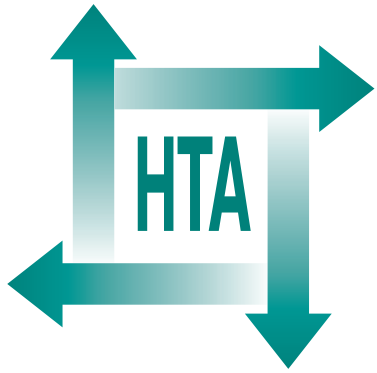
Health Technology Assessment 2005; Vol. 9: No. 46

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma

AJ Kwartz, DB Henson, RA Harper, AF Spencer and D McLeod

November 2005

Health Technology Assessment
NHS R&D HTA Programme



How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma

AJ Kwartz,* DB Henson, RA Harper, AF Spencer and D McLeod

Department of Ophthalmology, University of Manchester, UK

* Corresponding author

Declared competing interests of authors: none

Published November 2005

This report should be referenced as follows:

Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D. The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma. *Health Technol Assess* 2005;**9**(46).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE and *Science Citation Index Expanded (SciSearch®)* and *Current Contents®*/Clinical Medicine.

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA Programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, service-users groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including service users) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or conducting a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a short time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this monograph was commissioned by the HTA Programme as project number 95/18/04. The contractual start date was in April 1998. The draft report began editorial review in July 2004 and was accepted for publication in March 2005. As the funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

Abstract

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma

AJ Kwartz,* DB Henson, RA Harper, AF Spencer and D McLeod

Department of Ophthalmology, University of Manchester, UK

* Corresponding author

Objectives: To determine the potential of optic nerve head tomography [Heidelberg Retina Tomograph (HRT)] and scanning laser polarimetry (GDx) for identifying patients with glaucomatous visual field loss.
Design: Examinations were performed with the HRT, GDx and Humphrey Field Analyzer (HFA). Glaucoma was defined by the presence of a field defect. Patients within the cross-sectional groups underwent a single examination, whereas patients in the longitudinal groups were examined 6 monthly, for an average of 3.5 years.
Setting: Manchester Royal Eye Hospital, UK.
Participants: Patients with primary open angle glaucoma (POAG) or who were at risk of developing glaucoma.
Interventions: The diagnostic accuracies of the HRT and GDx were compared; specificity was set at 95%. The rate of change was determined by linear regression. To estimate the clinical application of the instruments, the proportion of an unselected group of patients on whom the examinations could be performed was calculated. Additionally, the time taken to perform and process each examination was measured.
Main outcome measures: The ability of the techniques to identify cases showing deterioration. The level of agreement and applicability of the techniques. Time taken to perform and process each examination.
Results: From the cross-sectional group, the maximum sensitivities of the HRT and GDx were 59% and 45%, respectively (at 95% specificity). From the two longitudinal cohorts, the level of agreement between the three instruments for identification of the development and deterioration of POAG was low. The applicability of the techniques was 80% (HRT), 88% (GDx) and 98% (HFA). The length of time to perform a full examination with each instrument was 12.3, 11.8 and 28.3 minutes, respectively. Agreement of HRT and GDx parameters between and within observers was largely good.
Conclusions: There is poor agreement for detection of glaucoma between the HFA, HRT and GDx. The techniques are amenable to use in the clinical environment, but no single examination has sufficient diagnostic precision to be used in isolation; also, the imaging techniques were not universally applicable. Neither the HRT nor GDx should be viewed as a replacement for visual field examination. Further research is needed into why most patients within the longitudinal arms of the study showed very little deterioration and into determining aspects of the structure versus function relationship in glaucoma that may explain why any one technique fails to detect a proportion of cases.

Contents

List of abbreviations

AGIS	Advanced Glaucoma Intervention Study
ARVO	Association for Research in Vision and Ophthalmology
C:D	cup-to-disc ratio
CI	confidence interval
CSLO	confocal scanning laser ophthalmoscopy
DC	dioptre cylinder
DFA	discriminant function analysis
DS	dioptre sphere
EMGT	Early Manifest Glaucoma Trial
FDD	floppy disk drive
FL	fixation loss
FN	false negative
FP	false positive
GHT	glaucoma hemifield test
HDD	hard disk drive
HES	Hospital Eye Service
HFA	Humphrey Field Analyzer
HRT	Heidelberg Retina Tomograph
ICC	intraclass correlation coefficient
IOP	intraocular pressure
LDF	linear discriminant function
MD	mean deviation
MPD	mean pattern deviation
MRA	Moorfields Regression Analysis
MREH	Manchester Royal Eye Hospital
MSD	mean standard deviation
NFA	Nerve Fiber Analyzer
NRR	neuroretinal rim
NTG	normal tension glaucoma
OHT	ocular hypertension
ONH	optic nerve head
POAG	primary open angle glaucoma
PPA	peri-papillary atrophy
PSD	pattern standard deviation
QUORUM	Quality of Reporting of Meta-Analyses
RNFL	retinal nerve fibre layer
ROC	receiver operating characteristic
SITA	Swedish interactive thresholding algorithms
SLO	scanning laser ophthalmoscopy
SLP	scanning laser polarimetry
UMIST	University of Manchester Institute of Science and Technology

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Objectives

The objectives of this study were:

- To compare the diagnostic accuracy of optic nerve head tomography [Heidelberg Retina Tomograph (HRT)] and scanning laser polarimetry (GDx) for identifying patients with glaucomatous visual field loss.
- To investigate the applicability of the instruments in an unselected population of hospital patients.
- To measure the length of time required for a full examination.
- To calculate between- and within-observer variability in HRT and GDx measurements.

Design

Examinations were performed with the HRT, GDx and Humphrey Field Analyzer (HFA). Glaucoma was defined by the presence of a field defect. Patients within the cross-sectional groups underwent a single examination, whereas patients in the longitudinal groups were examined 6 monthly, for an average of 3.5 years.

Setting

The study was carried out by the University of Manchester at Manchester Royal Eye Hospital.

Participants

Cross-sectional groups:

- 98 normal controls
- 152 patients with primary open angle glaucoma (POAG).

Longitudinal groups:

- 240 patients at risk of developing glaucoma (either due to high intraocular pressure, and/or a fellow eye with POAG)
- 75 patients with POAG.

Main outcome measures

For the cross-sectional groups, the diagnostic accuracies of the HRT and GDx were compared; specificity was set at 95%. The extent of agreement was determined. In the longitudinal cohorts, the rate of change was determined by linear regression. The ability of the techniques to identify cases showing deterioration was investigated.

To estimate the clinical application of the instruments, the proportion of an unselected group of patients on whom the examinations could be performed was calculated. Additionally, the time taken to perform and process each examination was measured.

Results

From the cross-sectional group, the maximum sensitivities of the HRT and GDx were 59% and 45%, respectively (at 95% specificity). From the two longitudinal cohorts, the level of agreement between the three instruments for identification of the development and deterioration of POAG was low.

The applicability of the techniques was 80% (HRT), 88% (GDx) and 98% (HFA). The length of time to perform a full examination with each instrument was 12.3, 11.8 and 28.3 minutes, respectively.

Agreement of HRT and GDx parameters between and within observers was largely good.

Conclusions

There is poor agreement for detection of glaucoma between the HFA, HRT and GDx. The techniques are amenable to use in the clinical environment, but no single examination has sufficient diagnostic precision to be used in isolation; also, the imaging techniques were not universally applicable. Neither the HRT nor GDx should be viewed as a replacement for visual field examination.

Implications for healthcare

All cases of suspect, incipient or progressing glaucoma cannot be detected by one form of examination (e.g. HRT, GDx or HFA) alone. Since agreement between the three techniques is low, several different tests are necessary to optimise diagnostic precision.

Further research

The following areas are recommended for further research:

- To determine why most patients within the longitudinal arms of the study showed very little deterioration.
- The determination of aspects of the structure versus function relationship in glaucoma, which may explain why any one technique fails to detect a proportion of cases.

Chapter 1
Introduction

This chapter provides a brief introduction to glaucoma and describes the methods of clinical assessment used in the typical NHS glaucoma clinic. Shortcomings of these techniques and the burden that glaucoma poses to the health service are also discussed.

Introduction to glaucoma

There is no universally accepted definition of glaucoma.[1] It can be considered as a group of diseases that result in progressive optic neuropathy, with characteristic morphological changes at the optic nerve head and associated visual field defects (the visual field has been defined as 'that portion of space in which objects are simultaneously visible to the steadily fixating eye'[2]). Although raised intraocular pressure (IOP) was traditionally used as a defining feature of glaucoma, contemporary opinion reflects the fact that increased IOP is by no means pathognomonic of the condition.[3]

Primary open angle glaucoma (POAG) is idiopathic, whereas secondary causes of the condition are, *inter alia*, pseudoexfoliation, pigment dispersion, trauma and neovascularisation. POAG has the greatest prevalence of all the glaucomas, accounting for 50–66% of cases,[4] and is the condition considered in this report.

Two clinical entities confound the diagnosis of glaucoma. The first is cases where IOP, a major risk factor, lies outside the normal range, but there are no clinically detectable signs of glaucomatous optic neuropathy – this condition is categorised as ocular hypertension (OHT).[5,6] The second is when individuals develop glaucomatous optic neuropathy in the absence of raised IOP. This is known as normal tension glaucoma (NTG) and accounts for 25–50% of cases of open angle glaucoma.[7]

Glaucoma affects many aspects of visual function. However, the two most significant are damage to the visual field and reduction in visual acuity. In the early stages of POAG, most patients are asymptomatic and detection of the disease is dependent on clinical examination of the optic nerve head (ONH), perimetric investigation of the visual field and measurement of IOP. Glaucoma differs from other common causes of visual impairment (e.g. cataract and age-related macular degeneration) in that it has the potential to cause complete and irreversible bilateral blindness.

Epidemiological studies have determined many factors that are associated with an increased prevalence of POAG. These include increased IOP, age, race, positive family history and myopia; their manifestation as risk factors for the development of glaucoma has been widely documented.[8,9]

Epidemiology

The precise epidemiology of blindness and glaucoma is confounded by a lack of universal definitions and disparities between sampling methods.

The glaucomas are a major cause of blindness in the UK, causing 15% of registerable blindness,[10,11] although this may underestimate the true prevalence owing to under-registration.[12] Many different definitions and classifications for glaucoma are used, making direct comparison between studies difficult. *Table 1* compares the prevalence of open angle glaucoma from a variety of studies.

Clinical diagnosis and management of glaucoma

Currently, the three principal forms of examination for patients with suspect or diagnosed primary open angle glaucoma are measurement of IOP, examination of the ONH and perimetric assessment of the central visual fields.[24] However, each of these techniques has significant limitations with regard to its ability to detect and/or monitor the condition.

Tonometric measurement of intraocular pressure

Within the Hospital Eye Service (HES), IOP is usually measured with a Goldmann tonometer.[25]

TABLE I *Prevalence of glaucoma*

Population studied	No. of subjects	Prevalence (%)	Age (years)
Skvode, Sweden[13]	7,275	0.4	>40
Oxford, UK[14]	2,000	1.5	40–75
Framingham, USA[15]	2,675	2.2	52–85
Dalby, Sweden[7]	1,511	0.9	55–70
Nottingham, UK[16]	874	1.7	>49
Beaver Dam, USA[17]	4,926	2.1	43–84
Roscommon, Ireland[18]	2,186	1.9	>49
Rotterdam, The Netherlands[19]	3,062	1.1	>55
Sydney, Australia[20]	3,654	3.0	>49
Egna-Neumarkt, Italy[21]	4,297	1.4	>40
Mobile Glaucoma Unit, Israel[22]	10,037	1.0	18–95
Egna-Neumarkt, Germany[23]	4,927	2.0	>40

However, measurement of IOP has poor diagnostic precision. There are several contributory factors to this shortcoming: the proportion of patients with normal tension glaucoma,[26,27] the dependence of the measurement on central corneal thickness[28] and the diurnal variation in IOP[29] (which is enhanced in some glaucoma patients[30]).

Assessment of the optic nerve head and retinal nerve fibre layer

Within the HES, these structures are usually examined by binocular indirect ophthalmoscopy and a photographic record is sometimes made. However, the diagnostic precision of examination of the optic disc is limited by high levels of inter- and intraobserver variability.[31,32]

Typical characteristics of the glaucomatous ONH (which is often referred to by its synonym of optic disc) include: an increase in cupping [clinically detected as an increase in cup:disc (C:D) ratio]; notching of the neuroretinal rim; peri-papillary atrophy (PPA); alteration to the configuration of blood vessels at the optic disc; and optic disc haemorrhages. Although it is widely felt that these changes precede the onset of visual field defects,[33] changes are often difficult to distinguish from the wide range of normal cases, and perimetric changes often precede detected morphological change at the ONH.[34,35]

Despite the wide range of ONH changes described above, there are significant limitations to their use as diagnostic features. One of the main reasons is the large degree of overlap between the appearance of normal and glaucomatous discs.[36] In addition, some of the alterations (e.g. changes to the configuration of blood vessels) occur mainly in the later stages of glaucoma, whereas others (e.g. optic disc haemorrhages[37]) have poor diagnostic precision because, although they have high specificity, they are only detected in between 4 and 7% of eyes with glaucoma.[38–40]

In addition to the ONH, glaucomatous changes are also encountered at the retinal nerve fibre layer (RNFL). However, ophthalmoscopic examination of the RNFL is notoriously difficult, and only yields qualitative results. Grading scales have been described,[41,42] but are infrequently used in the clinical setting. Photography of this structure is possible, but is often limited by the presence of media opacities.[43]

Perimetric examination of the visual field

There is a characteristic pattern of visual field loss in glaucoma. Glaucomatous defects include: diffuse loss;[44] arcuate scotomas within the central 25° (which classically respect the horizontal midline); nasal step; hemifield loss; and a small remaining central or temporal island in advanced cases.[45–49]

Visual field analysis in the secondary and tertiary care sectors is commonly performed with the Humphrey Field Analyzer (HFA), or similar instruments. The most commonly used examinations are the full threshold and SITA 24-2 strategies, which are lengthy (up to 30 minutes to examine both eyes) and, like all perimetric examinations, require an attentive patient with good observational skills. The test is universally disliked by patients, who find it stressful, tiring and arduous.

The ability of perimetry to monitor glaucoma is compromised by high levels of variability.[50] Consequently, small amounts of change cannot easily be detected. It is often necessary to repeat a visual field examination owing to unreliable results, with subsequent costs to both the patient and healthcare provider.

The clinical application of perimetry is limited by both learning[51–54] and fatigue effects.[55,56] The manifestation of the former is that novice patients may require multiple examinations (≤5 tests[51]) to achieve results that are of useful diagnostic value. The latter phenomenon is more marked in the elderly[57] and those with glaucomatous field defects,[58,59] which paradoxically reflects the profile of patients attending glaucoma clinics.

The HFA has the ability to generate indices that are specifically designed to detect glaucomatous visual field defects, for example, the glaucoma hemifield test (GHT). However, although promising sensitivity and specificity have been reported, they are dependent on repeating the test,[60] with the aim of achieving a confirmatory result. Yet again, there are cost implications for both the health service and the patient of the need to duplicate an examination.

Problems faced by the NHS

Glaucoma places an enormous burden on the NHS: in 2001, POAG affected an estimated 356,000 people in England and Wales.[61] The chronic nature of the condition and need for life-long treatment mean that glaucoma patients account for between 25 and 40% of all HES outpatient visits.

Referral and follow-up of glaucoma patients and suspects

Currently, the majority of cases of suspect glaucoma are detected by community optometrists.[62] Patients are then referred via their GP to the HES. Patients are usually subsequently managed in consultant-led clinics, although in some regions other management systems have been established.[63]

Following diagnosis, patients are typically reviewed every 6–12 months.[64] Visual field tests are performed at least annually on most patients. Specific patients may have more frequent examinations if it is indicated by their clinical condition. In addition, supplementary field tests are often performed when a patient appears to have shown perimetric deterioration. Supplementary/confirmatory field tests are often required owing to the large amount of variability inherent in perimetry.[65,66] Perimetric variability is of particular significance in the glaucoma clinic, since it is even higher in glaucoma patients than normals.[50] Additional indications for extra field tests are when the patient fails to perform the test adequately (e.g. owing to tiredness or anxiety) or the result was spoiled by a testing artefact.

Glaucoma patients versus suspects

Clearly, all individuals who are referred to the HES are not subsequently diagnosed with glaucoma: some transpire to be false positive referrals and others are diagnosed with ocular hypertension. In both of these scenarios, the patient may require multiple HES visits to confirm their status. The number of hospital episodes that these patients undergo varies widely between hospitals and between specialists but, in many departments, two or three field tests and clinical assessments are standard practice.

In some situations, it is very difficult, with current clinical techniques, to determine whether particular patients actually do have glaucoma. Some individuals repeatedly produce field tests with equivocal results, whereas others have idiosyncratic optic nerve heads that can be very difficult to interpret. In these circumstances, the patients are often monitored for long periods with re-examination every 6 or 12 months, which further compounds the workload of the glaucoma clinic.

Increased workload with an ageing population

The age-related prevalence of glaucoma,[17,20,67] coupled with demographic changes in the UK population (the number of people over pensionable age is projected to increase from 11.2 million in 2006 to 11.9 million in 2011 and to 13.1 million by 2021[68]), mean that the already considerable burden associated with the management of the condition is certain to rise.

Summary

Glaucoma is a common condition that accounts for a significant proportion of the outpatient workload within the HES. Current methods of clinical examination require that patients make lengthy and repeated visits to the hospital that may still not result in a definitive diagnosis. An ideal component in the clinical armoury would be an instrument that is capable of achieving rapid, objective examination with high sensitivity and specificity and low measurement variability. This report summarises research performed on two pieces of equipment, the Heidelberg Retina Tomograph (HRT) and laser diagnostic glaucoma scanning system (GDx), to determine their potential to fulfil this role.

Chapter 2

Instrumentation

This chapter outlines two instruments: the HRT and GDx and discusses the potential each has to detect and monitor glaucoma.

Some of the ideal properties of a clinical test, that may be useful in the glaucoma clinic, are listed below. This report describes the investigation of many of these variables:

- high sensitivity and specificity
- objective examination
- independent of operator effects
- widely applicable
- highly repeatable
- easy to use
- cheap, reliable, portable
- rapid examination with easy analysis and interpretation of results
- non-invasive test with no need for mydriasis
- ability to compare results with baseline
- well-established diagnostic algorithms
- good quality normative database.

