CHAPTER 9

General discussion
In this thesis, Scanning Laser Polarimetry (SLP) has been investigated. The clinical studies here have led to the development of image quality criteria and limits for statistically significant change. We have also found that the standard database of the GDx can be used for evaluating amblyopic eyes for glaucoma. Also, the GDx can measure damage to the NFL in non-glaucomatous diseases such as an AION. Further, we have developed a systematic approach for the clinical interpretation of GDx scans. We have found that split bundles (a specific pattern of NFL orientation) are a common finding in healthy eyes. We have shown that the Number is the best single standard parameter for discriminating between glaucomatous and normal eyes. The 4 recently developed GDx parameters did not seem to add to the Number in this respect. Still, the sensitivity and specificity of the Number could be further increased by subjective judgement of all available data on the printout. With this method, the sensitivity and specificity for two different observers was 92.0% and 82.5%, and 89.5% and 92.0% respectively.

Glaucoma remains a disease of which we do not know the precise origin. Many ophthalmologists may have little trouble recognizing the disease when they see its signs in most of their patients. Unfortunately, the inter-observer agreement of optic disc assessment is generally moderate. We still cannot exactly define glaucoma because there is no consensus on the criteria for the signs on which we base our diagnosis. The aim of this thesis was to find out whether scanning laser polarimetry provided better criteria for the definition and detection of glaucoma. The answers to this question have not been equivocal.

The GDx has been used on thousands of normal subjects, people with ocular hypertension and glaucoma patients and its advantages are clear. The applicability is high. We found that in a group of 527 subjects, high quality images could be obtained in even 94.5% of glaucoma patients and 96.7% of normal subjects (chapter 7). Perimetry is considerably less applicable. In one study, only 68.5% of subjects managed to produce a reliable visual field test result. Image acquisition with the GDx is fast and objective. The GDx has a high reproducibility of measurements. It has a normative database that is stratified for ethnicity and different age groups. The sensitivity and specificity found by us are comparable to those of visual field testing, which is currently used as a gold standard by many clinicians. Other topics, however, are still unclear.

First, sensitivity and specificity results for SLP differ greatly in the literature. Results roughly fall into two categories: sensitivity and specificity of automated analysis (e.g. the performance of the Number) and of subjective assessment of the printouts. We have investigated both of them (chapter 7 and 8). In our study, sensitivity and specificity values for the Number (at a critical value of 23) were 86% and 82%, respectively. The results from other studies vary greatly, and they are usually not as favorable as ours. This may be for several reasons. One reason may be that the maximum diagnostic power of the GDx could be limited by a poor quality of the images. We were the first to demonstrate that motion artifacts have a substantial effect on retardation values (chapter 2). This led to the development of 4 image quality criteria that we applied routinely to all our images. Almost none of the other reports mentioned image quality. Only one study group recalculated their data after...
exclusion of images with motion artifacts and found that sensitivity increased.\textsuperscript{34} Another reason for the differences may be that our results were based on the largest sample of test subjects (527 subjects) ever reported. Other sample sizes were smaller, with typically fewer than 50 glaucoma patients. Glaucoma may present as different types of damage in different severities. We therefore believe that a large database of test subjects is needed. Finally, other factors that may have contributed to the differences in findings include differences in population and recruitment methods, the severity of glaucoma and the cut-off point for the Number.

Apart from the sensitivity and specificity of the automated parameters, we investigated the performance of subjective expert judgement of the printouts (chapter 8). Choplin and Lundy\textsuperscript{47} 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. Choplin and Lundy did not compare their data to the sensitivity and specificity of the Number in their group. We found that two expert observers better discriminated between normal eyes and glaucomatous eyes than the Number did. This was in agreement with the findings of Sanchez-Galeana \textit{et al.}\textsuperscript{32} However, Nicolela \textit{et al.}\textsuperscript{34} 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). 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 \textit{et al.}\textsuperscript{34} selected their patients on the basis of typical visual field defects. Moreover, they judged images of only one eye per patient. We think that by comparing the 2 eyes one may obtain more useful clinical information, notably with interocular asymmetries, as was done in our study. The manufacturer now also recognizes that information on asymmetry between the 2 eyes may be valuable and in the latest versions of the GDx software all parameters are now compared between eyes with the differences presented on the printout. Whether this leads to a better separation between normal and glaucomatous eyes remains to be investigated.

