Population based glaucoma screening

Stoutenbeek, Remco

Chapter 3

Frequency doubling perimetry screening mode compared to the full threshold mode

FREQUENCY DOUBLING PERIMETRY SCREENING MODE COMPARED TO THE FULL THRESHOLD MODE

Remco Stoutenbeek, Govert P. Heeg and Nomdo M. Jansonius

Department of Ophthalmology, University Hospital Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands

Abstract

The diagnostic performance of the frequency doubling perimetry (FDT) C20-1 screening mode was compared to that of the C20 full-threshold mode. For the number of defects $p < 1\%$ in the total deviation plot, both modes appeared to perform similarly in terms of sensitivity, specificity, and area under the receiver–operating characteristic (ROC) curve. Different cut-off points should be applied for both modes to obtain equal sensitivity and specificity values, and – related to that – for most subjects more defects were found in full-threshold mode than in screening mode. For the screening mode, we found a sensitivity of 0.91 and a specificity of 0.88 at a cut-off point of >0 defects, and an area under the ROC curve of 0.93.

3.1 Introduction

Glaucoma is a chronic disease that may cause irreversible blindness, but the early stages of glaucoma do not cause any symptoms. Treatment of glaucoma may arrest or slow down its progress (e.g. Heijl et al., 2002). Therefore, screening for glaucoma might be advisable. Various studies have suggested that the frequency doubling perimeter (FDT) may be suitable for this purpose (Johnson and Samuels, 1997; Quigley, 1998; Cello et al., 2000). FDT can be applied in two different modes, a full-threshold mode and a screening mode. In full-threshold mode, a contrast threshold is determined for each of the test locations by means of a staircase procedure. In screening mode, on the contrary, a suprathreshold strategy is applied. The full-threshold mode produces two global indices (mean deviation and pattern standard deviation) and total and pattern deviation plots. The output of the screening mode is limited to a single plot that can be considered the equivalent of the total deviation plot in full-threshold mode. Therefore, the full-threshold mode might be the more informative one. The screening mode, however, has a much shorter testing time. Besides logistical advantages, the diagnostic performance of the screening mode could be better because a shorter concentration time span is required.

In an accompanying study (Müskens et al., 2004), we compared all previously published algorithms for the interpretation of FDT test results in full-threshold mode. We found that none of the algorithms performed substantially better than simply counting the number of defects $p < 1\%$ in the total deviation plot. Interestingly, this algorithm can also be applied in screening mode. The question is, however, whether the algorithm performs equally well in both modes.

Trible et al. (2000) found for the C20-1 screening mode a sensitivity of 0.39 (n = 51 eyes), 0.86 (n = 42 eyes), and 1.00 (n = 32 eyes), for early, moderate, and severe glaucoma, respectively, and a specificity of 0.95 (n = 95 normal eyes). In full threshold mode they found sensitivities of 0.59, 0.93, and 1.00, respectively, and a specificity of 0.82. Thomas et al. (2002) found for the C20-1 screening mode a sensitivity of 0.81 and a specificity of 0.95. According to their methods, they also performed an FDT N30 full-threshold mode test on all subjects. Unfortunately, these results were not reported in their article. Finally, a third comparison of both modes by Burnstein et al. (2000) was based on only 20 glaucoma patients and nine suspects.

The aim of the present study is to investigate whether the algorithm `counting the number of defects $p < 1\%$ in the total deviation plot' performs at least as well in screening mode as compared to the full-threshold mode. To this end we compared sensitivity, specificity, and the area under the receiver–operating characteristic (ROC) curve in a large group of glaucoma patients and normal subjects.

3.2 Methods

Patient data and study protocol
The glaucoma patients incorporated in the present study were recruited from a large cohort that is currently followed prospectively with FDT and GDx (Nerve Fibre Analyser; Laser Diagnostic Technologies, San Diego, CA, USA) in the glaucoma service of our outpatient department. In this cohort, 452 patients out of 1051 consecutive visitors to the glaucoma service were classified as glaucoma patients. Classification was based on conventional perimetry. Classification of patients can be done both on by-eye basis and on by-patient basis. We decided to perform by-patient analysis because results of this type of analysis are easier to extrapolate to clinical practice: any patient with an abnormality in any eye needs further attention whereas only patients with normal test results in both eyes can be reassured. Thus, in our study a glaucoma patient was defined as a patient with a reproducible visual field defect on conventional perimetry in at least one eye. Defects had to be compatible with glaucoma and without any other explanation. For further details see Heeg et al. (2003). The worse eye of the glaucoma patients had an average mean deviation (MD) Humphrey Field Analyser (HFA) of -11.5 dB (SD 8.6 dB). In 100 consecutive patients out of the 452 glaucoma patients we added an FDT measurement in screening mode to the usual measurement in full-threshold mode. The mean age of these patients was 68 years (SD 13 years; range 28–89). In addition to these 100 glaucoma patients from the glaucoma service, 108 normal subjects were recruited outside the hospital by advertising in local senior departments, among volunteers from the local blood bank, and in other public places. Subjects regularly visiting an ophthalmologist for glaucoma-related reasons were excluded. Both in the glaucoma patients and in the normal subjects no (further) exclusion criteria were applied in order to get samples as representative as possible of the populations studied. The mean age of the normal subjects was 61 years (SD 11 years; range 36–83). We alternated the sequence of the tests; subgroup analyses did not reveal any influence of the testing sequence.