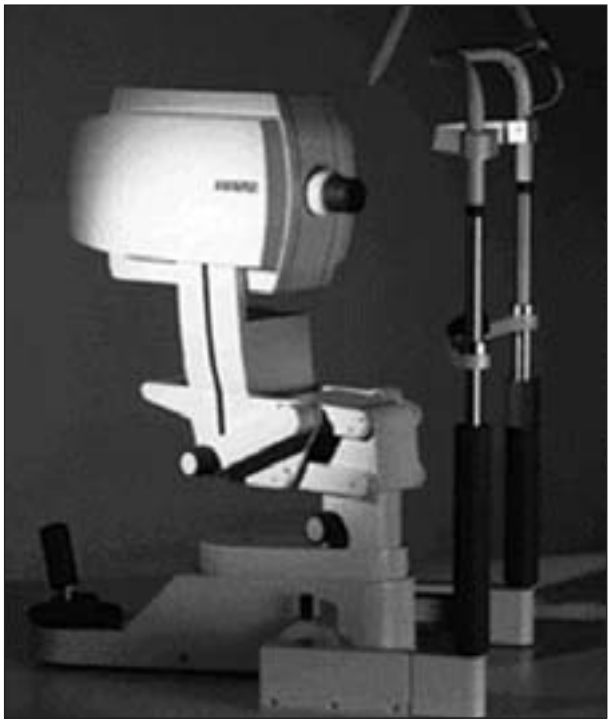


FIGURE 1 *The Heidelberg Retina Tomograph*

HRT

The HRT (Heidelberg Engineering, Germany) is a confocal scanning laser ophthalmoscope. The instrument (*Figure 1*) uses a 670-nm diode laser with a maximum output of 180 μW to acquire images of the posterior segment of the eye.[69,70] For the purposes of the diagnosis and monitoring of glaucoma, the HRT is used to generate three-dimensional topographic images of the optic nerve head. Two models of the HRT are currently available: the HRT I and HRT II. All research described in this report was conducted with the former. The latter was developed and released while the research described in this report was under way; it has been designed to be more 'user friendly' and is considered more suited to the clinical environment, whereas its predecessor is more appropriate for ophthalmic research.

Specification

Rotary controls on the operation panel can be used to select the focus setting [from +11.75DS to –11.75DS in 0.25DS increments (DS = dioptre sphere)], scan depth (from 0.5 to 4.0 mm in 0.5-mm increments) and scan width (10° × 10°, 15° × 15° or 20° × 20°). Intensity and sensitivity controls govern image brightness. Image acquisition and capture are controlled by 'Freeze' and 'Record' buttons, respectively. The duration of image capture is 1.6 seconds.

Following image capture, the image series, which consists of 32 planes, comprising 256 × 256 pixels, is stored in memory. The HRT software then processes this information into the intensity and topography images (as shown in *Figure 2*). (All sample images in this chapter are taken from the same eye to allow comparison to be made between output modes.) The topography map is a colour-coded topographic image of the optic disc: darker colours represent prominent structures, whereas more depressed areas are shown in lighter colours. Multiple topographies can be combined to form a mean topography image, from which the stereometric parameters are subsequently taken.

Throughout the glaucoma imaging study, the HRT software has been updated as successive versions have become available. Analysis that has

reviewed aspects of diagnostic precision (see Chapters 7–9) was performed with the latest Windows-based software (version 1.4.1.0). The predecessors to this version were DOS-based and, consequently, far more time consuming and cumbersome to use. Software improvements have been made not only to the user interface, but also to the method of outlining the margin of the optic disc (*Figure 3*) and to the software algorithm responsible for the alignment of single images when a mean image is composed. The latter has been shown to reduce significantly the variability in the mean image.[71]

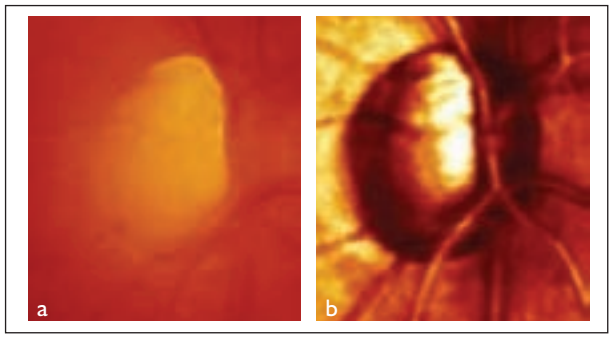


FIGURE 2 *HRT images. (a) Topography image; (b) intensity image. [This figure is shown in colour on the CD and on the website.]*

Operation

After activation of the laser, the illuminating beam is directed through the pupil; the operator then adjusts the scan settings until the optimum parameters are ascertained. Multiple scans are taken until the user is satisfied that the optimum scan settings have been achieved – this is attained by a combination of personal experience and the feedback messages that the instrument gives after the capture of a suboptimal image. Good-quality images are saved to the computer's hard disk. The issue of image quality is discussed further in the section 'HRT' (p. 18).

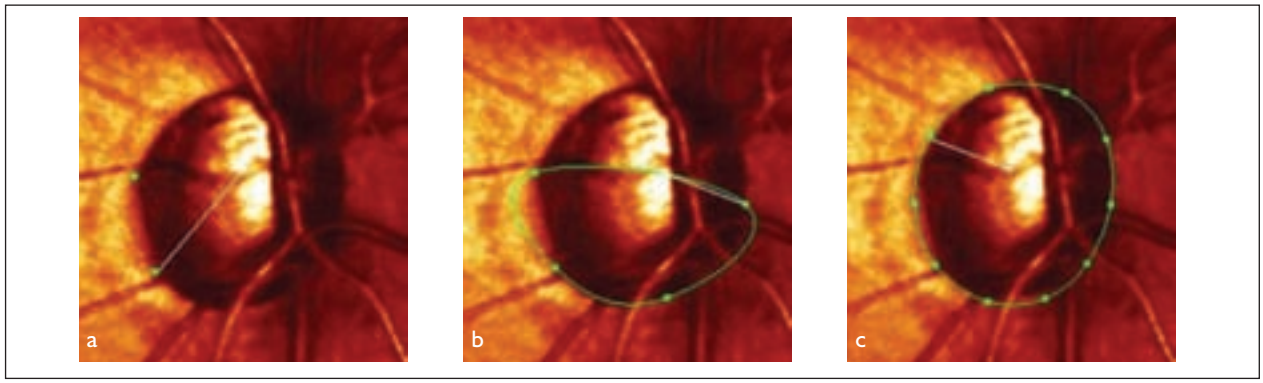


FIGURE 3 *Method for identification of optic disc margin. (a) Beginning, (b) middle and (c) end of the process. [This figure is shown in colour on the CD and on the website.]*

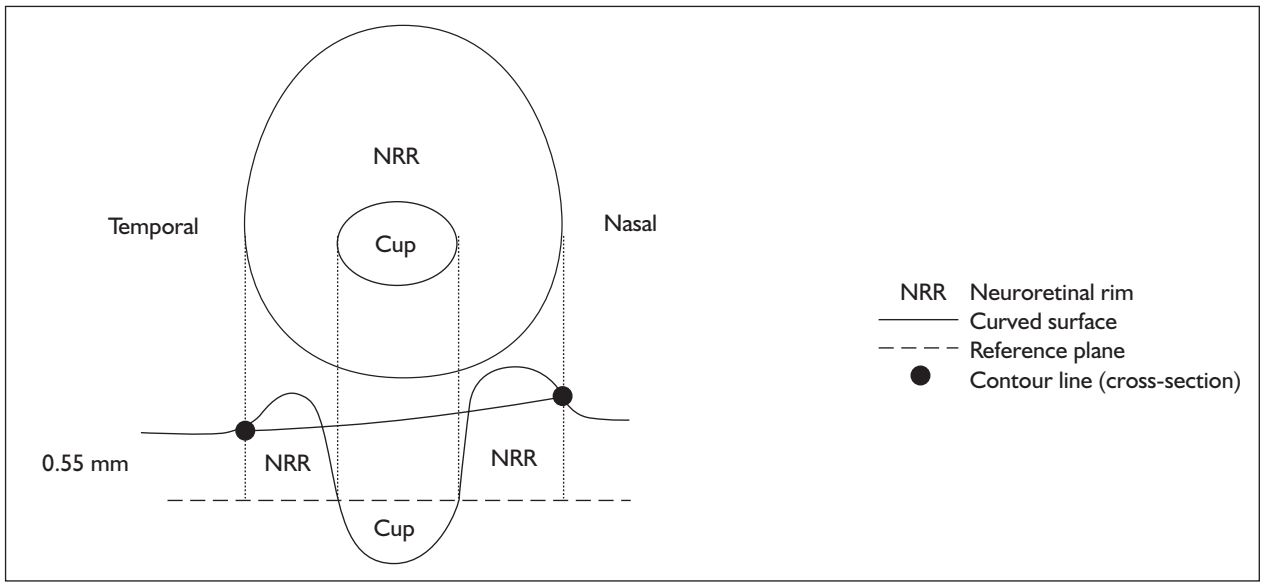


FIGURE 4 *Relative positions of HRT reference plane, curved surface and contour line*

Stereometric parameters

The stereometric parameters are calculated with respect to two reference surfaces: the reference plane and the curved surface (see *Figure 4*). All analysis described in this report was conducted via the default location of the reference plane (i.e. 50 μm below the retinal surface between –4° and –10°, where 0° is the temporal location and negative angles denote inferior locations).[72] The HRT software uses the height of the reference plane to differentiate the cup from the neuroretinal rim (NRR); the cup is defined as lying below the reference plane and the NRR as lying above. *Figure 5* shows the topography map after division by the reference plane. The curved surface subclassifies the NRR according to whether it is 'ascending' or 'stable'.

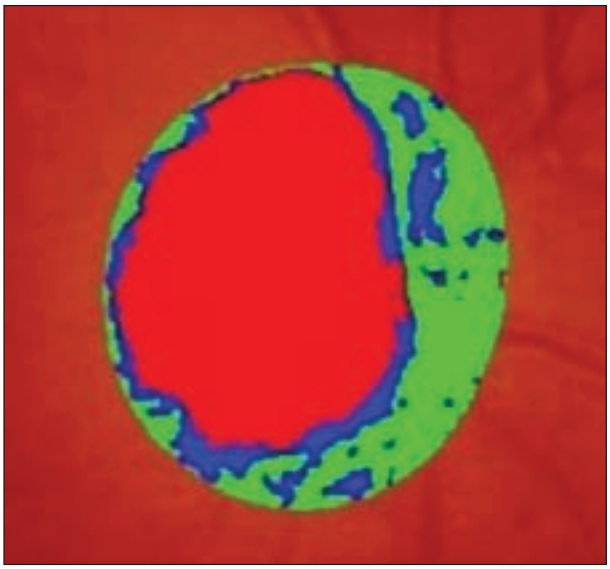


FIGURE 5 *HRT topography image. [This figure is shown in colour on the CD and on the website.]*

Table 2 gives the definition of a variety of HRT stereometric parameters;[73] the instrument computes both global and sectoral values. In addition to features with which the clinician is familiar (e.g. C:D ratio), an additional parameter expresses the 'steepness' of the excavation of the disc. This is quantified by the parameter cup shape measure, which is derived from the third central moment of the frequency distribution of depth values within the contour line and below the curved surface. It outputs a value between –1.0 and +1.0 (zero indicates a symmetric distribution of excavation depth measurements).[74] When applied to the optic nerve head, a shallow cup will give a negative value whereas a deep cup will yield a positive value

Reproducibility and variability

High repeatability and low variability are desirable attributes of a clinical test. HRT measurements of the optic disc have been shown to be highly reproducible;[75–77] for example, Mikelberg and colleagues found coefficients of reproducibility to range from 60.6 to 99.4% for different stereometric parameters.[78] Their findings corroborate those of other researchers that measurement variability is greater in glaucoma patients than normal individuals.[77–79] Other sources of increased variability in HRT measurements are accommodation,[80] incorrect alignment of the laser beam with the optic disc[81,82] and variations in the location of the reference plane.[83]

Factors affecting image quality

Theoretically, the technique of confocal scanning laser ophthalmoscopy (CSLO) is less likely to require pupillary dilation and is less sensitive to media opacities than other techniques of fundus imaging.[84,85] However, pupil diameter and clarity of the crystalline lens do affect image quality.

TABLE 2 *Definition of HRT stereometric parameters*

Parameter	Definition
Disc area (mm^2)	Total area within contour line
Cup area (mm^2)	Area below reference plane
Cup:disc area ratio	Cup area/disc area
Rim area (mm^2)	Area above reference plane
Height variation contour (mm)	Height variation of retinal surface around contour line
Cup volume (mm^3)	Volume below reference plane
Rim volume (mm^3)	Volume above reference plane
Mean cup depth (mm)	Mean depth inside contour
Maximum cup depth (mm)	Maximum depth inside contour
Cup shape measure	Measure for overall three-dimensional shape of cup
Mean RNFL thickness (mm)	Mean distance along contour line between retinal surface and reference plane
RNFL cross-sectional area (mm^2)	Mean RNFL thickness multiplied by the length of the contour line

It has been suggested that eyes with pupil diameter <3 mm and/or media opacities will benefit from pupil dilation.[86] Experience acquired during the period of this study has led to the adoption of a strategy where undilated imaging is attempted on all eyes, since in each situation, the exact effect on image quality of cataract morphology, cataract density and pupil diameter cannot easily be predicted. In cases of unclear media, a larger pupil diameter is associated with improved image quality.

Sensitivity and specificity

There are two fundamental conclusions from the vast body of literature that has investigated the diagnostic precision of the HRT parameters. First, combinations of parameters are superior to single measurements in separating glaucoma patients from normal individuals[87–89] – such algorithms are discussed in the next section. Second, the diagnostic ability of an instrument that assesses the optic nerve head is significantly compromised by the large amount of overlap in the appearance of the structure in normal and glaucomatous eyes.[90]

There is some degree of commonality between different research groups in the single parameters that perform well at diagnosing glaucoma. Features of the peri-papillary contour line,[36,90,91] cup shape measure[89,91,92] and cup volume[90,93] have all been identified for their superior diagnostic ability. However, various factors confound the comparison of results from different studies; these factors include the stage (and type) of disease, different inclusion criteria[94] and definitions of normality and abnormality.

Diagnostic algorithms

In 1995, Mikelberg and colleagues described a forward-stepping discriminant function analysis (DFA), which employed a combination of stereometric parameters to diagnose glaucoma.[87] The DFA used the parameters cup shape measure, rim volume and height variation contour. The procedure was validated using a jack-knife procedure in which categorisation functions were created from the entire data set, but excluded the case being tested. The DFA produced sensitivity and specificity values of 87% and 78%, respectively. Another DFA was described by Bathija and colleagues;[88] their formula used the parameters cup shape measure, rim area, height variation contour and RNFL thickness and yielded sensitivity and specificity values of 78% and 88%, respectively.

One of the newer forms of data analysis available with the HRT is the Moorfields Regression Analysis (MRA) described by Wollstein and colleagues in 1998.[95] They showed sensitivity and specificity of 96% and 84%, respectively, from the linear relationship of optic disc area and the logarithm of NRR area. MRA performs the analysis for both global and sectoral values and compares its results with a data set of 80 normals. Three diagnostic classifications are given: 'within normal limits' (results lie within the 95% prediction interval of normal); 'borderline' (between 95 and 99.9%); and 'outside normal limits' (beyond 99.9%).

Subsequent research that has investigated these diagnostic algorithms has always yielded lower diagnostic precision than the original research. This apparent disparity in diagnostic precision is most likely due to variations in test populations.[36,96,97] The outcome of the application of the three diagnostic algorithms described in this section to our study population is described in Chapter 7.

Longitudinal research

Longitudinal follow-up (3.0 ± 1.5 years) of glaucoma patients has shown that scanning laser tomography is able to detect progression prior to perimetric change.[98] A significant loss of rim area and an increase of both cup area and cup volume were detected in eyes that had undergone clinical progression. If the HRT is to be used for longitudinal follow-up, compensation must be made for variability both during image capture and in the placement of the peri-papillary contour line. Tan and Hitchings proposed a method of exploiting this shortcoming, by measuring the extent of variability, over time, of a control group.[99] Patients within their glaucoma group who showed variability outside these limits were considered to have undergone structural change of the optic nerve head. A positive result in two sequential tests had a sensitivity of 90% and specificity of 97.2%.

Kamal and colleagues reviewed two sets of images spaced 16–21 months apart from patients with ocular hypertension who were considered at risk of developing glaucoma.[93] They found that 62% of eyes that had undergone perimetric conversion showed glaucomatous change prior to confirmed visual field defects. In the group of eyes with ocular hypertension, 29% showed glaucomatous changes in global and segmental parameters that were outside the expected amount of variability. This study highlights the potential use of the HRT in identifying those at risk of conversion. A criticism of the methodology is that normal variability over the research period was established by the examination of only 21 normal eyes (with

pathological change defined as a difference lying outside the 95% confidence limit).

Chauhan and colleagues proposed a method of serial analysis of height measurements from topographic images, using variability estimates for 'superpixels', each comprising a square of 4×4 pixels.[100] The method determines locations where significant change from the baseline has occurred and assigns probabilities to the magnitude and temporal aspects of the change.

There is a scarcity of reports of large-scale longitudinal studies. Research described in this report will, therefore, make an important contribution to the body of knowledge.

GDx

The GDx (Laser Diagnostic Technologies, USA, now manufactured by Carl Zeiss Ltd, Oberkochen, Germany), is a scanning laser polarimeter that is designed to measure the peri-papillary RNFL. The instrument (*Figure 6*) uses a diode laser with a wavelength of 780 nm, whose state of polarisation has been modulated. The GDx utilises the fact that the RNFL is birefringent, that is, it changes the state of polarisation of a beam of light. A linear relationship has been shown to exist between retardation (amount of change in state of polarisation) and RNFL thickness.[101] However, there is a major confounding factor in inferring RNFL thickness from ocular retardation. The cornea and, to a small extent, the crystalline lens are also birefringent. The GDx assumes that the birefringent properties of the cornea are the same for all patients and, therefore, makes uniform compensation when generating RNFL thickness measures. This issue is discussed in greater detail in the section 'Corneal polarisation' (p. 13).

The model of the GDx that was used in this study is no longer manufactured; the reason is that significant changes to both the hardware and software have been necessary to allow individualised adjustments in corneal compensation to be made. Our instrument cannot be updated. The contemporary version of the equipment was initially marketed as the GDx Access 3000, and is now known as the GDx VCC, so-called because it is capable of providing **v**ariable **c**orneal **c**ompensation.

Specification

The GDx examines an area of $15° \times 15°$, centred on the optic disc, and each image consists of 256×256 pixels. Each pixel is measured with 20 different states of polarisation, from which the maximum retardation value is calculated. Instrument controls govern focus and gain, which are adjusted prior to image capture, which takes approximately 0.7 seconds. The GDx software has been upgraded as new editions have become available. Research that has investigated the diagnostic precision of the GDx (Chapters 7, 8 and 9) was all performed with the most recent software, version 2.0.0.9.

