RO, Robinson JC, Rothe TL. Accuracy of scanning laser polarimetry in the diagnosis of glaucoma. *Arch Ophthalmol* 1999;117:1298-304.
13. Vitale S, Smith TD, Quigley T, et al. Screening performance of functional and structural measurements of neural damage in open-angle glaucoma: a case-control study from the Baltimore Eye Survey. *J Glaucoma* 2000;9:346-56.
14. Kamal DS, Bunce C, Hitchings RA. Use of the GDx to detect differences in retinal nerve fibre layer thickness between normal, ocular hypertensive and early glaucomatous eyes. *Eye* 2000;14:367-70.
15. Essock EA, Sinai MJ, Fechtner RD, et al. Fourier analysis of nerve fiber layer measurements from scanning laser polarimetry in glaucoma: emphasizing shape characteristics of the 'double-hump' pattern. *J Glaucoma* 2000;9:444-52.
16. Lauande-Pimentel R, Carvalho RA, Oliveira HC, et al. Discrimination between normal and glaucomatous eyes with visual field and scanning laser polarimetry measurements. *Br J Ophthalmol* 2001;85:586-91.
17. Poinoosawmy D, Tan JC, Bunce C, Hitchings RA. The ability of the GDx nerve fibre analyser neural network to diagnose glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2001;239:122-7.
18. Choplin NT, Lundy DC. The sensitivity and specificity of scanning laser polarimetry in the detection of glaucoma in a clinical setting. *Ophthalmology* 2001;108:899-904.
19. Tjon-Fo-Sang MJ, Lemij HG. The sensitivity and specificity of nerve fiber layer measurements in glaucoma as determined with scanning laser polarimetry. *Am J Ophthalmol* 1997;123:62-9.
20. Kogure S, Chiba T, Kinoshita T, et al. Effects of artefacts on scanning laser polarimetry of retinal nerve fibre layer thickness measurement. *Br J Ophthalmol* 2000;84:1013-7.

21. Hoh ST, Greenfield DS, Liebmann JM, et al. Factors affecting image acquisition during scanning laser polarimetry. *Ophthalmic Surg Lasers* 1998;29:545-51.
22. Colen TP, Lemij HG. Motion artifacts in scanning laser polarimetry. *Ophthalmology* 2002;109(8):1568-72.
23. Colen TP, Lemij HG. Prevalence of split nerve fiber layer bundles in healthy eyes imaged with scanning laser polarimetry. *Ophthalmology* 2001;108:151-6.
24. Anderson DR, Patella VM. Automated Static Perimetry. 2nd ed. St. Louis: Mosby; 1999:104.
25. Sanchez-Galeana C, Bowd C, Blumenthal EZ, et al. Using optical imaging summary data to detect glaucoma. *Ophthalmology* 2001;108:1812-8.
26. Nicolela MT, Martinez-Bello C, Morrison CA, et al. Scanning laser polarimetry in a selected group of patients with glaucoma and normal controls. *Am J Ophthalmol* 2001;132:845-54.
27. Colen TP, Lemij HG. The Rotterdam GDx Course [Computer program/ CD-ROM]. Rotterdam, The Netherlands: The Rotterdam Eye Hospital. 2002.
28. Greaney MJ, Hoffman DC, Garway-Heath DF, et al. Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci* 2002;43:140-5.
29. Bowd C, Zangwill LM, Berry CC, et al. Detecting early glaucoma by assessment of retinal nerve fiber layer thickness and visual function. *Invest Ophthalmol Vis Sci* 2001;42:1993-2003.
30. Zangwill LM, Bowd C, Berry CC, et al. Discriminating between normal and glaucomatous eyes using the Heidelberg Retina Tomograph, GDx Nerve Fiber Analyzer, and Optical Coherence Tomograph. *Arch Ophthalmol* 2001;119:985-93.
31. Greenfield DS, Knighton RW, Huang XR. Effect of corneal polarization axis on assessment of retinal nerve fiber layer thickness by scanning laser polarimetry. *Am J Ophthalmol* 2000;129:715-22.
32. Zhou Q, Weinreb RN. Individualized compensation of anterior segment birefringence during scanning laser polarimetry. *Invest Ophthalmol Vis Sci* 2002;43:2221-8.