Frequency doubling perimetry
Testing was performed with the frequency doubling perimeter (Carl Zeiss Meditec AG, Jena, Germany) using the C20 full-threshold mode and the C20-1 screening mode. FDT was performed in both the eyes, and a patient was considered positive if there was an abnormal test result or an unfeasible test in at least one eye. The test result was based on the number of defects $p < 1\%$ (called `mild' in C20-1 screening mode) in the total deviation plot. In case of only one functional eye, only that eye was used for classifying the patient and for determining the test result. For details on frequency doubling perimetry see e.g. Maddess and Henry (1992); Johnson and Samuels (1997), and Maddess et al. (1999).

Analysis
Sensitivities, specificities, and Areas Under the ROC Curve (AUC) were calculated for both modes. Confidence limits for sensitivity and specificity were calculated with PEPI 4.0 (Abramson and Gahlinger, 2001); AUCs were calculated using SPSS 10.0.07. Possible differences in sensitivity or specificity values between both modes were analysed with McNemar's test (paired proportions); differences in AUC values were compared with the technique described by Hanley and McNeil (1983). Analyses were performed both for the group as a whole and after stratifying the glaucoma patients into two groups (early and moderate/severe) according to the level of damage. Early glaucoma was defined as MD (HFA) -6 dB or better in the worse eye; moderate/ severe glaucoma as MD(HFA) <-6 dB.

3.3 Results

Table 3.3.1 shows sensitivity and specificity values of both the full-threshold mode and the screening mode at various cut-off points. Figure 3.3.1 presents the corresponding ROC curves. To establish a sensitivity of 0.90, the full-threshold mode and the screening mode appeared to need different cut-off points, >1 and >0, respectively. At these cut-off points, specificities were 0.83 (95% confidence limits 0.75–0.89) and 0.87 (0.80–0.91) for the full-threshold mode and the screening mode respectively. This difference is not significant ($p = 0.30$). The area under the ROC curve was 0.93 (95% confidence limits 0.89–0.96) for the full-threshold mode and also 0.93 (0.90–0.97) for the screening mode (no significant difference). A useless test has an area under the ROC curve of 0.50; a faultless test has an area of 1.00.

Table 3.3.1 Sensitivity and specificity of FDT full-threshold mode and screening mode at various cut-off points.

cut-off	full-threshold mode		screening mode	
	sensitivity	specificity	sensitivity	specificity
>0	0.96	0.72	0.91	0.88
>1	0.91	0.83	0.84	0.94
>2	0.87	0.85	0.78	0.95
>3	0.82	0.90	0.68	0.97
>4	0.75	0.92	0.65	0.98

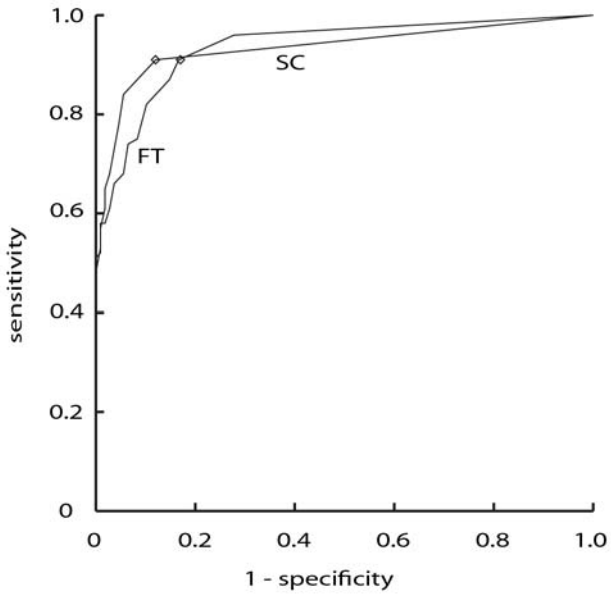

Figure 3.3.1 Receiver–operating characteristic (ROC) curves of the FDT full-threshold mode (FT) and screening mode (SC). Diamonds correspond to cut-off point >1 for full-threshold mode and >0 for screening mode.