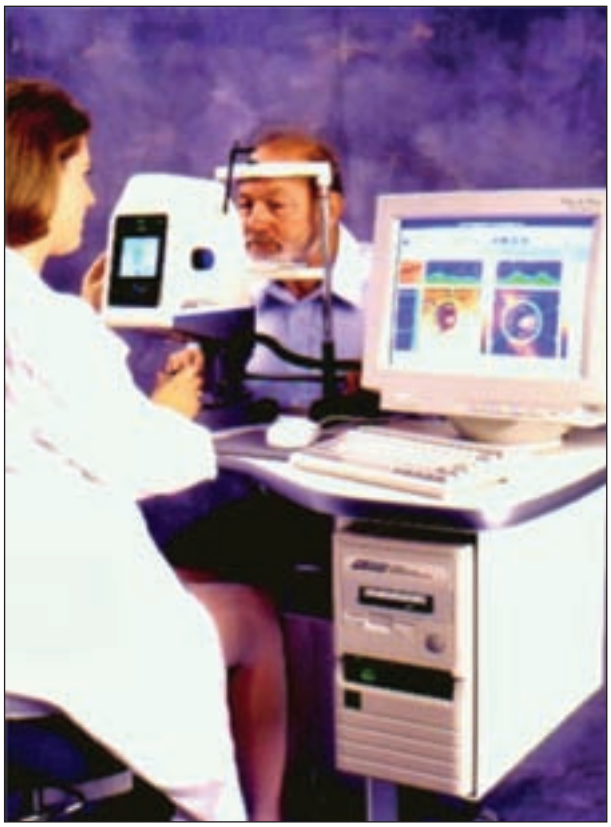


FIGURE 6 *The GDx [This figure is shown in colour on the CD and on the website.]*

Operation

The instrument is aligned with the pupil centre, using semicircles of light projected on to the cornea. Centration of the optic disc on the viewing screen is achieved by adjusting the position of gaze of the contra-lateral eye with an external fixation target. The focus and gain settings are adjusted and the image capture button is pressed to acquire an image. A 'quality score' is calculated (maximum = 100) for each image, taking into account intensity, centration, evenness of illumination, lack of vignetting and contrast, and also two other parameters that examine the relative thickness of a combination of quadrants. The operator is given feedback as to whether the

image has 'passed' or 'failed'. Pupil dilation is not required and is stated as being undesirable on account of it potentially conflicting with the corneal compensator.[102]

Graphical and numerical outputs

After processing, two main images are displayed (*Figure 7*): a reflectance image of the optic nerve head and colour-coded map of RNFL thickness; the latter is calculated by Fourier analysis of retardation values. Definitions of the GDx parameters are shown in *Table 3*; however, this list does not detail all of the extensive number of parameters. New parameters have been described[103] and appear to show promising diagnostic precision[104] but, at the time of writing, are not available with the current version of the software.

For many of its parameters, the GDx analysis uses ratios of values obtained from different regions of the RNFL to reduce intrasubject variability. A variety of the parameters utilise the mean of the median 1500 pixels in the nasal and temporal quadrants and the mean of the 1500 thickest pixels in the superior and inferior zones. These quantities have been selected to reflect the distribution of the superior and inferior arcuate bundles of nerve fibres. In the lateral quadrants, the median value is considered optimal to prevent the inclusion of any values from the RNFL bundles and to exclude the noise associated with the thinnest pixels.

The parameter known as the number is derived from a two-layered back-propagation neural network analysis (a fundamental feature of neural networks is that they are not programmed, but are trained by example; the system is educated by presenting a series of data together with the anticipated outcome; the neural network formulates a relationship between input and output data) of 128 parameters that include (but are not limited to) the results that are presented to the operator. Precise details of the experimental group are not available, but a brief description is

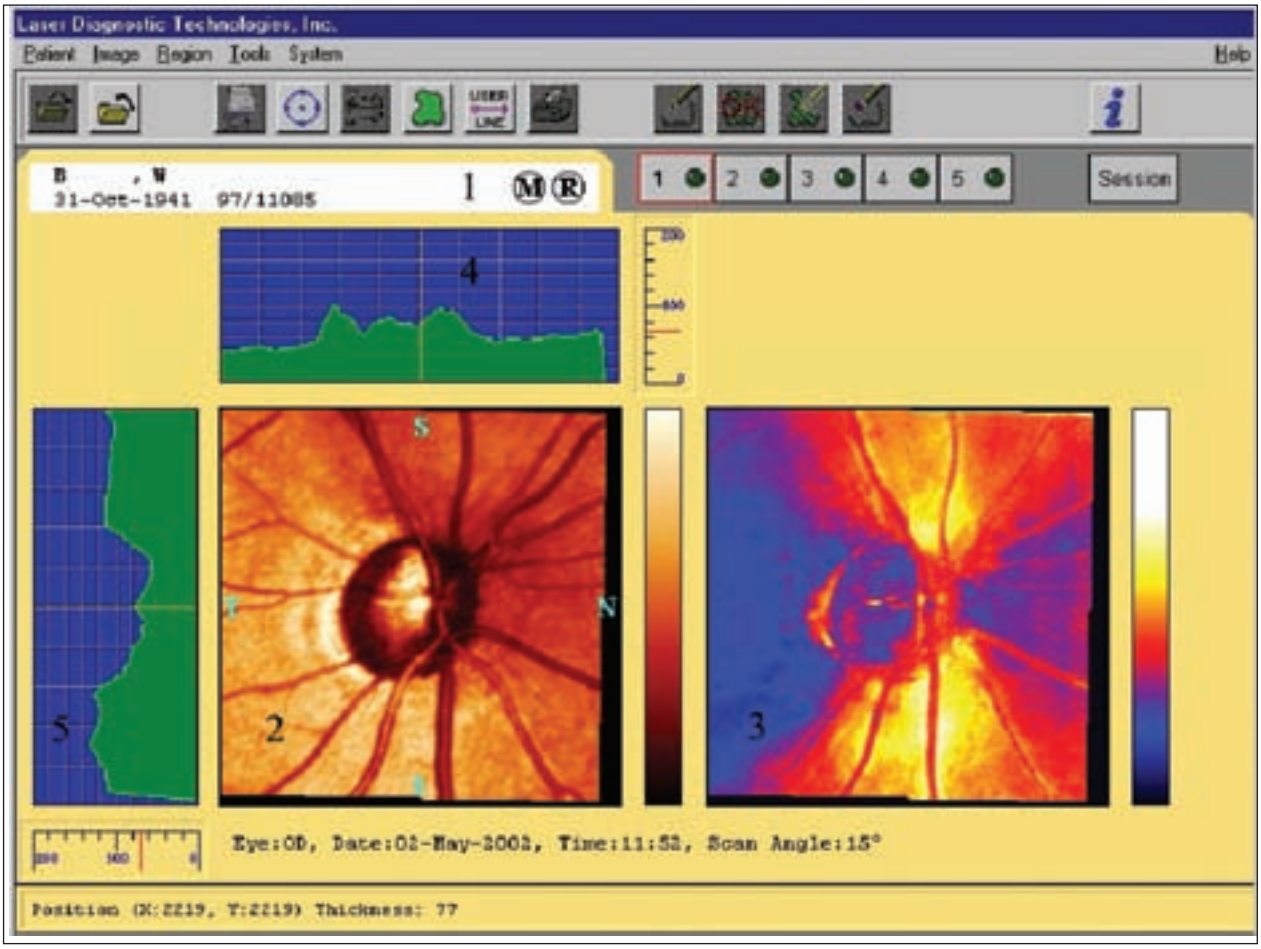


FIGURE 7 *GDx output following image capture. 1, patient details; 2, intensity plot; 3, retardation plot; 4, horizontal profile; 5, vertical profile. [This figure is shown in colour on the CD and on the website.]*

TABLE 3 *GDx parameters*

Parameter	Definition
Symmetry	Ratio of mean of 1500 thickest pixels in superior quadrant to mean of 1500 thickest pixels in inferior quadrant
Superior ratio	Ratio of mean of 1500 thickest pixels in superior quadrant to mean of 1500 median pixels in temporal quadrant
Inferior ratio	Ratio of mean of 1500 thickest pixels in inferior quadrant to mean of 1500 median pixels in temporal quadrant
Superior/nasal ratio	Ratio of mean of 1500 thickest pixels in superior quadrant to mean of 1500 median pixels in the nasal quadrant
Maximum modulation	Difference between the thickest and thinnest areas of RNFL
Ellipse modulation	Same as maximum modulation but around the ellipse
The number	Derived from neural network (see below)
Average thickness	Mean thickness of all pixels outside optic nerve head
Ellipse average[a] (μm)	Mean thickness of RNFL beneath ellipse
Superior average (μm)	Mean thickness of RNFL beneath superior portion of ellipse
Temporal average (μm)	Mean thickness of RNFL beneath temporal portion of ellipse
Inferior average (μm)	Mean thickness of RNFL beneath inferior portion of ellipse
Nasal average (μm)	Mean thickness of RNFL beneath nasal portion of the ellipse
Superior integral (μm^2)	Total area of RNFL along superior portion of ellipse
Superior maximum (μm)	Mean of 1500 thickest pixels in superior quadrant
Inferior maximum (μm)	Mean of 1500 thickest pixels in inferior quadrant
Total integral (μm)	Total area of RNFL beneath ellipse around optic disc
Superior integral (μm)	Total area of RNFL beneath superior portion of ellipse around optic disc
Inferior integral (μm)	Total area of RNFL beneath inferior portion of ellipse around optic disc
Nasal integral (μm)	Total area of RNFL beneath nasal portion of ellipse around optic disc

[a] Also known as total polar average.

available (Rennie-Cassell JR, Laser Diagnostic Technologies, California, USA: personal communication, 2001). One eye from 197 normal individuals and 51 age-matched glaucoma patients was imaged three times with the GDx and a mean image created. A subset of subjects (130 normal individuals and 34 glaucoma patients) was used to train the network. Training was done until the answers were 95% correct. The remaining subjects were used to test the network, achieving 94% sensitivity and 97% specificity. A limitation of this methodology is that the final sensitivity estimate was derived from only 17 glaucoma patients, which may limit its generalisability.

The GDx number ranges from 0 to 100 for each eye, where 0 indicates total normality and 100 represents advanced glaucoma. Guidelines state that a result ≤ 30 is within normal limits and ≥ 71 is indicative of glaucoma. A score lying between 31 and 70 suggests a suspect case.[105] A threshold of 27 is thought to be superior for differentiating between normal and glaucoma,[106] whereas other research has shown 35[107] and 39[108] to be the optimum threshold. The 'number' indicates a probability of the diagnosis and is not a linear scale suitable for monitoring progression. The manufacturers stress that this parameter is experimental and is currently under evaluation, although it seems somewhat paradoxical that a third-generation instrument with wide commercial availability contains a feature that is still considered experimental. A further limitation of this parameter is its inability to detect localised RNFL defects, because it was not trained to do so.[109]

Reproducibility and variability

Reproducibility of the GDx has not been investigated as thoroughly as the HRT [much of the research was performed with the Nerve Fiber Analyzer (NFA) I and NFA II, which are the predecessors to the GDx]. Colen and colleagues found variable consistency, with straight parameters appearing more robust than ratio values.[110] They found that most parameters had good repeatability with intraclass correlation coefficients >90%.

Factors affecting image quality

More than a 10% change in GDx parameters was found in 20% of eyes following dilation.[111] The authors concluded that a consistent approach should be taken for patients undergoing longitudinal follow-up, which is in line with the approach taken in this study. Miotics are thought to have little effect on image quality,[112] although personal experience has found that it can be difficult to image eyes with pupil diameters

<1.0 mm, although this is also dependent on media clarity. Eye movements during image capture have been shown to cause erroneously thick RNFL results.[113] The authors recommend deleting any images containing motion artefacts.

It is estimated that media opacities allowing visual acuity of 6/60 or better yield adequate image quality.[109,114] However, the authors do not discuss cataract morphology, although posterior subcapsular cataract is particularly implicated by other researchers.[115] A recent study found that removal of cataract resulted in greater absolute measurements of RNFL thickness but ratio values were unchanged. Scanning laser polarimetry (SLP) measurements can change significantly after cataract extraction and may necessitate the establishment of new baseline measurements.[116]

Experience from using the HRT and GDx for 6 years has shown that pupil dilation is required much less frequently for GDx examination than HRT. The GDx seems less sensitive to media clarity and small pupil diameter. Aspects of this difference are discussed further in Chapter 6.

Normative database

When the Glaucoma Imaging Study began, a novel and unique feature of the GDx was its normative database. Subsequently, such features have become more frequently incorporated into other imaging devices (e.g. HRT, STRATUSOCT optical coherence tomographer). Values from the normative database are used to make age- and race-specific comparisons when a patient is examined.

Data for the GDx normative database were collected from a multicentre study (six centres) in the USA and Europe. Although the precise details of the population examined are not available, outline information of the study group is available (Rennie-Cassell JR, Laser Diagnostic Technologies, California, USA: personal communication, 2001). Data were collected on a total of 400 eyes from up to 400 persons; when both eyes from an individual satisfied inclusion criteria, then data from both eyes were included. Individuals ranged in age from 18 to 80 years with approximately 60 persons in each 10-year age bracket. Briefly, inclusion criteria were: refractive error <5.00DS of ametropia and <2.00 DC
(DC = dioptre cylinder) of astigmatism; no history of ocular disease; no family history of ocular hypertension or glaucoma; visual acuity of 20/40 or better; no visual field defect with the Humphrey 24-2 full threshold programme; IOP ≤21 mmHg; and no abnormality of the optic disc on ophthalmoscopy. Elsewhere, it is stated that eyes with cup to disc ratio >0.5 were also excluded.[102] Each person had three images taken with the NFA and was realigned after each measurement to ensure independence of the readings.

These protocols raise several issues: the inclusion of both eyes of some patients; the constraints on refractive error when the instrument is clearly capable of examining patients well outside this range; and the fact that the first-generation instrument was used to collect data for its third-generation counterpart. Additionally, the potential consequence of excluding patients with physiological cupping, when these are the patients for whom GDx results may be clinically important. It is also stated that further data have been collected and added to the normative database (Rennie-Cassell JR, Laser Diagnostics Technologies, California, USA: personal communication, 2001) but supplementary information is not given.

Sensitivity and specificity

Over the last 6 years, many papers have been published concerning polarimetry; of these, the most frequently quoted gives sensitivity and specificity of 96% and 93%, respectively.[117] Although the research was performed with the NFA, it is included in this report since the results are very widely quoted and the methodology is subject to several criticisms. The experimental group was not stratified according to the extent of visual field loss and the results may have been biased by the inclusion of a large number of patients with advanced glaucoma. In addition, the authors used a calculation method that employed a previously described 'squares method'.[118] This technique differs from the default GDx analysis and does not relate directly to the currently available GDx parameters. A further criticism of the methodology is that the same patient group was used for both the analysis and establishment of the cut-off criteria; this approach overestimates test performance.

Paczka and colleagues found sensitivity and specificity of 62% and 96%, respectively, for the number,[119] whereas the estimates from other research are 82% and 80%, respectively.[120] Poinoosawmy and colleagues found very promising values of 92% and 96%, but their experimental group comprised a large number of cases with very advanced glaucoma.[108] Zangwill and colleagues[36] agreed with other researchers[89] that the number featured as one of the GDx's optimum

parameters and found 41% sensitivity at 90% specificity for this parameter. Issues such as inclusion criteria and the subsequent effects on apparent diagnostic precision are discussed in greater detail in the introduction to Chapter 7 (p. 47)

Of the other GDx parameters, some are considered to be of minimal utility and may give clinically misleading information when they are flagged as outside normal limits.[121]

Corneal polarisation

The GDx compensates for corneal and lenticular polarisation. This compensation is fixed with the slow axis of corneal birefringence 15° nasally downward and a magnitude of 60 nm.[122] The magnitude and direction of polarisation have been confirmed experimentally, but the measurements have a wide range of inter- and intraindividual variability.[123] Improved diagnostic accuracy has been shown when the corneal compensation is individualised.[124–128]

The most contemporary version of the GDx, the GDx VCC, is capable of customising the magnitude and direction of the corneal compensation for each patient. This adjustment is achieved by the capture of an image from the macular region at the baseline examination. Lenticular polarisation is rarely discussed in the literature, although it has been suggested that the birefringence of lens proteins may affect SLP results.[129] However, with the GDx VCC, the origin of the polarisation is probably immaterial, since the instrument adjusts RNFL measurements for overall anterior segment birefringence.

Since all data presented in this report were collected with the original GDx, prior to the development of the GDx VCC, there may well be some disparities between our results and those obtained with the GDx VCC. Unfortunately, it is not possible to quantify this effect. However, it is likely that some parameters will be more susceptible than others. It has been proposed that the number is reasonably robust; this is likely to be because the parameter expresses a statistical likelihood rather than giving a direct measurement of RNFL characteristics. Similarly, GDx ratio parameters should be relatively unaffected since, owing to the nature of the reading, they will be less susceptible to anomalous values of corneal polarisation.

Longitudinal research

Agreement between GDx measurements and visual field results in patients who were reviewed over 2 years has been already been reported.[130] However, the study was limited by the short period of follow-up and the lack of criteria for establishing progression.

Researchers from Moorfields showed alteration in GDx results in patients with normal tension glaucoma who were also reviewed over a period of 2 years.[131] During their study, they observed differences in the rate of RNFL loss between glaucoma patients and normal individuals; global RNFL thickness showed the greatest loss, compared with sectoral values.

As is the case with the HRT literature, there is a definite lack of major large-scale longitudinal GDx studies. The aim of the glaucoma imaging study is to address this issue.

Summary

The HRT and GDx both have the potential to be useful tools in the glaucoma clinic. They each provide rapid, objective and repeatable examination and previous research indicates that they have reasonable diagnostic precision for detecting glaucoma.

This report investigates the possible role of the instruments in long-term monitoring of glaucoma patients and suspects.

Chapter 3

Glaucoma Imaging Study

Within Manchester Royal Eye Hospital (MREH), the study described in this report became known as the 'Glaucoma Imaging Study'. This chapter describes many aspects of the day-to-day running of the study, the groups of patients examined, methods of examination, staffing of the study and problems encountered during the research. Results are also given for the number of subjects in each experimental group and the length of follow-up of patients under longitudinal review.