There is a second topic that lacks clarity. At present it may still be unclear how to interpret clinical GDx data and there is no consensus on how to define an abnormal GDx scan. In most studies, the Number is the best separating parameter between normal and glaucomatous eyes. We showed that the four newly devised parameters add little in that respect. There is some agreement on what cut-off point for the Number should be used. We found a value of 23 to yield the best trade-off between sensitivity and specificity. This agrees with the values between 17 and 35 found in the literature.\textsuperscript{34,38-40} In contrast, there is almost no information on how the GDx printouts are best subjectively assessed. We were the first to describe how we interpreted the printouts. Later, this was more extensively covered in the GDx course we developed. Some clinicians have copied this approach but it remains to be investigated what sensitivity and specificity values these clinicians reach with this method of interpretation.

Also, SLP has not been histologically validated in humans. Retardation values were correlated to the histological thickness of the NFL in two monkey eyes with their corneas removed,\textsuperscript{48} and in one monkey eye with an intact cornea.\textsuperscript{5} No human data
is presently available. One might argue that this is not a problem as long as the measurements discriminate well between normal subjects and glaucoma patients as well as reliably detect change over time. Yet, before a claim can be made that SLP measures NFL thickness, this needs to be validated first. A human validation study has recently been initiated by our group.

There is insufficient data on whether the GDx can detect glaucomatous progression. Thus far, we have evaluated parameters only for case finding. Parameters with a high sensitivity and specificity are not necessarily also the best parameters for follow up. Although we found some evidence that the GDx can detect glaucomatous change, we do not presently know what parameters are most suited to monitor this. A good parameter would need to be valid, and have a high reproducibility of measurements.\textsuperscript{49} Validity (also called accuracy) means that the measurement (e.g. NFL thickness) is correct for the patient being studied and relevant for the disease being studied. This means that the parameter will change as the state of the disease changes. Reproducibility of measurements (also called reliability or precision) is the extent to which repeated measurements of a relatively stable phenomenon fall closely to each other. In case of glaucoma, this means that the parameter will not change when the state of the disease remains stable. With the GDx, reproducibility of measurements is high.\textsuperscript{44-46,50-52} Our study was the first to express reproducibility of measurements in terms of limits of agreement.\textsuperscript{53} These limits indicate how much a GDx parameter from two consecutive measurements is allowed to change before becoming statistically significant. Our study identified those parameters that had, on average, the best limits of agreement. It is, however, unlikely that the same limits apply to every subject. Ideally, one would like to obtain limits of agreement per parameter per subject, but this would be very elaborate. Even though the parameters with the best limits of agreement have been identified in chapter 3, a large follow-up study would be needed to establish whether these parameters hold sufficient validity to detect glaucomatous change. Such a study has already been launched by our group.

The output parameters merit additional research. Although there are different parameters that relate to different locations of the retardation image, they do not take into account the patients individual pattern of NFL orientation. We identified one such pattern and named it a split bundle (chapter 4). We showed that the outcome of some parameters could depend on the pattern of NFL orientation.\textsuperscript{54} Specificity of those parameters may increase when, in the future, parameters would be normalized for different patterns of the imaged NFL. Whereas a split bundle reflected a normal NFL variation, a wedge shaped defect is recognized as an early sign of glaucoma.\textsuperscript{55} In our experience, wedge shaped defects are easily and regularly observed in the retardation image. However, there is no algorithm available that can detect them. Such an algorithm may greatly improve the sensitivity of the automated analysis of GDx images. Another parameter that may be worth developing would focus on the temporal part of the arcuate bundles instead of the entire bundle. Glaucomatous damage tends to be more prominent over the temporal part than over the nasal part of the NFL, notably in the earlier stages of the disease. Over the past few years, there have been attempts to create new parameters.\textsuperscript{25,56} Four of these have been tested in chapter 7. It was found that they discriminated well
between normal eyes and glaucomatous eyes, but none of the 4 new parameters, on their own, discriminated better than the Number. The high sensitivity and specificity of the new parameters, however, does suggest that they may add to a new algorithm of the Number. Such a contribution is less likely for the Discriminant Analysis (DA), since it is calculated from 3 already existing parameters. It is unclear what the performance of such a discriminant analysis would be with 4 (or more) input variables. Anyhow, the high performance of the DA was already a surprising finding; being calculated from only 3 existing parameters (average thickness, ellipse modulation and ellipse average) the area under the ROC curve was the same as for the Number. In contrast, the algorithm for the Number consists of a combination of 28 parameters, of which only 13 are included on the printout. This suggests that all the information the Number needed up to now, was already contained in those three parameters. This also suggests that some of the parameters that have a low area under the ROC curve might better be left off the printout. On today’s version of the GDx Access (a portable version of the GDx, but with comparable hardware) printout, 5 parameters that performed poorly in chapter 7 have been omitted in the printout.