Table 3.3.2 shows 2 x 2 matrix form in which the test results of the full-threshold mode (cut-off > 1) and the screening mode (cut-off > 0) are compared in glaucoma patients and normal subjects. As can be seen in this table, four of 100 glaucoma patients and 15 of 108 normal subjects were classified differently by the two modes. Figure 3.3.2 presents a scatter plot of the screening mode and the full-threshold mode results of all 208 subjects. This plot illustrates that for most subjects more defects were found in full-threshold mode than in screening mode.

Table 3.3.2 Test results of the full-threshold mode and the screening mode compared in glaucoma patients and normal subjects.

	glaucoma patients		normal subjects	
	SC = 0	SC > 0	SC = 0	SC > 0
FT ≤ 1	7	2	85	5
FT > 1	2	89	10	8

FT, full-threshold mode; SC, screening mode.

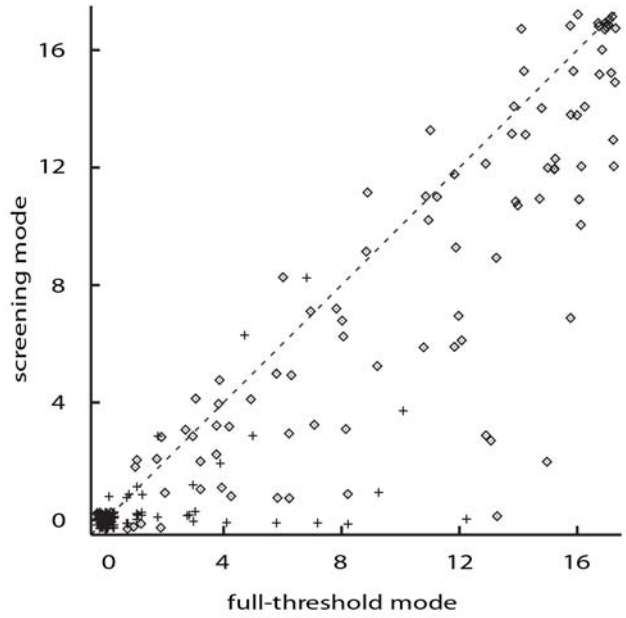

Figure 3.3.2 Scatter plot showing the screening mode and the full-threshold mode test results (number of defects p < 1% in the total deviation plot) of both the glaucoma patients (◊) and the normal subjects (+). Noise with a maximum amplitude of 0.3 was added in order to avoid overlapping dots.

Table 3.3.3 presents the sensitivity of both modes after stratifying the sample into two groups according to the level of damage. The stratifying process yielded 34 patients with early glaucoma and 66 patients with moderate/severe disease. At the cut-off points as mentioned above (>1 for full-threshold; >0 for screening mode), both modes achieved a sensitivity of 1.00 for moderate/severe disease, i.e. they missed none of the moderate and severe glaucoma cases. The sensitivity for early glaucoma did not differ between both modes.

Table 3.3.3 Sensitivity of both modes after stratifying the sample into early and moderate/severe glaucoma.

cut-off	early glaucoma		moderate/severe glaucoma	
	FT	SC	FT	SC
> 0	0.88	0.74	1.00	1.00
> 1	0.74	0.59	1.00	0.97
> 2	0.62	0.47	1.00	0.94
> 3	0.56	0.29	0.95	0.88
> 4	0.41	0.26	0.92	0.85

FT, full-threshold mode; SC, screening mode.

3.4 Discussion

In this study, we compared the diagnostic performance of the FDT screening mode to that of the full-threshold mode. For the number of defects $p < 1\%$ in the total deviation plot, both modes appear to perform similarly in terms of sensitivity, specificity, and area under the ROC curve. However, different cut-off points should be applied for both modes and – related to that – for most subjects more defects were found in full-threshold mode than in screening mode.

In the present study we found, like Trible et al. (2000), that the diagnostic performance of the screening mode is at least equal to that of the full-threshold mode. Moreover, we confirmed their interesting finding that both modes need different criteria to obtain equal sensitivity and specificity values. Trible et al. used >0 defects $p < 0.5\%$ in the total deviation plot to define an abnormal test result in the full-threshold mode and >0 defects $p < 1\%$ in the screening mode. We found >1 defects $p < 1\%$ for the full-threshold mode and >0 defects $p < 1\%$ for the screening mode. These findings are important because they are essential for a proper comparison of FDT test results obtained with different modes in the same subject.