Introduction

The glaucoma imaging study has been run by the University of Manchester at MREH; the latter is the largest provincial eye hospital in the UK. Data collection for the study commenced on 28 May 1998 and continued until January 2004. The data collection period was extended beyond that originally planned to compensate for data loss [see the section 'Data loss' (p. 24)] that occurred during the early stages of the project. An application for a 1-year extension was submitted in August 2002 and subsequently granted.

The study was granted approval by the Manchester Health Authority Research Ethics Committee (Central) – reference CM/97/182. An amendment to this was later sought to allow a subgroup of patients to undergo photography of the retinal nerve fibre layer.

Within the study, two groups were studied cross-sectionally:

- a group with a clinical diagnosis of glaucoma
- a normal control group;

and two groups longitudinally:

- a group clinically diagnosed as high risk
- a group clinically diagnosed with glaucoma.

Further details of the recruitment and assessment of these patients are given in subsequent sections.

Evolving concepts and definitions

A fundamental problem with long-term research is that technology, views, opinions and definitions may well change whilst the study is under way. The original concept of this study was to categorise patients as 'glaucomatous' or 'normal' on the basis of the appearance of their ONH and/or visual field status. The definitions used were in accordance with a committee of ophthalmologists specifically convened for the purpose of developing such criteria by Dr P Sample, Glaucoma Center and Research Laboratories, University of California, San Diego, CA, USA (Dr Sample is a well-established expert in this field).

However, subsequently, an important issue arose: if the sensitivity and specificity of a device are to be investigated, then features of the structure that they measure should not be used as diagnostic criteria, in case sensitivity estimates be artificially elevated. The consequence of this concept for the glaucoma imaging study is that ONH and RNFL characteristics should not be used to categorise patients as normal or abnormal. This issue has been described in the literature[94] and, latterly, has become widely accepted.

The study adopted the strategy of categorising patients solely on the basis of their visual field results and not on the appearance of the ONH or RNFL. This approach highlighted a further issue of which visual field index to use as the classifying criterion. The HFA generates its own indices (e.g. the GHT), which has been shown to have very good diagnostic precision for glaucoma when repeat measures are considered.[60] However, the limitation of the GHT is that it provides a categorical output: within normal limits, borderline or outside normal limits. Although this may be useful for a clinical classification of patients, it is less appropriate for research, which would benefit from a quantitative output. The other HFA indices [e.g. mean deviation (MD), pattern standard deviation (PSD)] are unsuitable for monitoring subtle deterioration since, although they quantify visual field loss, they do so for the whole visual field. Global measures have been shown to be insensitive to early glaucomatous field loss, which often produces areas of localised damage, whose presence is diluted when values are averaged over the whole visual field. Other systems [e.g. the method used

by the Advanced Glaucoma Intervention Study (AGIS)][132] have been described. These techniques are normally based on an empirical set of rules and are not readily applicable to very large data sets since they require meticulous inspection of the printouts from the HFA with documentation of the number and location of symbols which indicate the severity of visual field damage. To circumvent these problems, a system of analysing perimetric results has been written by the lead applicant [see the section 'Analysis of visual field data' (p. 20)].

A further modification in scientific thinking that occurred since the original study was designed is a change in attitude towards visual field reliability indices. In brief, during a visual field test, the program performs infrequent catch trials in an attempt to estimate patient reliability. Previously, it was thought prudent to exclude from analysis any visual field test that exceeded the widely used criteria of 20% fixation losses and 33% false positives and false negatives. (Of interest, the scientific basis to these levels is unclear and has not been defined within the literature.) The reliability indices have been criticised, since infrequent sampling makes the confidence limits of the estimates very wide.[133] Subsequent research by Bengtsson[134] has indicated that reliability indices contribute very little to repeatability and, therefore, this study took the view of including all visual field test results, without considering the number of failed catch trials.

When the Glaucoma Imaging Study was initially designed, the experimental protocols indicated that visual field tests should be performed with HFA 24-2 full threshold program. At the time, this test was considered the 'gold standard' of perimetric examinations for glaucoma. However, subsequent to the beginning of the study, third-generation algorithms have been described, which aim to decrease testing time whilst preserving accuracy: the Swedish interactive threshold algorithms (SITA)[135] have become widely used in clinical and research environments. However, much of the work evaluating their reliability and repeatability occurred after the planning stages of the glaucoma imaging study.[136–141] SITA were not, therefore, used in the research described in this report.

Patient recruitment

Within the first 2 years of the study, considerable time was spent recruiting patients. The following two subsections detail the method of recruitment for hospital patients and normal control individuals, respectively.

Hospital patients

All patients in the cross-sectional, longitudinal high risk and longitudinal glaucoma groups were recruited from the glaucoma and general clinics at MREH. Two patients were enrolled from nearby Wythenshawe Hospital, where an outreach clinic is run by MREH. On several occasions, assistance with recruitment was requested from all the medical staff at MREH, which was done at the weekly regional postgraduate meeting.

Potentially suitable patients were identified by several methods:

1. Case notes were reviewed by the research workers and names and addresses of likely patients noted. All suitable patients were sent a letter, which explained the aims and methods of the study and invited them to participate. A prepaid envelope was enclosed, together with an acceptance form on which the patient could indicate convenient days/times for study appointments. [In order to optimise participation from as many patients as possible, appointments were available in the early mornings (prior to 8.00 a.m.), late evenings (up to 8.00 p.m.) and Sundays.]

   A total of 1499 letters were sent. The positive response rate was 32%. However, on examination, it transpired that not all these patients were suitable for recruitment. In some instances, inspection of the case notes had not disclosed that the patient was unsuitable for inclusion (e.g. owing to high refractive error, non-Caucasian race), so the final enrolment rate was somewhat lower than 32%.
2. Direct referral from ophthalmologists. Ophthalmologists at MREH occasionally made direct referrals to the study. When patients were recruited via this route, the enrolment rate was much higher, at around 85%. However, this method was not without complications. In some cases, ophthalmologists referred patients who they wished to discharge clinically but still wished to undergo eye-care follow-up. Some of these patients were not suitable for inclusion since they did not adequately fulfil the criteria for being 'at risk' of developing glaucoma on account of their IOP being too low or the fact that they were too young.

Normal control individuals

The original aim of the study was to examine, as normal control individuals, the partners of

patients attending the study. However, this strategy yielded an insufficient number of people. In some cases, partners were unwilling to participate, and others who did volunteer were themselves diagnosed with ocular pathology [see the section 'Normal control group' (p. 25)].

Enormous efforts were made on behalf of the research team to recruit normal control individuals. Strategies included:

- Fourteen local GPs were contacted, requesting access to their patient databases in order to target suitable patients.
- Leaflets and posters were left in waiting rooms of local GPs and dentists.
- Forty local optometrists were contacted, requesting assistance with recruitment.
- The clinic manager at the optometry department at the University of Manchester Institute of Science and Technology (UMIST) was contacted and asked to make appropriate referrals (at UMIST there is a Department of Optometry with which the University has an excellent working relationship).
- 'Professional patients' (individuals who regularly volunteer for clinics, assessments, examinations, etc.) at UMIST were invited to participate.
- All ophthalmologists at MREH were asked to assist with recruitment.
- All optometrists at MREH were asked to assist with recruitment.
- Leaflets and posters were placed in all clinic areas in MREH.
- Leaflets and posters were placed in local churches.
- Leaflets and posters were placed in local charity shops.
- An article about the study was placed in a local newspaper and individuals without any eye problems were asked to participate.
- E-mails were sent to all staff at both Manchester University and Central Manchester and Manchester Children's University Hospitals NHS Trust (of which MREH is a component). However, this system was only partially successful owing to regulations that restrict blanket contacting of staff. To overcome this problem, the research workers targeted a small number of individuals each day from the staff directory.

Again, problems were encountered: some potential control patients attended but had previously been suspected of having glaucoma and viewed the study as a way of achieving a 'definitive' diagnosis; indeed, some of these patients were subsequently diagnosed with glaucoma. In addition, a number of patients volunteered, but actually had refractive error that lay significantly outside the inclusion criteria and so were not eligible for recruitment [see the section 'Normal control group' (p. 25) for further details].

Patient groups

A total of four groups of patients were recruited and examined. Details are given in the next four subsections.

Cross-sectional patients

The original aim of this group was to review the use of the HRT and GDx on an unselected population of patients attending a glaucoma clinic. No other inclusion criteria (e.g. refractive error, coexisting ocular pathology) were imposed other than that the patient was aged ≥40 years, of Caucasian origin and had no secondary cause of glaucoma (e.g. pseudoexfoliation, pigment dispersion). (The restriction on ethnic origins of patients was imposed because of well-documented racial differences in both ONH and RNFL characteristics and the prevalence of glaucoma.[142–147]) Patients were recruited into the cross-sectional study irrespective of whether or not they were receiving medical treatment for glaucoma or had previously had surgical treatment. Patients with all stages of glaucoma were invited to participate in the study, that is, no consistent clinical diagnostic criterion for glaucoma was used. In order to adhere to ethical research standards, an additional criterion was imposed: patients were not invited to participate if their hospital notes indicated that they had problems with communication or memory. This is in accordance with the tenets of the Declaration of Helsinki.[148]

Normal control individuals

Patients were included as normal control individuals if they conformed to the following criteria:

1. Inclusion criteria
   (a) open angles of the anterior chamber
   (b) visual acuity better than 0.5 (logMAR)
   (c) spherical ametropia <5.00 DS and cylinder <3.00 DC
   (d) age ≥40 years
   (e) Caucasian race
   (f) no visual field defect.

2. Exclusion criteria
   (a) history of intraocular surgery (except uncomplicated cataract surgery)
   (b) current or previous intraocular eye disease

(c) 'general reduction' or 'too high sensitivity' on GHT
(d) history of systemic illness that might give rise to visual field defect, such as previous cerebrovascular accident.

Longitudinal high risk

1. Inclusion criteria
   (a) as normal control individuals, plus
   (b) untreated IOP ≥23 mmHg on two or more occasions
   (c) fellow eye with glaucomatous visual field loss.
2. Exclusion criteria
   (a) As normal control individuals, plus
   (b) history of intraocular surgery (except uncomplicated cataract surgery or uncomplicated trabeculectomy)
   (c) secondary causes of raised IOP (e.g. pseudoexfoliation, pigment dispersion, trauma, intra-ocular inflammation)
   (d) life-threatening illness.

Longitudinal glaucoma

Inclusion and exclusion criteria were as above, except all patients had repeatable glaucomatous visual field defects.

Data collection

The schedule of examinations for patients in the longitudinal groups was that they were examined twice within 1 month at baseline and, thereafter, every 6 months. This timetable was adhered to as strictly as possible. However, there were some instances, such as when patients had long distances to travel, that their appointments were more widely spaced. Many patients were only willing to attend for research appointments that were scheduled to coincide with their regular clinic visits. In these cases, postponement of routine clinic appointments for patients whose condition was stable frequently disrupted the pattern of examination for the study. Details of the length of follow-up are given in the section 'Longitudinal high risk and longitudinal glaucoma groups' (p. 25).

In order to investigate suspected perimetric deterioration further, a repeat visual field test was made for individuals who showed a change in category in their GHT analysis. In most cases, the supplementary examination was performed within 1 month.

At each visit, all patients underwent the following evaluations:

- full ocular history
- full threshold 24-2 examination with the HFA
- HRT imaging
- GDx imaging
- measurement of pupil diameter under the lighting conditions used for HRT examination
- logMAR visual acuity
- measurement of IOP (Goldmann tonometry).

At the baseline visit, the aims, objectives and methods of the study were thoroughly explained to all patients. In all cases, both the patient and researcher signed the ethics committee approved consent form.

Preliminary examinations

At the baseline visit, corneal curvature was measured with a manual keratometer (Bausch and Lomb) and spectacle prescription determined either by automated focimetry (AccuRx Lens Analyzer, Humphrey) of a recently dispensed pair of glasses (≤1 year old) or by auto-refraction (Ref R-1, Canon). For longitudinal patients, details of refractive correction were regularly updated – the information was taken from the results of patients' routine sight tests. Refraction and keratometry were repeated during follow-up if patients underwent any form of ophthalmological intervention, such as surgery. Knowledge of a patient's refractive status was necessary in order to calculate the appropriate corrective lens for perimetry. The HRT and GDx do not actually rely on the input of this information: the former infers the patient's refraction from the scan focus setting established on image capture and GDx measurements are independent of ocular magnification, and so do not require refractive data.

HRT

Scanning laser tomography of the optic nerve head was performed with the HRT by experienced technicians. Five images were taken of each with a 10 × 10° scan. At the operator's discretion, a 15 × 15° scan was used for patients with large optic discs.

The decision regarding pupillary dilation was made at baseline and a consistent approach taken throughout the study. No threshold for minimum pupil diameter was defined; pupil dilation was indicated as a result of unsuccessful imaging through the undilated pupil and is also dependent on the clarity of the ocular media. Although intraocular pathology was an exclusion criterion for the longitudinal arms of the study, patients with early cataract were not excluded, owing to the high prevalence of the condition in patients in their seventh and eighth decades.

Imaging was performed in dim illumination with a vertex distance of 15 mm from the tip of the laser head to the patient's cornea, using a distant light source as a fixation target. At the first visit, prior to recording data, images were taken with various scan setting and the latter adjusted until the message 'Settings of acquisition parameters OK' was obtained. It transpired that a proportion of eyes exists on which this cannot be achieved; in this instance, the operator used their judgement to produce series of images with the optimum appearance. In cases where feedback from successive scans gave conflicting advice regarding the scan depth, it was always set to the deeper level. When contradictory feedback was given regarding the scan focus, the user ascertained the interval between the recommended focal powers and used the median value for image capture.

Image capture was repeated until five high-quality images had been captured and saved. Reasons for rejection included poor centration, suboptimal focus, illumination that was too high, too low or uneven or significant eye movement. In the event of not being able to achieve a sufficient number of adequate images, the reason (or reasons) for this was documented – the incidence of failed examination, together with causative factors, are discussed in Chapter 5.

Scan focus and depth were adjusted at each visit (analysis is not affected by alterations of these settings), but scan width remained constant for all examinations.

Analysis of imaging data was performed towards the end of the study, and was performed independently from data collection. While analysis was performed, the technician did not have sight of patients' research notes, clinical records or diagnostic categorisation.

During analysis, the operator reviewed the five original images and selected the best three for composition of the mean image. Favourable features were considered to be minimal eye movement during image capture, even illumination and optimum centration. On the mean image, the inner aspect of Elschnig's ring was identified and a smooth peri-papillary contour line constructed. HRT data were exported to a data file and subsequently imported into Microsoft Excel. For patients in the longitudinal studies, the same disc outline was used for all examination sessions.

For each mean image, the HRT software computes the mean standard deviation (MSD) of the height of all the pixels. Guidelines are given as to acceptable image quality,[149] with images which have MSD >50 μm considered unacceptable. Therefore, this study excluded mean images that fell outside this threshold.

GDx

SLP of the peri-papillary RNFL was performed with the GDx, operated by experienced technicians. Five high-quality images were recorded on each eye with the default 15° × 15° scan width. Imaging was performed in moderate room illumination, which enhanced the appearance of the corneal focussing rings.

Pupil dilation was indicated as a result of unsuccessful imaging through the undilated pupil, with a consistent approach being maintained throughout the study; dilation was seldom required, except for patients using pilocarpine. During image capture, focus and gain settings were adjusted to obtain optimum image quality; alterations were made during and between examinations. Repeated images were captured until the operator was satisfied that five high-quality images had been recorded. Images that showed poor centration, illumination that was to high, too low or uneven or eye movements during image capture were rejected.

Analysis of imaging data was performed towards the end of the study. Whilst analysis was performed, the technician did not have sight of patients' research notes, clinical records or diagnostic categorisation.

A mean image was composed from the three best single images. Selection was based on the numerical 'image quality' results although, occasionally, the operator over-rode these values when the quality indicator appeared to be in error [see the section 'Results' (p. 99)]. The inner aspect of Elschnig's ring was defined by the placement and resizing of the preformed ellipse on the mean image. When outlining the discs of eyes with peri-papillary atrophy, operators over-rode this rule in order to prevent the measurement ellipse overlying the atrophic area; this strategy is in accordance with established guidelines.[150] GDx data were exported to a data file and subsequently imported into Microsoft Excel. For patients in the longitudinal studies, the same disc outline was used for all examination sessions.

Humphrey field analyzer

All patients underwent examination with the 24-2 full threshold programme on an HFA Model 740. Patients were tested using the appropriate

refractive correction (taking into account their distance prescription and an age-appropriate near correction for the testing distance of the HFA). The refractive correction was inserted into the Perimetric Trial Lens Set.[151] During each perimetric test, the patient was given rest breaks as frequently as required, at the discretion of the operator.

Patients requiring referral

Any patient who showed a clinically significant deterioration in the status of their visual field, visual acuity, clinical appearance of the ONH or the development of a non-related pathology was referred back to their consultant within MREH, as were subjects who were non-compliant with their ophthalmic medication or who were troubled by adverse side-effects.

Analysis of visual field data

The HFA software does not incorporate adequate facilities for the batch processing of visual field data. Data were, therefore, exported from the HFA to a Windows-based visual field database, Peridata (Peridata Software, Hurth, Germany), through a serial port connection. An early, DOS-based version of the Peridata software package was used to export visual field data as text files that could be read in to Excel and/or custom software developed for this project (later versions of the Peridata software do not incorporate this facility). Custom software was written in the Delphi programming language (Borland Software). The custom software performed the following operations:

- Flipped around the vertical meridian all left eye data so that left and right eyes could be pooled.
- From the normal data set calculated the age corrected threshold values for each test location.
- From the normal dataset calculated the pattern deviation values for each test location. Pattern deviation values are designed to reduce the effects of overall shifts in sensitivity and to be more sensitive to localised patterns of loss. The seventh most sensitive test location from each visual field record is compared with that from a normal group to derive an offset that is then applied to all the test points. Certain assumptions are made in the calculation of pattern deviations, the major one being that the seventh most sensitive location is not affected by any visual field loss. This assumption is valid for early visual field defects, that is, those which do not affect a large proportion of the test locations.
- The MPD was calculated (excluding two test locations which frequently fall within the blind spot area) and a distribution of values generated (*Figure 8*). From this distribution, the 95% confidence limit was derived.
- The software calculated the MPD for each of the abnormal group and established whether these were outside the 95% limits for the normal group.