We know little about the NFL thickness in subjects who have other diseases than glaucoma. In chapter 5 we reported measurements in amblyopia and found no differences in NFL thickness between amblyopic eyes and normal eyes.\(^{57}\) This was later confirmed by a second report.\(^{58}\) We also documented SLP measurements in an AION patient (chapter 6).\(^{59}\) Others have documented GDx measurements in patients with intracanalicular and intracranial lesions to the optic nerve,\(^{60}\) traumatic optic neuropathy,\(^{61}\) demyelinating optic neuritis,\(^{62}\) myelinated retinal nerve fibers,\(^{63}\) and Alzheimer’s dementia.\(^{64}\) Of even more clinical interest, would be to know whether and how more common disorders such as diabetes or systemic hypertension would affect the SLP measurements. Only one study reports on changes found in diabetic patients, but the studied group was small (12 patients).\(^{65}\)

Despite the promising results that are presently obtained with the GDx, the clinician is still faced with over 10% of inaccurate test results. Part of these misclassifications can probably be explained by the cornea. The GDx compensates for retardation arising from the cornea and crystalline lens on the assumption that a corneal birefringence of 60 nanometer with a slow axis of 15 degrees nasally downward exists in all patients. Greenfield et al. showed that the corneal birefringence varies considerably between subjects.\(^{66}\) It was estimated later that in eyes with a corneal polarization axis (CPA) that differed substantially from average, the compensation was inadequate.\(^{67,68}\) Garway-Heath et al. showed that a software correction of the retardation images, based on retardation data of the macular region, could narrow the variability of retardation readings in normals and glaucoma patients.\(^{69}\) Also, the sensitivity for discriminating between normal subjects and glaucoma patients increased. Greenfield et al. measured the CPA in a group of patients using a slit lamp mounted device that incorporated two crossed linear polarizers and an optical retarder.\(^{70}\) They found that the incorporation of the CPA in their logistic regression model increased the power to discriminate between normal and glaucomatous eyes. Later, Zhou and Weinreb reported results using a modified GDx that accommodated a variable corneal compensator.\(^{71}\) A polarimetry image of the Henle fiber layer was
used to set the variable compensator before acquiring peripapillary data. Using a macular scan to verify the effectiveness of compensation, they showed that individualized anterior segment compensation was feasible. A new generation GDx that features this variable corneal compensation is called GDx VCC and has recently become commercially available. We expect that the sensitivity and specificity for detecting glaucoma will increase further with the GDx VCC, although this is still under investigation.

SLP has evolved at a high pace over the past decade. This has added to the need for practical instruction to those who are new to or inexperienced with the technology and has prompted us to develop The Rotterdam GDx Course. This course has been presented to about 800 ophthalmologists and ophthalmic technicians worldwide. From feedback of the course participants we know that hands-on imaging training and instruction for interpretation of clinical data were highly appreciated. This might be a signal to the manufacturer that new techniques may not reach their full potential until they are accompanied by clinically relevant instructions and training. Now that the Rotterdam GDx Course has been translated onto an electronic platform, it will hopefully be available to a wider audience. We speculate that the need for systematic training also applies to other imaging techniques in glaucoma.

In summary, the GDx provides fast, objective and quantitative data on NFL thickness. The applicability and reproducibility of measurements are high, and measurements are user and patient friendly. The GDx yields useful sensitivity and specificity values for the detection of glaucoma, whereas its role in follow-up remains to be investigated. As it stands, the GDx holds insufficient validity to serve as a single test for glaucoma. It does, however, provide a very useful addition to the existing tests we run in patients to make the correct diagnosis.
References