Figure 3.3.2 suggests a rather poor correlation between the full-threshold mode and the screening mode. This might be caused by the fact that the two modes are actually measuring different things. Part of the apparently poor correlation, however, could also be caused by a large test–retest variability of FDT, independent of the mode used (see Heeg et al., 2003). Figure 3.3.2 also shows that most points are below the diagonal, i.e. there is apparently more damage in full-threshold mode. Differences in testing strategy (staircase vs supra-threshold) could be a possible explanation. Fatigue in the longer lasting full-threshold mode could be another factor.

In our data, the mean duration of a test in full-threshold mode was 272s (SD 24s) in normal subjects and 282s (SD 37s) in glaucoma patients; the mean duration in screening mode was 45s (SD 7s) in normal subjects and 79s (SD 36s) in glaucoma patients. The duration of a test in screening mode depends on the number of defects because only missed stimuli are repeated at a higher contrast value. For that reason, Patel et al. (2000) suggested that any test in screening mode lasting longer than 90s could be considered abnormal irrespective of the test result. Alternatively, a new screening mode could be designed that skips the repeated testing at $p < 0.5\%$ and at full contrast. Information regarding the depth of a defect can be skipped without loss of diagnostic performance for screening (Müskens et al., 2004). For other purposes (e.g. disease staging or follow-up), however, information on defect depth might be very useful.

In conclusion, the FDT C20-1 screening mode saves time and performs for screening purposes at least as well as the FDT C20 full-threshold mode. A criterion for an abnormal result in screening mode with a reasonable trade-off between false-positive and false-negative test results is: one or more defects $p < 1\%$ in the total deviation plot.

Acknowledgements

The authors wish to thank Johan Drost, José Eisses, and Wietse Wieringa, technicians in our outpatient department, for performing the FDT measurements. This research was supported by the Dutch Health Care Insurance Council (CVZ) through the Department of Medical Technology Assessment (MTA) of the University Hospital Groningen, the Netherlands.

References for chapter three

1. Abramson JH, Gahlinger PM. Computer Programs for Epidemiologists PEPI Version 4.0. Sagebrush Press 2001; Salt Lake City, UT.

2. Burnstein Y, Ellish NJ, Magbalon M, Higginbotham EJ. Comparison of frequency doubling perimetry with humphrey visual field analysis in a glaucoma practice. Am J Ophthalmol 2000;129:328–333.

3. Cello KE, Nelson-Quigg JM, Johnson CA. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. Am J Ophthalmol 2000;129:314–322.

4. Hanley JA, McNeil BJ. A method of comparing the areas under receiver–operating characteristic curves derived from the same cases. Radiology 1983;148:839–843.

5. Heeg GP, Ponsioen TL, Jansonius NM. Learning effect, normal range, and test–retest variability of frequency doubling perimetry as a function of age, perimetric experience, and the presence or absence of glaucoma. Ophthal Physiol Opt 2003;23:535–540.

6. Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. Arch Ophthalmol 2002;120: 1268–1279.

7. Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. Invest. Ophthalmol. Vis. Sci. 1997;38:413–425.

8. Müskens RPHM, Heeg GP, Jansonius NM. An evaluation of algorithms designed to classify the results from frequency doubling perimetry. Ophthal Physiol Opt 2004;24:498–503.

9. Maddess T, Henry GH. Performance of a nonlinear pathway in subjects with ocular hypertension and glaucoma. Clin Vis Sci 1992;7:371–382.

10. Maddess T, Goldberg I, Dobinson J, Wine S, Welsh AH, James AC. Testing for glaucoma with the spatial frequency doubling illusion. Vision Res 1999;39:4258– 4273.

11. Patel SC, Friedman DS, Varadkar P, Robin AL. Algorithm for interpreting the results of frequency doubling perimetry. Am J Ophthalmol 2000; 129:323–327.

12. Quigley HA. Identification of glaucoma-related visual field abnormality with the screening protocol of frequency doubling technology. Am J Ophthalmol 1998;125: 819–829.

13. Thomas R, Bhat S, Muliyil JP, Parikh R, George R. Frequency doubling perimetry in glaucoma. J Glaucoma 2002;11:46–50.

14. Trible JR, Schultz RO, Robinson JC, Rothe TL. Accuracy of glaucoma detection with frequency doubling perimetry. Am J Ophthalmol 2000; 129:740–745.