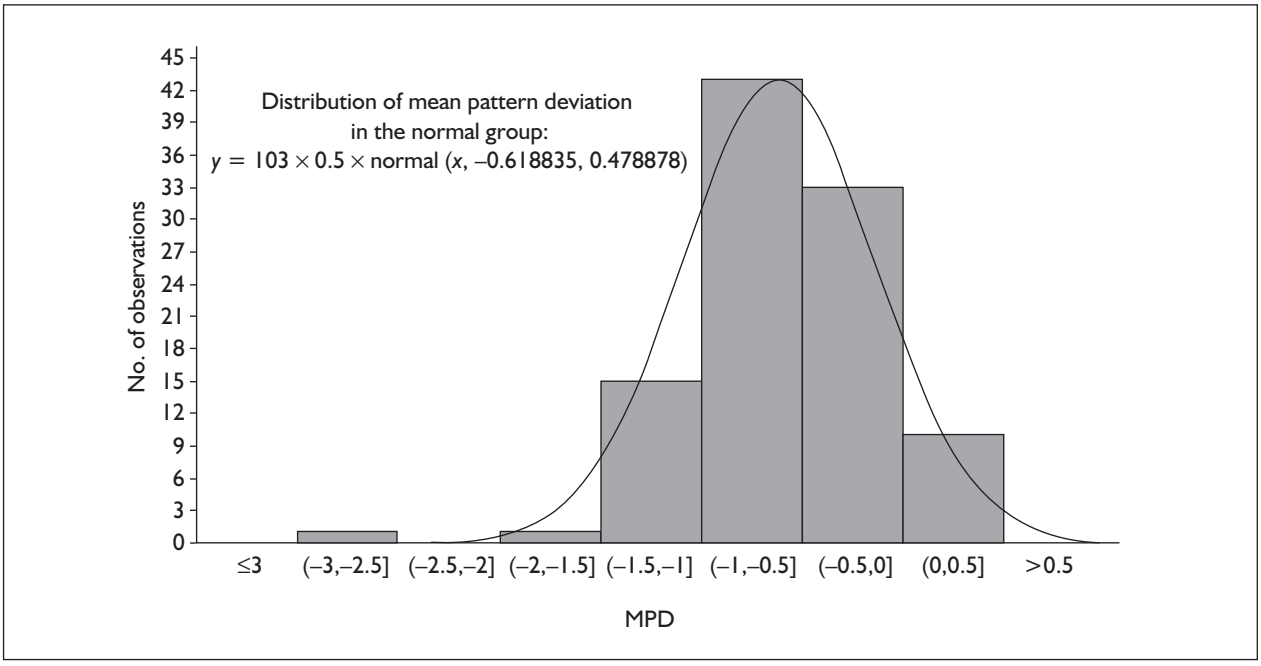


FIGURE 8 *MPD of normal control individuals*

The above calculations are similar to those used within the HFA software and are widely understood by those involved in glaucoma research/management. The use of our own normative database, rather than that incorporated within the HFA, will result in some differences.

For the calculation of sector differences, the custom software:

- Calculated the MPD for each of the five sectors in the superior and inferior visual fields (sectors defined by the GHT see *Figure 9*).
- Derived the distribution of vertically paired sector differences in MPDs for the normal group (*Figure 10*).
- From these distributions derived the 95 and 99% limits.
- Calculated the sector differences for all eyes in the abnormal group and established whether or not they fell outside the 95 and 99% cut-offs.

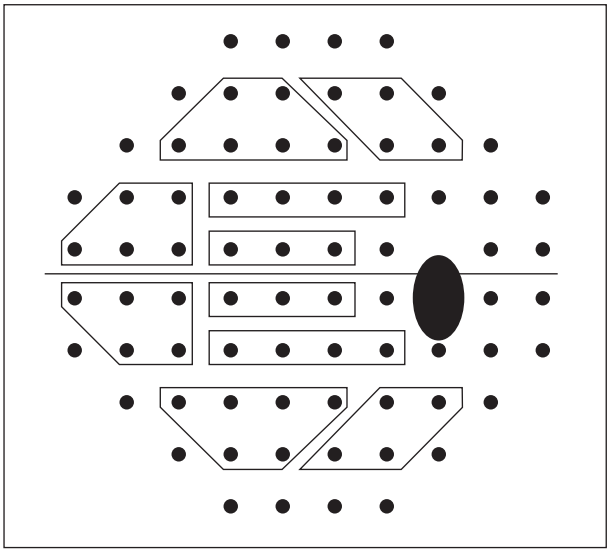


FIGURE 9 *GHT sectors*

The above is based on the HFA GHT, which uses the same sector definitions. Our calculations differ from those used within the perimeter in that the HFA software computes a score for each sector based on pattern deviation probability values rather than MPDs. When we tried to duplicate the HFA computations we found that the distribution of differences between vertically paired sectors was very leptokurtic, with large numbers of normal eyes having zero scores. The calculation of suitable cut-offs from these distributions was compromised by the lack of any suitable defining distributions and the small spread of results. In comparison, the difference in MPDs for vertical sector pairs is well represented by the normal distribution, as shown in *Figure 10*.

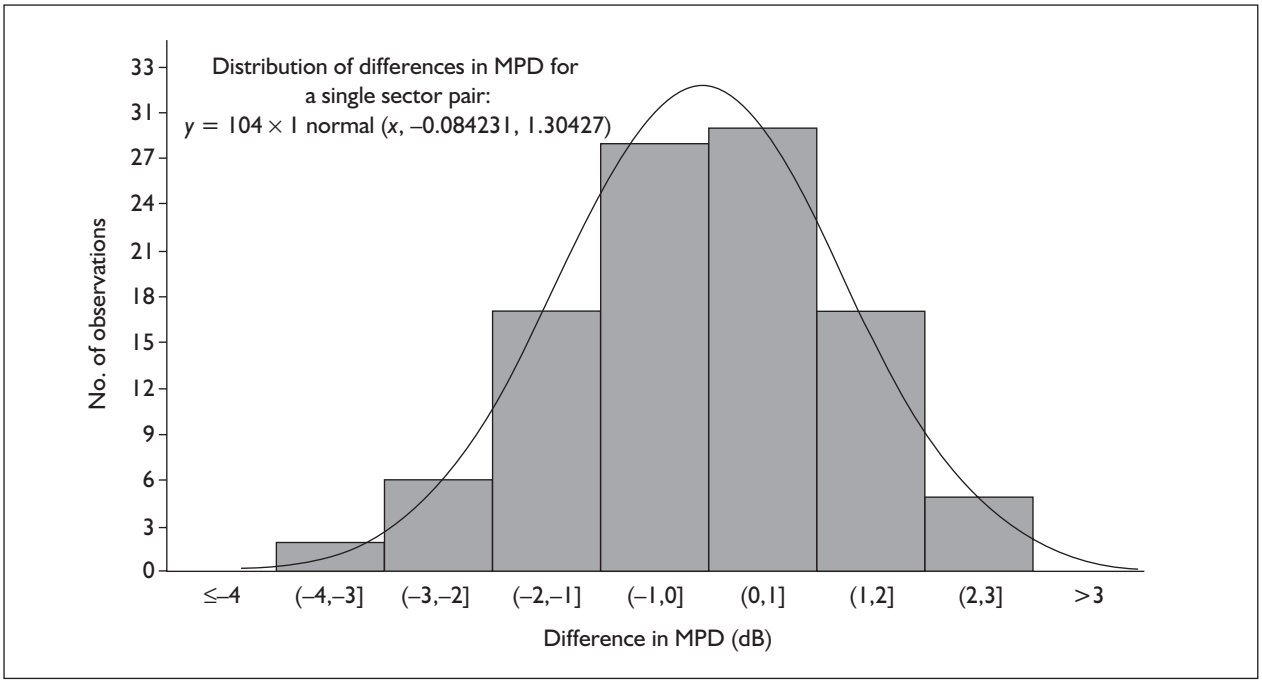


FIGURE 10 *Distribution of differences in MPD for a visual field sector pair*

Problems encountered during the study

The Glaucoma Imaging Study overcame a variety of problems: insufficiencies in patient recruitment, staffing and data loss. Details are given below.

Patient recruitment

The target for the cross-sectional group was met and exceeded early in the study. However, this achievement was not the case for the longitudinal high risk group. The goal of 490 patients in this category was not achieved, because a sufficient number of suitable patients could not be identified and successfully recruited. The initial estimate was that from 1100 possibly suitable patients per annum within MREH, 50% would fail to meet entry requirements and a further 20% would be unwilling to participate. However, in reality, this was not the case: our experience was that >60% did not meet recruitment criteria and of that group, only 32% were willing to take part [see the section 'Hospital patients' (p. 16)]. A further proportion of the remainder was not eligible for recruitment.

At the end of the study, we were able to enter into our analysis data from 240 longitudinal high risk patients; this represents the total amount of data that remained following data loss, discontinuation of patients and so on. Further details are given in the section 'Data loss' (p. 24).

The number of patients in the longitudinal glaucoma group was 75, which exceeded the original target of 50. The normal control group comprised 98 individuals, which fell short of the target of 150. In all, 122 potential control individuals were examined but, owing to unsuitability, 24 were excluded.

Figure 11 shows the overall number of patients initially recruited into the study, the final size of the experimental groups, together with the targets for each experimental group. The disparities between the numbers of patients examined and entered into the analysis are detailed in the section 'Characteristics of experimental groups' (p. 24).

Staffing

Three main issues affected the staffing situation during the study: changeover of research technicians, changeover of administration staff and maternity leave of the research manager.

All technicians who performed data collection were fully qualified optometrists or orthoptists and each had a career background that combined both clinical and academic experience. All had completed an MSc in Investigative Ophthalmology and Vision Sciences (UMIST and University of Manchester), which covered practical and theoretical aspects of all instruments used within the Glaucoma Imaging Study. An adverse consequence of using highly qualified individuals as research workers is that they frequently wish to move on to further their careers elsewhere. Therefore, there was a reasonably rapid turnover of technicians. *Table 4* shows the entire team who

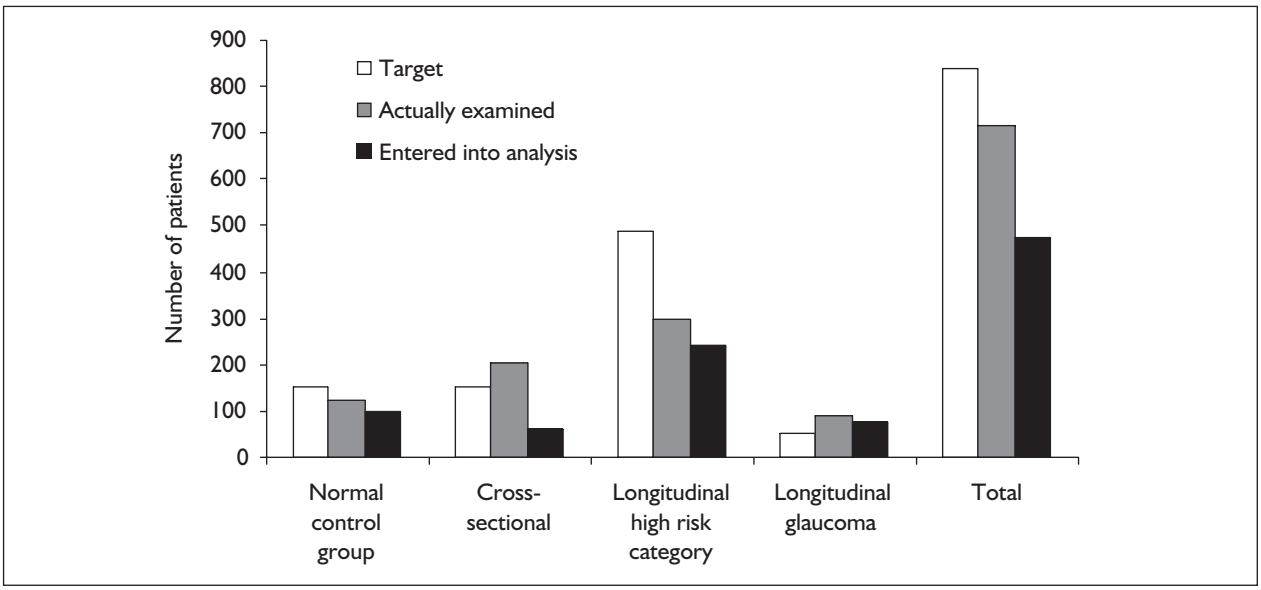


FIGURE 11 *Recruitment target and number of subjects recruited. The data in this graph differ very slightly from those submitted in the final progress report.*

Table 4 *Research staff, April 1998 to March 2004*

Name	Post
Prof. David Henson	Lead applicant
Dr Anna Kwartz	Research manager
Annemiek Coops	Full-time research student
Harpreet Josan	Part-time research technician
Anne Bjerre	Part-time research technician
Amanda Jones	Part-time research technician
Kathryn Dandy	Part-time research technician
Dr Inma Pérez-Gomez	Part-time research technician
Emma Gowan	Part-time research technician
Derek McPhee	Part-time research technician

have contributed to the study over the last 6 years; clearly, all these individuals were not employed simultaneously.

Each technician required a period of training, during which they were not able to collect data. This period of instruction was time consuming for the existing staff and decreased the rate at which patients could be examined.

Once the study was well under way, it became apparent that the research manager's workload of examining patients and dealing with the administration was impractical [for example, in the period up to April 2000, the weekly schedule was: examination of approximately 40 patients (at approximately 75 minutes per examination), 3–4 hours for patient recruitment, 1–2 hours filing and 6–8 hours administration (mailing recruitment letters, dealing with appointments, etc)]. Therefore, through additional funding and reallocation of funds, a part-time administrator was recruited.

The original administrative assistant, Jeanette Allen, left after around 10 months owing to problems with transport and availability of parking at the hospital site. Her successor, Joanne Bradley held the post for approximately 3 years before leaving for a job as a research coordinator elsewhere in the University. During summer 2003, the administration was performed by a variety of temporary staff in combination with the research workers. The vacancy was filled, in August 2003, by Katherine Dickinson, who has proved to be a tremendous asset to the department.

The research manager took two episodes of maternity leave (April 2000–January 2001 and October 2002–July 2003) during the study. Although these periods were covered by the recruitment of additional research staff, the study was adversely affected in that processing of data and analysis was suspended during these times. Subsequent to her return in April 2000, the research manager has worked part-time, on the basis of 2 days per week.

Technological developments

In a rapidly advancing area of clinical and scientific knowledge, such as ophthalmic imaging, there are bound to be technological advances within the period of a longitudinal study. Advances made in the hardware and software for the HRT and GDx have been detailed in the sections 'HRT' (p. 5) and 'GDx' (p. 9), respectively.

Overall, developments incorporated in the HRT have assisted the study: the new system for identifying the margin of the ONH and the new Windows-based software have greatly facilitated our analysis. Although a successor to the HRT was developed while the study was under way, our equipment is still capable of capturing images and producing data that are relevant and comparable to those from the newer device.

Unfortunately, this situation is not the case with the GDx: while the study was under way, our instrument was superseded by the GDx VCC. The latter is capable of adjusting RNFL measurements for individual variations in anterior segment birefringence, which significantly improves the diagnostic precision of the technique.[124–128] Our GDx is not amenable to such modification because the new instrument incorporates both hardware and software developments. The fact that an instrument cannot be upgraded is a major disadvantage to a longitudinal study, and seriously compromises the feasibility of such research. The development of the new instrument is especially disappointing, since at the outset of the study, the GDx was not a prototype instrument but a recently produced third-generation device with wide commercial availability. Lack of backward compatibility and frequent upgrading of equipment is a major drawback in chronic conditions such as glaucoma, where longitudinal data over many years are required for good-quality management.

Technological developments during the study have caused some inconsistencies within the report, depending on when different aspects of the analysis were performed. An example of such a discrepancy is shown in Chapter 11, which reports the effects of inter- and intraobserver variability in the placement of the peri-papillary contour line on HRT and GDx parameters – data

for this part of the study were analysed fairly early in the study, when more contemporary functions, such as the MRA [see the 'Diagnostic algorithms' (p. 8)] were not yet available via the HRT software. The MRA is reported in Chapter 7, but is not incorporated in other chapters, where the analysis was performed several years previously.

Data loss

The glaucoma imaging study was set back by significant data loss, as detailed below. The losses affected data from all the patient groups; the data loss was in no way systematic. When data were compared between instruments, longitudinal series were truncated to equalise the length of follow-up between the different techniques.

HRT

In the first 6 months of the study, data were lost from an estimated 150 patients owing to instability of the database. Subsequently, data were lost on all patients examined between May and July 2000. In November 2001, processed data for approximately 350 longitudinal patients were lost owing to problems encountered while improvements were being made to the HRT's computer.

The initial back-up system for the HRT was a series of 1-GB Jaz disks (Iomega, San Diego, CA, USA). Unfortunately, it transpired that around 20% of these disks were faulty and their data could not be read; consequently, archived data could not be retrieved. In some cases, data were rescued, as our strategy was to make two copies as back-up. The consequence of these problems is that many patients have incomplete data sets, with gaps in the 6-monthly examination schedule. These deficits are detailed in the section 'Longitudinal high risk and longitudinal glaucoma groups' (p. 25).

While our analysis of longitudinal data was under way, it became apparent that a small proportion of patients had series of examinations where the baseline images appeared to be from a different eye (a different patient, rather than just reversal of right and left eyes). Meticulous inspection of patient files and the HRT database revealed that there was a 3-day period in October 1998 when the incorrect association between HRT images and database entry had been made. This probably occurred when the HRT database was rebuilt when the database and all the images were transferred to a new computer in November 2002. All data from this time period (both longitudinal and cross-sectional patients) were excluded from the analysis.

GDx

On 30 July 1998, representatives from Laser Diagnostic Technologies Inc (the manufacturers of the GDx) visited the department. They noticed a fundamental malfunction with the GDx: when the instrument was moved laterally to swap between right and left eyes, the corneal compensator unit should rotate. In fact, this had not been occurring and the research staff had been unaware of this malfunction. It must be noted that the GDx does not incorporate a feedback system to warn the user of this type of hardware failure. Therefore, all data collected until that time (number of subjects = 48, number of visits = 55) were invalid and data collection had to begin afresh. Consequently, useful baseline data for the longitudinal subjects had been lost and, for the cross-sectional patients, some data collection sessions were irrevocably lost. In all cases, attempts were made to contact patients to arrange re-examination. However, many individuals were not willing to attend for a supplementary session, as they were only prepared to make visits to the study that coincided with their regular clinic appointments.

HFA

Data loss with the HFA has been minimal. The only problem encountered with the instrument is that very occasionally (e.g. once a year) it requires the database to be rebuilt. If this function occurs while a patient is being examined, there is a time penalty to the process, but no data loss occurs. While we have been reviewing our data, there appear to be a very small number (<6) of patients for whom no database entry appears for a certain examination, even though the printout exists within our notes. The possible explanations for this finding are either (the unlikely event) of the technician forgetting to save the result or of instrument malfunction.

Characteristics of experimental groups

Demographic characteristics (e.g. age, gender) of each experimental group will be given in the relevant chapter. This section describes the number of patients who were suitable for inclusion, reasons for exclusion, the amount of data that could not be incorporated in the analysis and the overall period of follow-up of the longitudinal patients. Clearly, the number of subjects initially examined in each category does not represent the number of

individuals who were finally included in the experimental groups. Some patients were excluded (from the longitudinal groups) on account of the development of ocular or systemic disease and others owing to poor image quality or because of insufficient data. The importance of reporting such quantities in scientific papers has become widely recognised and a set of protocols for disclosing this information has been established by the Quality of Reporting of Meta-Analyses (QUORUM) protocols.[152] On account of the specific nature of the data from the Glaucoma Imaging Study, the information is presented in tabular form in the following sections, rather than in flow diagrams as recommended.

It must be noted that patients from whom data were lost in the early stages of the study [see the sections 'HRT' and 'GDx' (p. 24)], where HRT and GDx data were lost for 150 and 48 patients, respectively, are not represented below since, after these episodes, data collection was started anew.

Cross-sectional group

The total number of subjects examined in the cross-sectional group was 206, which exceeded the initial target by 51 subjects. *Tables 5* and *6* detail the outcome of recruitment into the cross-sectional group.

Normal control group

Although the final number of normal control individuals entered into our analysis was 98, a total of 122 was initially examined. *Table 7* gives reasons for exclusion of control individuals.

Longitudinal high risk and longitudinal glaucoma groups

Characteristics of the number of visits and length of follow-up of the longitudinal groups are described in *Table 8* and *Figure 12*, which represent the actual data that was included in the analysis, rather than the total amount of data collected. *Table 9* summarises reasons for discontinuation of patients from both longitudinal arms. In some

TABLE 5 *Recruitment of cross-sectional patients*

Outcome	No. of subjects (original *n* = 206)
Examined but not recruited[a]	30
Could not be examined with one or more instrument(s)[b]	40
HRT data lost[c]	37
HRT data lost but has database entry[d]	3
Poor-quality raw HRT data	8
Poor-quality raw GDx data	6
Poor-quality raw HRT and GDx data	14
Possible misallocation of image and database entry on HRT[e]	2
Unable to compute HRT mean image with MSD ≤ 50 µm	5
Total remaining	61

[a] Reasons for patient examination not resulting in recruitment are given in *Table 6*.
[b] Reasons for unsuccessful examination are described in detail in Chapter 6.
[c] No HRT data exists owing to data loss or absence of back-up owing to faulty Jaz disk.
[d] Patients have HRT database entry but no images exist – possibly owing to errors while the database was rebuilt when the HRT computer was upgraded.
[e] Data excluded because of possible misallocation of database entry and image [see the section 'HRT' (p. 24)].

TABLE 6 *Reasons for non-recruitment of cross-sectional patients*

Reason	No. of subjects
Non-Caucasian race	14
Unwilling to sign consent form[a]	3
Recruited in error, e.g. too young, secondary glaucoma or IOP too low	9
Could not tolerate examination	4
Total	30

[a] Patients were concerned about the possible risks of their eye being examined with a laser or were deterred by the statement 'potential risks' in the consent form.

TABLE 7 *Reasons for exclusion of normal control individuals*

Reason	No. of subjects (original *n* = 122)
Refraction >5.00DS and/or >3.00DC	11
Diagnosed with ocular pathology	5
Repeatedly failed field test[a]	4
Unable to achieve good-quality images with HRT and/or GDx	3
Unable to compute HRT mean image with MSD <50 μm	1
Total remaining	98

[a] Patients in this category did not show a field defect typical of a specific pathology, but on account of extremely poor fixation and attention during the test, produced spurious results.

TABLE 8 *Length of follow-up of longitudinal patients*

Characteristic	Longitudinal high risk	Longitudinal glaucoma
Number of subjects	240	75
Median length of follow-up (years)	3.41	3.51
Standard deviation, length of follow-up (years)	1.01	0.98
Minimum length of follow-up (years)	1.04	1.06
Maximum length of follow-up (years)	5.14	5.12
Median number of visits	7	7
Standard deviation, number of visits	1.91	1.99
Minimum number of visits	4	4
Maximum number of visits	13	11

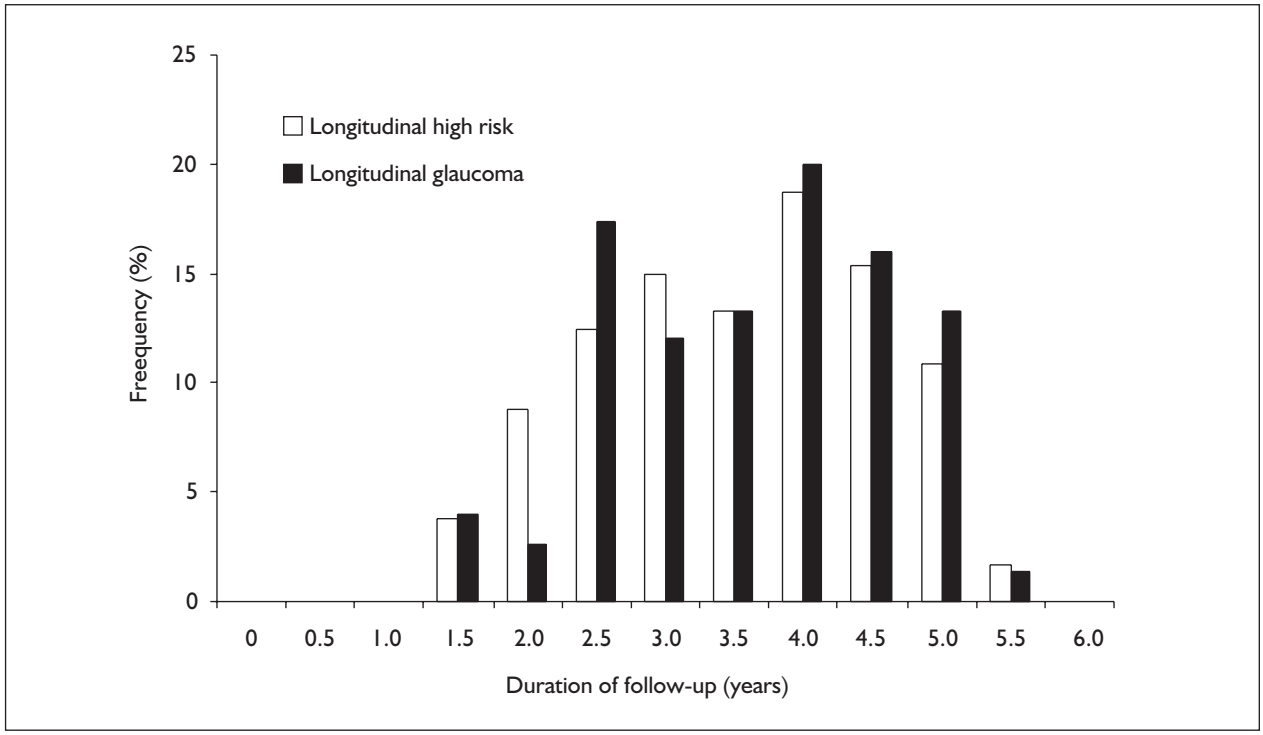


FIGURE 12 *Duration of follow-up*

TABLE 9 *Reasons for discontinuation of longitudinal patients*

Reason for discontinuation	Longitudinal high risk	Longitudinal glaucoma
Not enough time to commit to study	1	2
Personal reasons	6	1
None given	7	1
Repeated non-attendance[a]	11	6
Deceased	4	1
Became too ill/frail to continue[b]	3	0
Moved away	8	4
Developed systemic/ocular pathology[c]	8	2
Other	4	3

[a] Multiple attempts (by telephone and letter) were made when patients failed to attend appointments.
[b] Patients were still eligible for inclusion but felt they could not participate because of ill-health or frailty.
[c] This category included patients who had to be excluded from the study because they developed a systemic (e.g. Cushing's disease, cerebrovascular accident) or ocular (e.g. iritis, retinal vein occlusion) disease that precluded continued involvement in the study.

TABLE 10 *Number of HRT data sessions not included in analysis*[a]

Reason	Longitudinal high risk	Longitudinal glaucoma
Raw image inadequate quality	105	53
Mean standard deviation >50 μm	13	5
Contour line did not align on follow-up images	4	3
Wrong patient data recorded	12	10
Data lost	68	2

[a] The data represent the number of examination sessions that were not incorporated in the analysis. Some individuals are represented in more than one category.

TABLE 11 *Number of GDx data sessions not included in analysis*[a]

Reason	Longitudinal high risk	Longitudinal glaucoma
Raw image inadequate quality	59	58
Software could not align single images	34	8
Could not compose adequate mean image	28	11
Data lost	1	1
Contour line did not align on follow-up images	4	0
Insufficient raw data to compute mean image	5	0
Patient data stored in two separate folders	2	0

[a] The data represent the number of examination sessions that were not incorporated in the analysis. Some individuals are represented in more than one category.

cases, sufficient data had already been collected to allow the patient's data to be utilised in the analysis. In others, where ≤ 3 visits had been made, patients were excluded from the longitudinal analysis and their data were subsequently used for calculation of sensitivity and specificity (see Chapter 7).

Table 10 shows the number of HRT data sessions that were not included in the longitudinal analysis; comparable data for the GDx are shown in *Table 11*. It must be noted that the data shown in these two tables represent the number of examination sessions that could not be incorporated into the analysis, rather than the number of patients. Additionally, data from some patients are represented more than once in situations where multiple examination sessions from one individual were excluded from the analysis.

Subjects who had ≤ 3 visits were not incorporated in the longitudinal analysis. This strategy was

TABLE 12 *Outcome of patients recruited for longitudinal examination*

Outcome	Longitudinal high risk (*n* = 297)	Longitudinal glaucoma (*n* = 92)
Included (≥ 4 data points remaining)		
Complete set of data	98	32
≥ 1 session lost with sufficient data	129	34
Early discontinuation with sufficient data	13	9
Total included in analysis	**240**	**75**
Excluded (≤ 3 data points remaining)		
Early discontinuation	39	11
Data loss	18	6

adopted because the study uses linear regression to detect change and it is inappropriate to perform such analysis when patients have <4 visits. In cases where data was deficient from one instrument, the data series with the other two techniques were truncated to ensure that the start and end dates were identical. The outcome of these modifications on the size of the experimental groups is shown in *Table 12*.

Back-up procedures

All three instruments are rigorously backed up. The following three subsections detail the procedures for each piece of equipment. In order to provide the capacity to store data off-site, the facility has been arranged for us to utilise a University server. In all cases where removable media (e.g. CD-ROM or floppy disks) are used for back-up, they are stored remotely from the equipment.

HRT

The HRT's computer has been fitted with a RAID level 4 system. Basically, this system involves two hard disks, which, when the computer is switched on, automatically inspect each other's contents and copy, between each other, any missing elements. This system ensures data integrity in the event of a hard disk drive failure.

In addition, each week, all new images, together with the latest version of the database, are backed up on to CD-ROMs. Once a month, a further back-up procedure is instigated; another copy of all the images, together with the database, is made on to a University server. This provides a supplementary echelon of security. The server's data are further backed up on to tapes.

GDx

Each week a copy of all new data is made to the GDx computer's second hard disk, with a further copy to a University server.

HFA

Each week the HFA is backed up on to a series of floppy disks. Two sets of disks are used, which are alternated each fortnight. Data have also been serially exported to other computers running Peridata software.

Publications and conference presentations

Data from the Glaucoma Imaging Study have been presented at a number of international conferences, as detailed below:

Conference:	The Association for Research in Vision and Ophthalmology (ARVO) Annual Meeting, Fort Lauderdale, FL, USA, May 1999
Authors:	McLeod D, Joseph AJ, Henson DB, Harper RA, O'Donoghue EP, Spencer AF
Title:	Applicability of scanning laser polarimetry, retinal tomography and threshold perimetry to an unselected population of patients attending a glaucoma clinic
Presentation type:	Poster
Conference:	Glaucoma Society (UK and Eire) Meeting, November 2001
Authors:	Mathew P, Kwartz AJ, Henson DB, Spencer AF
Title:	Comparison of examination times for different methods of detecting glaucoma
Presentation type:	Paper

Conference:	ARVO Annual Meeting, Fort Lauderdale, FL USA, May 2002
Authors:	Kwartz AJ, Jones AK, Henson DB, Harper RA, Spencer AF
Title:	The effect of inter- and intra-observer variability in the definition of the optic disc margin on HRT and GDx parameters
Presentation type:	Poster

Conference:	Optical Group of the Institute of Physics: Imaging in the Eye: Technologies and Clinical Applications, May 2002
Authors:	Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D
Title:	Relationship between regression residuals in longitudinal analysis of HRT parameters and image quality
Presentation type:	Paper

Conference:	International Perimetric Society, June 2002
Authors:	Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D
Title:	Ability of the Heidelberg Retina Tomograph and GDx to detect patients with early glaucoma
Presentation type:	Paper

Conference:	International Perimetric Society, June 2002
Authors:	Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D
Title:	The ability of the Heidelberg Retina Tomograph and GDx to predict the development of visual field loss
Presentation type:	Poster

Conference:	Glaucoma Society (UK and Eire) Meeting, November 2003
Authors:	Kwartz AJ, Henson DB, Spencer AF, Harper RA, McLeod D
Title:	Linear regression of global and sectoral HRT parameters from glaucoma converters
Presentation type:	Paper

Conference:	Glaucoma Society (UK and Eire) Meeting, November 2003
Authors:	Coops A, Henson DB, Kwartz AJ, Spencer AF, Harper RA
Title:	The role of imaging in the detection of glaucoma
Presentation type:	Paper

Conference:	ARVO Annual Meeting, Fort Lauderdale, FL, USA, May 2004
Authors:	Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D
Title:	Ability of HRT summary measures to identify perimetric glaucomatous progression
Present type:	Poster

Conference:	ARVO Annual Meeting, Fort Lauderdale, FL, USA, May 2004
Authors:	Henson DB, Coops A, Kwartz AJ, Harper RA, Spencer AF, McLeod D
Title:	Rate of functional and structural change in glaucoma suspects
Presentation type:	Poster

Conference:	European Association for Vision and Eye Research, Portugal, September 2004
Authors:	Coops A, Henson DB, Kwartz AJ, Fenerty C, Harper RA, Spencer AF, McLeod D
Title:	Differences between glaucomatous patients with predominantly functional or structural deterioration
Presentation type:	Submitted for publication

A number of publications have resulted from the Glaucoma Imaging Study.[153–159]

Chapter 4

Comparison of examination times with the Heidelberg Retina Tomograph, GDx and Humphrey Field Analyzer 24-2 program

Introduction

This component of the Glaucoma Imaging Study was designed to measure examination times with the HRT, GDx and HFA in a normal clinical environment. Lengthy examination time is a major disadvantage of perimetry, with obvious cost implications for both patients and the NHS. In addition, the length of the test is very demanding for elderly and frail patients. The results of this study have important implications for clinic scheduling and for the calculation of cost–benefit analysis.

In addition to considering the actual time taken to perform the measurements, the study also considers 'total examination time' by including the length of time to position the patient at the instrument, perform the analysis and print the results. This distinction was made to prevent the potentially favourable appearance of a test modality that allowed measurements to be taken relatively quickly, but required lengthy analysis to achieve the final results.

Method

Patients were examined according to the methods described in the section 'Data collection' (p. 18).

Each examination was divided into three phases: 'Preparation' (of both the patient and instrument), 'Examination' and 'Analysis'. The duration of the 'Total examination' was determined from the sum of these components. Exact definitions are given in the sections 'HFA' and 'HRT' (below) and 'GDx' (p. 32).

Mean spherical refractive error (calculated from: spherical power + 0.5 × cylindrical power), pupil diameter and visual field PSD were recorded for each eye. The last value was taken from the HFA printout, rather than calculated via the method described in the section 'Analysis of visual field data' (p. 20).

Patients

The sample consisted of consecutive patients attending the Glaucoma Imaging Study between 4 June 2001 and 4 July 2001. During this time, a combination of new and follow-up patients was assessed.

HFA

'Preparation' was defined as the time taken from seating the patient comfortably at the instrument until commencing either the demonstration or test programme. This component included either entering patients' details into the database (first visit) or recalling information (follow-up patients), explaining the examination procedure and setting up and positioning the perimetric trial lens set.

The length of the 'Examination' included the actual testing time, together with any breaks instigated by either the patient or examiner. This interval also incorporated time spent changing over between the two eyes. For those unfamiliar with perimetry, a 1-minute demonstration programme was run prior to testing the first eye at the patient's first visit. At the examiner's discretion, a full or shortened demonstration was sometimes also included for follow-up patients.

The 'Analysis' phase was measured from the end of the examination to obtaining printouts for the two eyes on an external laser printer (LaserJet 4, Hewlett Packard).

HRT

'Preparation' included the time taken to position the patient comfortably at the instrument, enter or recall information from the database and lining up the instrument ready to capture the first image.

The length of time taken to capture and save five images for each eye, together with any rest periods, was defined as the duration of the 'Examination'. This element included time spent capturing images that were discarded owing to unacceptable quality.

The 'Analysis' procedure involved deriving the topography images, the operator selecting the three best from the five original series from each eye, composing mean topography images, drawing a contour line around each disc margin and printing the results. For subjects who were followed up from previous sessions, this component included accessing the baseline images in order to export the disc contour lines into the current examination session. The termination of this phase was marked by the completion of the printout for each eye (Deskjet 1600C, Hewlett Packard).

GDx

'Preparation' and 'Examination' times were defined as for the HRT.

The 'Analysis' element involved determining the three highest quality images and combining them into a mean image. The disc margins were defined or the outlines imported from baseline examinations, as appropriate, for new or existing patients. Time spent printing the results was included (Deskjet 1600C, Hewlett Packard).

Results

A total of 40 patients (aged 60.4 ± 9.3 years) was examined during the period of this study. Of these examinations, 33 were follow-up visits and seven were first visits. Of patients attending follow-up appointments, the mode of the number of previous visits was six (mean 3.81 ± 2.02).

Examination times

All results were converted to decimal form. The means (and standard deviations) of examination durations are shown in *Table 13* and *Figure 13*. The Shapiro Wilks *W*-test shows most of the components were not normally distributed.

TABLE 13 *Duration of examination components*[a]

	Duration (minutes)			
	Preparation	**Examination**	**Analysis**	**Total**
HFA	2.83 (1.08)	23.62 (3.94)[b]	1.89 (0.38)	28.34 (4.60)[b]
HRT	1.39 (0.56)	4.92 (2.15)[b]	5.98 (2.27)[b]	12.29 (3.39)[b]
GDx	1.61 (0.89)[b]	6.45 (3.44)[b]	3.73 (1.56)[b]	11.79 (4.33)[b]

[a] Data shown as mean (standard deviation).
[b] $p < 0.01$ (Shapiro Wilks *W*-test) suggesting data are not normally distributed.

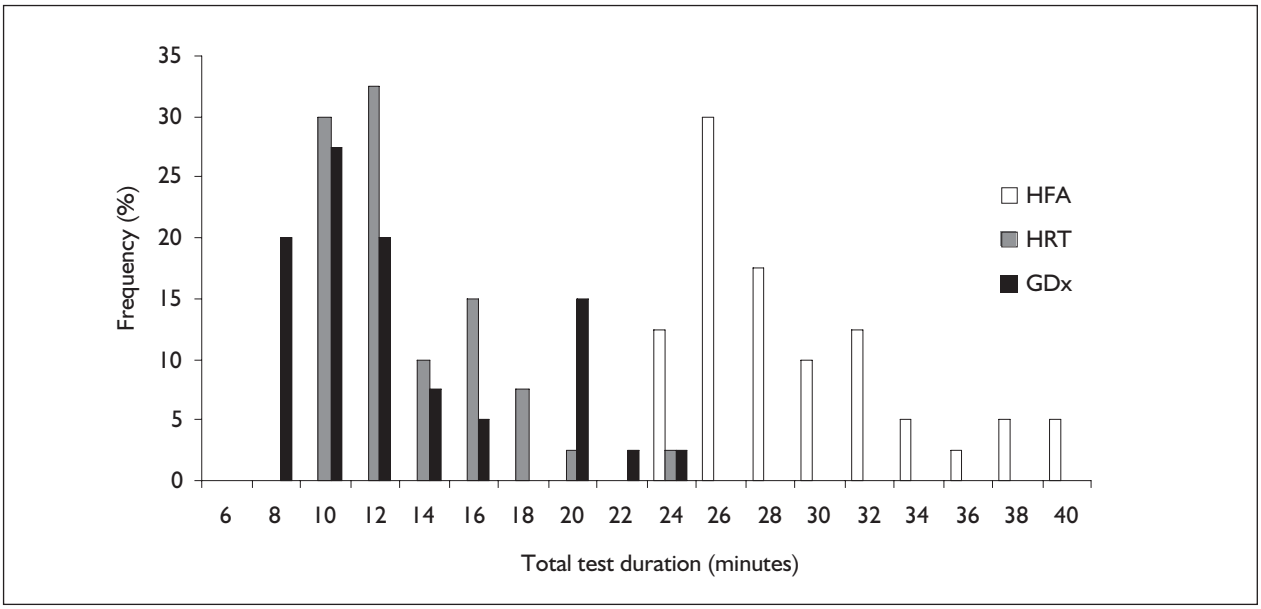


FIGURE 13 *Frequency distribution of examination durations*

TABLE 14 *Duration of preparation plus examination phases*[a]

Instrument	Preparation plus examination Duration (minutes)
HFA	26.44 (4.47)
HRT	6.31 (2.42)
GDx	8.06 (4.00)

[a] Data shown as mean (standard deviation).

In order to consider the length of time the patient is required to spend with the examiner, 'Preparation' and 'Examination' times were combined, and the results are shown in *Table 14*. This value is of interest since analysis and printing could be conducted at a later time, although there would be a time cost to restore data from file.

When considering the combination of 'Preparation' and 'Examination' times, the HRT was quickest ($p = 0.021$; Mann–Whitney *U*-test), closely followed by the GDx and then the HFA ($p \ll 0.01$; Mann–Whitney *U*-test). However, on comparison of 'Total examination' time, there was minimal difference between the HRT and GDx, on account of the former's longer analysis time.

New versus follow-up patients

Seven of the patients had not been previously examined within the study and, therefore, required two supplementary measurements: focimetry and keratometry. The times taken to perform these were 0.77 ± 0.65 and 1.13 ± 0.44 minutes, respectively.

Table 15 compares the durations of the different examination phases between new and follow-up patients; these figures do not include the time needed to perform keratometry and focimetry on new patients. 'Preparation' time was longer for new than follow-up patients with the HFA ($p = 0.023$; Mann–Whitney *U*-test) and HRT ($p = 0.011$). The 'Examination' time was longer for novice patients for perimetry ($p < 0.001$). For each of the other examination components, there was no statistically significant difference between new and follow-up patients. For the two imaging techniques, there was no significant difference in total examination duration between new and follow-up patients, whereas new patients required significantly longer for complete HFA examination than those attending follow-up appointments ($p = 0.004$).

With the HFA, the duration of testing time (recorded from HFA printouts) for both eyes combined was longer for new (25.51 ± 3.85 minutes) than follow-up patients (20.75 ± 3.26 minutes) ($p = 0.009$; Mann–Whitney *U*-test). These data reflect the length of time the patient actually spent performing the examination and exclude time spent performing the demonstration programme, resting during the examination and swapping between the two eyes.

Refractive error

The mean spherical equivalent refractive error of the experimental group was –0.40 ± 2.57DS right eye and –0.39 ± 2.63DS left eye. No significant correlation was found between refractive error and examination duration for any of the three instruments (Kendall's rank correlation coefficient).

TABLE 15 *Duration of examination components – comparison between new and follow-up patients*[a]

	Duration (minutes)								
	HFA			HRT			GDx		
Component	New	Follow-up	p^b	New	Follow-up	p^b	New	Follow-up	p^b
Preparation	3.37 (0.55)	2.70 (1.12)	0.023	1.83 (0.46)	1.28 (0.54)	0.011	1.72 (0.69)	1.58 (0.95)	0.651
Examination	27.37 (3.81)	22.68 (3.18)	<0.001	4.72 (2.13)	4.97 (2.19)	0.773	6.29 (3.64)	6.49 (3.45)	0.910
Analysis	1.89 (0.56)	1.89 (0.37)	0.986	5.28 (0.82)	6.16 (2.49)	0.182	3.25 (0.69)	3.85 (1.96)	0.168
Total	32.63 (3.84)	27.26 (3.95)	0.004	11.82 (2.06)	12.41 (3.66)	0.560	11.26 (3.95)	11.92 (4.47)	0.692

[a] Data shown as mean (standard deviation).
[b] Mann–Whitney *U*-test.

During the study, only one patient required HRT imaging to be performed through their spectacles on account of the cylindrical component of their refractive error exceeding 1.00DC.

Pupil diameter

No relationship was found between pupil diameter and examination time with the perimeter. Kendall's rank correlation coefficient showed a negative correlation between pupil diameter and HRT examination time for the right eye ($\tau = -0.331$, $p = 0.005$) and left eye ($\tau = -0.409$, $p \ll 0.001$). A similar relationship was found for GDx examination duration for the right ($\tau = -0.409$, $p = 0.005$) and left ($\tau = -0.331$, $p < 0.001$) eyes, respectively.

Pattern standard deviation

A positive correlation was found between PSD and visual field examination length for the right ($\tau = 0.433$, $p < 0.001$) and left ($\tau = 0.442$, $p < 0.001$) eyes. For this calculation, the test duration recorded from the HFA printouts was used. This approach was taken to exclude confounding factors of time spent resting or demonstrating the examination.

Discussion

The results of this study are important because they would provide a vital contribution to any future cost–benefit analysis in looking at improved outcomes resulting from the routine inclusion of the HRT and GDx.

In the literature, little mention is made of the issue of examination duration for the two methods of imaging. One source states that the time needed to perform imaging with the HRT on both eyes of a new patient, including data entry and the time taken for supplementary examinations, is about 20 minutes;[160] however, no details are given regarding the scientific basis of this estimate. The test duration is somewhat longer than the mean value found in this study (approximately 12 minutes). However, the latter figure does not consider the extra measurements required for new patients (keratometry and determination of refraction), the average duration of which was 1.9 minutes.

The number of patients examined in the study was 40. The aim was to perform an audit of examination duration for a period of 1 month. The chosen period was 4 June 2001 to 4 July 2001. A far better strategy would have been to calculate optimum sample size, rather than limit data collection to a specified duration.

1. Preparation
   This phase was much longer for the perimeter than either of the imaging devices, which is probably a reflection of the fact that the HFA software is rather slow in comparison with the HRT and GDx.
   The recorded 'Preparation' for perimetry from this study might be slightly longer than necessary, since we used the HFA's integral touch screen for data input. The provision of an external QWERTY keyboard (optional extra for the HFA) might reduce preparation time.

2. Examination
   Although the actual time to acquire a single image is shorter with the GDx than HRT (0.7 versus 1.6 seconds), the longer 'Examination' time found with the GDx is a reflection of the fact that the equipment takes longer to save the images. It must be noted that, for both the HRT and GDx, our thorough approach of taking five images on each eye has led to this phase being longer than that in other clinics where a smaller number of images are collected. The design of the GDx software additionally contributed to this: it requires the electronic patient 'folder' to be closed after the first eye has been examined, because the maximum number of images that it can contain is six. If only three images had been taken on each eye, they could have all been saved in one 'folder'.

3. Analysis
   The 'Analysis' with the HFA is clearly far quicker than with the imaging devices since it requires no reviewing of images or operator input. Timing data from the HRT and GDx might differ from a normal clinical setting on two counts. First, our strategy of selecting the three best quality images will have increased the time taken for this phase. Second, the process of recalling the baseline images for the purposes of exporting the reference ring might have been quicker than elsewhere, owing to our policy of keeping all data on very large hard disks, rather than archiving data and having to recall them at a later date.

4. Total examination
   The result of greatest potential significance from this study is the determination of the maximum number of patients who could be examined during one clinical session with each

instrument. If an average clinical session is considered to be 3.5 hours, then seven, 17 and 18 patients could be examined with the HFA 24–2 program, HRT and GDx, respectively. These estimates have been derived for a mix of patients (seven new and 33 follow-up), which is probably a realistic scenario in a glaucoma clinic.

The maximum number of perimetric tests within one session is possibly an underestimate of the real value that would be relevant to a contemporary glaucoma clinic. We used the full threshold algorithm, whereas the SITA Standard strategy, which is now widely employed, reduces testing time by an estimated 56% in normal individuals and 45% in glaucoma patients.[161] To calculate the theoretical number of patients who could be examined per 3.5-hour session with SITA Standard, halving the examination phase from this study (from 23.62 to 11.81 minutes), gives a total examination time of 16.5 minutes. This would mean that the number of patients who could be examined per session would increase to 13. (This calculation is slightly imperfect because SITA would marginally lengthen the 'Analysis' phase owing to a longer processing time.)

Another consideration is that in our 'ideal' research environment each perimetric patient is given the full attention of one operator, whereas in other settings, one technician may supervise several patients simultaneously (Poole BN, Fields Technician, Central Manchester and Manchester Children's University Hospitals NHS Trust; personal communication, 2001). Such multi-taking could potentially increase the number of perimetric patients examined per session, but there would be implications for the quality of the examinations, as many patients require continuous input from the perimetrist to ensure that they maintain central fixation.

5. New versus follow-up patients
   For patients who have not been examined previously, additional time is required to measure their spectacle prescription and corneal curvature (the former is required for examination with the HFA and the latter for the HRT). Although this measurement takes a relatively short time, it is nonetheless worthy of consideration, since two of the methods of examination cannot be performed without these data. (It must be remembered that the sample size for these measurements was small, $n = 7$.)

   The time required to perform these measurements is highly dependent on the equipment used. Focimetry was performed with an automated instrument; clearly, a manual device would have taken longer. Conversely, this study used a manual keratometer, whereas a more rapid examination could have been performed with an auto-keratometer. If a new imaging clinic were to be established, it is possible that it would be equipped with the most up-to-date equipment (i.e. auto-focimeter and auto-keratometer). Although this study neglected to measure the time spent on moving the patient to and from the keratometer, this is unlikely to be an issue with a more contemporary instrument, where the device is taken to the patient. Relocating the patient is also not an issue for focimetry, as the patient remains seated while their spectacles are carried to the instrument by the technician.

   Examination duration with the HFA was longer for new patients than those attending follow-up appointments (32.63 ± 3.84 versus 27.26 ± 3.95 minutes). A contributory factor is potentially the inclusion of the full demonstration phase (1 minute) for those unfamiliar with perimetry (five of the seven new patients were normal control individuals who had not previously undergone a full threshold perimetric examination). However, this phase does not fully account for the differences between the two groups, since the testing time (excluding demonstration and resting time both within and between eyes) recorded from the HFA printouts was also significantly longer for new than follow-up patients (25.51 ± 3.85 and 20.75 ± 3.26 minutes, respectively).

   For the imaging techniques, the differences in the total examination time between new and experienced subjects did not achieve statistical significance, although the preparation phase with the HRT was marginally longer for new than follow-up patients (1.83 ± 0.46 versus 1.28 ± 0.54 minutes).

   In an average 3.5-hour clinical session, the HFA, HRT and GDx could each be used to examine six, 17 and 18 new patients or seven, 17 and 17 follow-up patients, respectively. As far as clinic scheduling is concerned, it would be reasonable to book perimetric appointments at 30-minute intervals and imaging appointments at 15-minute intervals, with no discrimination between new and follow-up patients.

6. Refractive error
   It was felt that the different types of fixation targets with the HRT (distant light) and GDx (proximal light-emitting diode) might have been preferred by hypermetropic and myopic patients, respectively. Patients who are

presbyopic hypermetropes often report that the GDx's proximal target is difficult to use. However, it transpired that there was no significant effect on testing time. This might be a reflection of the high proportion (>80%) of patients who have previously undergone the examination and become accustomed to the procedure. Patients who were unable to undergo the examination owing to an inability to view the fixation targets had already been excluded from the imaging study (this issue is discussed further in Chapter 5).

7. Pupil diameter
   Smaller pupil diameters were associated with longer examination times for the HRT and GDx. This is due to the critical alignment required, because even image illumination is more difficult to achieve with small pupils. Mydriasis is, however, not a solution to decreasing examination time, since there is greater scope for misaligning the laser beam from the visual axis in a widely dilated pupil.
   Additionally, for patients with small pupil diameters, any eye movement during image capture has a greater effect on the quality of the image. Consequently, a greater number of images might have to be taken in order to achieve five of a good standard.

8. Pattern standard deviation
   The relationship of increased testing time and PSD has been documented previously.[162,163] This is because, at the outset of the test, the thresholding procedure is lengthier for points of reduced sensitivity and also a reflection of the inefficiency of the 'next points' algorithm when presented with an uneven hill of vision.

9. Other factors
   This study was performed prior to the upgrade of our HRT to the latest Windows-based software version. While the timing study was conducted, the original DOS-based software was used. The use of the earlier version will have affected elements of the examination, such as the length of time taken to identify the margin of the ONH and the speed with which baseline examinations could be identified from the database. Therefore, the results are not necessarily applicable to the current technology.

If these results are to be used in planning clinic schedules, several other issues need to be considered. All three instruments are computer driven and, consequently, the pace with which they function is dependent on their processor speed. Our values are based on new equipment of a high specification. An example is the HRT, which is equipped with the latest (Windows) capture card and a PC that has 256 MB RAM and fast disk access. Prior to upgrading to this specification, the previous system had considerably longer processing times.

Another factor for consideration is the 30-minute warm-up period that the HRT requires prior to use. The HFA also cannot be used immediately, as it takes several minutes to 'boot up'. All three instruments require stored data to be backed up [see the section 'Back-up procedures' (p. 28). The approximate times taken to perform the back-up of the HFA, HRT and GDx are 30, 20 and 45 minutes, respectively. Of course, the duration of the process is dependent on how much data has been collected. In addition, the transfer of data to the off-site server is usually performed overnight. To increase efficiency, the three pieces of equipment can be backed up simultaneously, with the operator switching between devices; this process takes >1 hour, during which the instruments and the technician are unavailable for examining patients.

Data presented in this chapter may not be applicable to the more contemporary version of the HRT and GDx, the HRT II and GDx VCC. Examination duration with both instruments may be significantly different than with their predecessors on account of differences in the software interface, methods of image capture and analysis. The GDx VCC also has a significantly different software interface to the original GDx, and has been designed to take only one image of each eye (the latter will, of course, lead to greatly reduced examination times; however, no independent research has yet been performed to investigate the implications of this approach for precision and repeatability). At the baseline examination, the GDx VCC involves a system for compensating for individual variations in anterior segment birefringence, which involves the capture of an additional macular image at the baseline examination. This additional measurement may have consequences for differences in examination duration between new and follow-up patients. The GDx's successor also contains a modified system for identifying the disc margin.

The effects of developments incorporated in both the HRT II and GDx VCC will need to be considered in calculations of the number of patients to be seen per session with the more contemporary equipment.

Chapter 5

Economic evaluation

A fundamental consideration in the modern NHS is financial expediency. Several aspects of the clinical application of the HRT, GDx and HFA are described in this report. However, it is also important to report the fiscal feasibility of incorporating such equipment into the normal clinical armoury. Chapters 4 and 6 examine the time taken to perform a complete examination with each piece of equipment and the proportion of an unselected population of glaucoma patients that can successfully undergo the tests, respectively. Although a formal economic evaluation is not given, this chapter will examine various economic aspects of the equipment. It will do so by taking a largely differential approach, by examining the differences and similarities in running and operating costs between the HRT, GDx and HFA.

Costs of computer consumables were correct on 1 April 2004 and were taken from a major UK supplier.[164] Prices of hardware and maintenance contracts were also confirmed, on the same date, by the manufacturers' UK representatives, as was the estimate of a technician's salary.

Labour

Each of the three instruments requires a skilled technician to perform the examination. Although this study used optometrists and orthoptists who were qualified to MSc standard (or higher), this calibre of staff is not actually necessary for successful operation of the equipment. There is no large disparity in the skill level required to perform any of the three tests and a trained technician could reasonably conduct all three methods of examination (especially with the newer versions of the imaging devices, the HRT II and GDx VCC, whose operation is far simpler than that of their predecessors). From our experience in a newly developed glaucoma scheme at MREH, a trained technician can easily and successfully perform perimetric examination and optic disc photography (which is roughly equivalent in complexity to digital imaging).

Labour costs for each technique are, therefore, dependent on the number of patients who can be examined within a clinical session. As described in the previous chapter, the HRT, GDx and full threshold 24-2 program of the HFA could each be performed on 17.1, 17.8 and 7.4 patients, respectively, in a 3.5-hour period. These figures reflect the length of time required to position the patient at the instrument, perform the test and complete the analysis (e.g. compose mean images, define the optic disc margin). In calculating labour costs, discrimination could be made between actual examination of the patient and subsequent analysis of data, as separate individuals could perform the two tasks. However, in reality, it is most likely that the same member of staff would complete the entire examination.

The employment cost of an appropriate technician (grade MT02) is £17,082.00. This amount includes basic salary (£14,105.00), employer's national insurance contribution (£861.00) and superannuation (£2116.00) (Chesters I, Accountant, Central Manchester and Manchester Children's University Hospitals NHS Trust: personal communication, 2004).

In addition to the actual examination of patients, consideration must be taken of the time taken to back up the three instruments. For the HFA, HRT and GDx, this process takes approximately 30, 20 and 45 minutes, respectively. Of course, the duration of the process is proportional to the amount of data that has been collected since the previous back-up and the size of the database files. To increase efficiency, the three pieces of equipment can be backed up simultaneously, with the operator switching between devices; this process takes more than 1 hour, during which all three instruments and the technician are unavailable for examining patients. To save time, the transfer of data to the off-site server can be performed overnight. There is potential for improved back-up, such as using more automated systems, in the future.

Consumables

The main consumables for the instruments are paper, print cartridges and removable back-up media.

The amount of paper used for all three instruments is equal as, for the most comprehensive printout, each requires one sheet per eye. Each device has the facility to generate and print a summary of longitudinal examinations but, again, there is little difference in the number of sheets required.

Currently, the HRT and GDx use a colour inkjet printer (Deskjet 1600C, Hewlett Packard), compared with the HFA's monochrome laser printer (Laserjet 4, Hewlett Packard). However, if a new clinic were to be established, it is probable that a colour laser printer would be used for the imaging devices, since the cost of these devices has dropped considerably since the beginning of the study. (The HRT and GDx require a colour printer, since some elements of the printout are colour-coded.) The price of colour and black toner cartridges for a Hewlett Packard Laserjet 3700 are £115.00 and £95.23 (including VAT) each, respectively; however, the printer requires three different colour cartridges plus a black cartridge (total cost = £440.23). Cartridges for a comparable monochrome printer (e.g. Hewlett Packard Laserjet 1300) are £52.87 (including VAT) each. Projections on the number of pages printed per cartridge are impossible to make, owing to multiple confounding factors such as amount of text per page, density of print and absorbency of paper.

The cost of backing up data is highly dependent on the medium employed. Initially, for HRT and GDx data, the study used 1-GB Jaz disks (Iomega, San Diego, CA, USA). However, this system proved to be both expensive and unreliable. Therefore, the HRT's computer was modified to incorporate a RAID level 4 system and a CD writer, and computers controlling both imaging devices were networked in order to store data on a university server. To compare the prices of removable storage media, the cost (including VAT) of a 1-GB Jaz disk is £62.27. For a pack of 50 CDs (each with 700 MB capacity), the price is £11.75. Therefore, the relative costs are £62.27 and £0.34 for Jaz disks and CDs, respectively.

Costings should allow for data security following a hard disk drive (HDD) failure and secure back-up of data to a remote file server or to removable media. For the HRT and GDx, back-up can be achieved by RAID systems on the PC and back-up to a networked file server. For the HFA, the current system involves saving all data to the HDD and floppy disk drive (FDD) (standard feature of the HFA) and backing up to FDDs. Differences between the HFA and the imaging devices reflect the much large storage capacity requirements of the latter. Costings for file server and networking are highly dependent on the status of existing facilities, such as whether or not the room where the instrument is housed is already networked and whether or not spare disk capacity is available on existing file servers. The cost of file security is, however, much greater for the imaging devices and can easily run into several thousands of pounds.

Overheads

The overheads for the three instruments are similar: the floor space needed for each is comparable. In an ideal situation, each test would be performed in isolation in a quiet room, where the lighting and heating conditions can be easily controlled. These limitations apply equally to all three devices, so there is no overall difference in the overheads for the HRT, HRT and GDx.

Capital and maintenance

The purchase price and cost of annual maintenance contracts are given in *Table 16*. For the HRT and HFA, prices are given for different models. However, data presented in this study could be collected with the most basic version of each, as features contained in the more sophisticated models are mainly employed within research environments and are not frequently used within normal clinical practice. This study used an HFA Model 740.

Discussion

This chapter has considered various elements of the cost of purchase and day-to-day operation of the HRT, GDx and HFA. However, the results from our study do not necessarily correspond to values relevant to a newly established imaging clinic, where a greater number of patients could potentially be examined. There is a variety of reasons for this disparity:

- If an HRT II were to be used, rather than its predecessor, then image capture times would be reduced. This decrease would also apply to the time taken to perform analysis, since when the latter duration was calculated from this study, the previous (lengthier) method of defining the margin of the optic disc was used.

TABLE 16 *Equipment costs*

	Purchase price (including VAT) (£)	Maintenance cost/year (including VAT) (£)
HRT I	No longer available	Bronze:[a] 1412.35 Silver:[b] 1777.78 Gold:[c] 2225.45
HRT II	30388.34[f]	Bronze:[a] 822.50 Silver:[b] £1036.35 Gold:[c] £1239.63
GDx	No longer available	N/A
GDx VCC	29369.13[f]	Basic contract: 1410.00 Gold standard contract: 2585.00[d]
HFA 720	13605.33	Basic contract: 548.73 Superior contract: £1175.00[e]
HFA 740	26226.38	
HFA 745	28399.75	
HFA 750	34577.90	

N/A, not applicable.
[a] Does not cover call-out charge, parts or consumables.
[b] Does not cover call-out charge or consumables.
[c] Does not cover consumables.
[d] Includes download of all data on to loan equipment.
[e] Includes emergency call-out and loan of equipment.
[f] Does not include costs of RAID system (or equivalent).

- Image capture would also be more rapid with the GDx VCC, since it only requires the capture of a single image for each eye. There would be a corresponding reduction in the time required to perform the analysis, because the operator would not have to review the original images and select, then compose, a mean image.
- Since the original design of the study, the SITA[135] on the HFA have become widely used in both clinical and research environments. Implementation of these algorithms would cause a reduction in perimetric testing time. This issue is discussed further in the section 'Discussion', point 4 (p. 34).
- The duration of analysis for the imaging devices (described in Chapter 4) included time taken to generate the printout with a colour ink jet printer. If a laser printer were used, the process would be more rapid.
- The PCs controlling the HRT and GDx were both replaced during the study and were of high specification at the time. However, functions such as recalling images from the hard disk are dependent on processor speed and would be quicker with more contemporary equipment.
- This study used the approach of capturing five images with the HRT and GDx, then selecting the best three for composition of the mean image. This very thorough strategy is not necessarily appropriate for a clinical environment, and capturing fewer images may lead to an increased rate of examination.
- The study used highly qualified/motivated staff, with strong backgrounds in both clinical practise and academia. If trained technicians were to be employed, then the rate of examination may be lower.

Overall, the purchase and running costs of the HRT and GDx are fairly comparable. Small disparities exist between the costs of maintenance contracts but these differences are small compared with the overall cost of the instruments. The cost of a basic HFA (Model 720) is less than half that of the imaging devices, as is its basic maintenance package. Back-up costs are also far cheaper for the HFA than either the HRT or GDx. Differences in the costs of print cartridges are relatively small. The main source of disparity in the financial efficiency of the equipment is the rate at which the examinations can be performed.

Chapter 6

Applicability of perimetry, the Heidelberg Retina Tomograph and GDx to an unselected population of patients

Introduction

In order to evaluate thoroughly the potential utility of a clinical test, it is important to derive an estimate of the proportion of an unselected group of patients who can successfully undergo the examination. This analysis is an essential component of appraising a system and is often not reported.[165,166]

Chapter 2 described some of the wide body of research that has investigated the HRT and GDx. A common factor of these studies is the use of strict exclusion criteria (e.g. constraints on refractive error, an absence of coexisting ocular pathology, level of minimum visual acuity). A feature of the cross-sectional arm of the Glaucoma Imaging Study that fundamentally differs from the above is its broad inclusion criteria: any Caucasian patient aged ≥ 40 years who was able to attend for examination was eligible to participate.

The aim was to attempt to perform perimetry, confocal SLO and SLP on all subjects and analyse the reasons that prevented the collection of good-quality data.

One of the original aims of examining this cross-sectional group of patients was to derive sensitivity and specificity estimates for the HRT and GDx; these calculations will be reported in Chapter 7. It must be noted that data from a different group of patients have been used to calculate diagnostic accuracy. This disparity resulted from the data loss suffered in the early stages of the study, where data from many of the patients reported in this chapter had been lost.

Method

Patients were examined according to the method described in the section 'Data collection' (p. 18). Patients were not excluded if their previous perimetric examinations had yielded unreliable results, even highly unreliable, as the aim was to produce findings that are relevant to a typical hospital-based glaucoma population.

Every reasonable attempt was made to complete each component of the examination. Mydriasis was performed if the pupil diameter was clearly too small for imaging (e.g. <1 mm) or if imaging was unsuccessful through the undilated pupil. (Perimetry was always performed prior to dilation.) When adequate HRT image quality could not be achieved, the decision that the examination had 'failed' was made by the operator. A threshold value of the GDx image quality score was not used to categorise acceptable images, since experience has shown that the GDx sometimes gives a high numerical value to an image that is clearly substandard [see the section 'Results' (p. 99)].

With the HRT and GDx, a 'failed' attempt at imaging was recorded if there was a complete failure to record an image or if the image quality was considered inadequate. Reasons for the latter included: images that were too dim and could not be remedied by increasing the intensity of the laser beam; images that consistently showed evidence of excessive eye movement (despite prompting the patient); and cases where the image could not be adequately centred (e.g. during GDx imaging when the patient could not follow the fixation target with the contra-lateral eye). Thresholds of features of adequate images were established for the research team to ensure consistency between operators; for example, an HRT image was judged to show too much eye movement if there was a shift of one-third of the disc diameter or more on playing the 'movie' of all 32 HRT sections.

When a patient's refractive error fell outside the range of the HRT or GDx, imaging was attempted through the subject's spectacles. The operators had experience of this technique, since the HRT requires patients with >1.00DC astigmatism to be examined through their refractive correction.

TABLE 17 *Reasons for unsuccessful examination (some patients are represented in more than one column)*

	Instrument		
Reason	**HFA**	**HRT**	**GDx**
Could not maintain position at instrument	3[a]	3[b]	2[c]
Cataract or media opacity	0	16	5
Contra-lateral eye unable to fixate	0	8	8
Head or eye movement/blinking	0	5	3
High refractive error	0	2	3
Instrument malfunction	0	1	0
Total	3	35	21

[a] One patient due to arthritis and two due to spondylitis.
[b] One patient due to arthritis, one due to spondylitis and one due to a previous cerebrovascular accident.
[c] One patient due to arthritis and one due to previous cerebrovascular accident.

TABLE 18 *Frequency of failed catch trials*

	Frequency of failed catch trials (%)		
Selection of patients	**≥ 20% FL**	**≥ 33% FP**	**≥ 33% FN**
Whole group (n = 176)	17.6	0.6	6.3
Patients unable to undergo HRT[a] (n = 35)	18.8	3.1	3.1
Patients unable to undergo GDx[b] (n = 21)	15.8	5.2	10.5

[a] $p = 0.0014$ (χ^2 test).
[b] $p \ll 0.001$ (χ^2 test).

In the eventuality of an unsuccessful examination with any of the instruments, the reason for this failure was recorded. The visual acuity and perimetric reliability indices were also noted. Patients were examined by experienced technicians; consequently, occurrences of failed examination cannot be attributed to operator inexperience.

Results

A total of 179 patients were examined, mean age 71.7 ± 12.1 years. The experimental group comprised 23 OHT patients and 156 patients with POAG, 82 females and 97 males, 94 right eyes and 85 left eyes. The mean logMAR visual acuity was 0.27 ± 0.25 (the score for subject 140 was excluded as his acuity was count fingers at a distance of 1 m). The mean spherical equivalent refractive error was –1.10 ± 6.20DS (range –23.50DS to +13.00DS).

Perimetry, HRT and GDx could not be performed on three (1.7%), 35 (19.6%) and 21 (11.7%) of patients, respectively. A total of 40 patients (22.3%) were unable to undergo examination with at least one of the three instruments; reasons are summarised in *Table 17* and shown graphically in *Figure 14*. Some patients were unable to be examined with more than one instrument - the Venn diagram (*Figure 15*) illustrates this distribution.

Perimetry was performed successfully on 176 subjects. Using widely accepted standards of reliability [<20% fixation losses (FLs), <33% false positives (FPs) and <33% false negatives (FNs)], reliable results were achieved from 144 subjects (81.8%). Of those with unreliable results, 22 cases (12.5%) were due to FLs, one case (0.6%) to FP and 11 cases (6.3%) to FN responses. One patient exceeded reliability thresholds for both FLs and FPs. The frequencies of failure to reach recommended reliability criteria, together with results for patients who were unable to undergo imaging, are shown in *Table 18*.

Of the patients who had difficulties with their positioning at the instruments, two were completely unable to achieve the correct posture at the perimeter owing to spondylitis; another was able to achieve the correct alignment, but could not maintain it owing to arthritic back pain. Also,

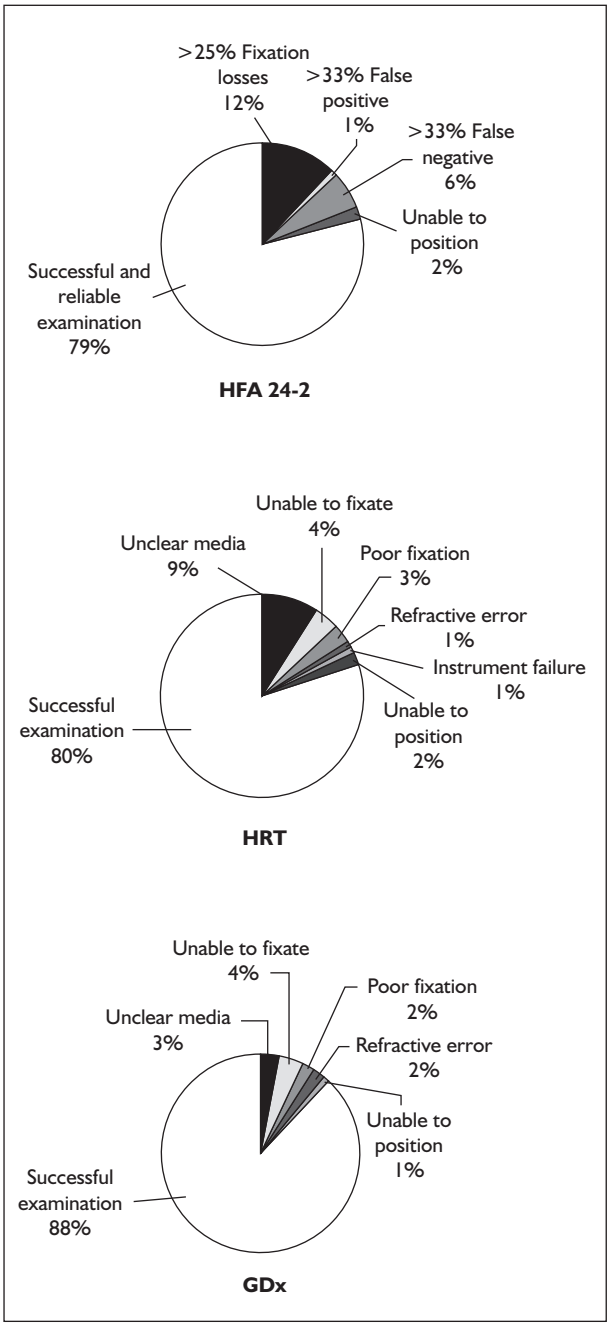


FIGURE 14 *Fate of examination of unselected population with HFA, HRT and GDx. Proportions are shown to the nearest 1%.*

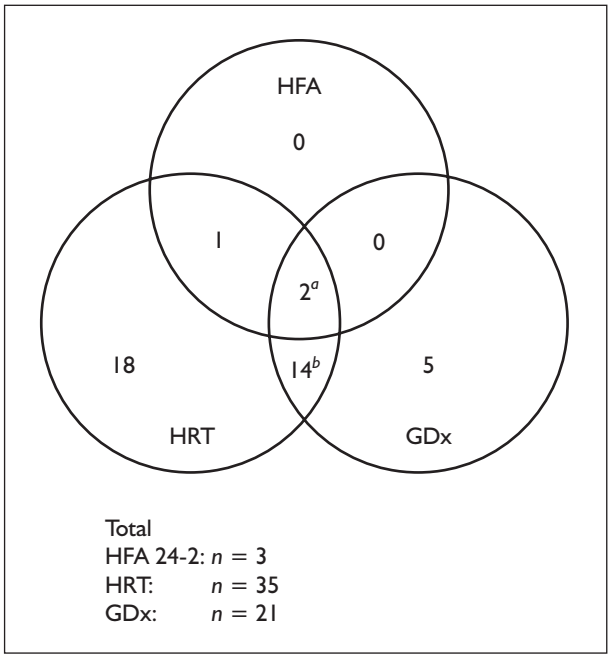


FIGURE 15 *Distribution of subjects unable to undergo examination procedures.* [a] *One subject due to arthritic back pain and one due to spondylitis (fields) and blinking (imaging).* [b] *One subject due to cerebrovascular accident, three to cataract, five to fixation, two to high refractive error and three to movement/blinking.*

imaging with both the HRT and GDx could not be performed on one patient who had previously suffered a cerebrovascular accident (this patient was, however, able to perform perimetry).

The presence of cataract or multiple vitreous floaters was a common cause of unsuccessful examination with the imaging devices, causing nearly half (HRT) and one-quarter (GDx) of all failures. (During GDx examination, the latter could potentially be confused with artefacts attributable to the tear film, but the author is familiar with both of these phenomena and could easily differentiate the two.) One patient was pseudophakic and had undergone a laser capsulotomy. However, the clear zone in the posterior capsule was not large enough to allow passage of the laser beam with either instrument.

A frequent cause of unsuccessful imaging was either very poor acuity or a grossly restricted field in the contra-lateral eye (one patient had a prosthetic eye). This occurrence accounted for 22.9% and 38.1% of failures with the HRT and GDx, respectively. Causes of poor acuity or field loss in the contra-lateral eye were (in decreasing order of prevalence) glaucoma, age-related macular degeneration, amblyopia and congenital macular scar.

Both imaging devices were equally affected by patients who were unable to keep their head and/or eye sufficiently still, or to control their blinking, causing 14.3% of unsuccessful attempts, which is clearly a problem, especially in older patients who may have a physiological tremor. Excessive blinking was often caused by anxiety, worsened by the proximity of the instrument and the brightness of the laser.

The lack of constraint on refractive error as an exclusion criterion in this study meant that the range of refractive errors examined was high (–23.25DS to +13.00DS). Performing imaging through the patient's spectacles (none of the experimental group wore contact lenses) was largely unsuccessful, owing to reflections from the lens surface. The mean absolute value of refractive error in the group who could not be imaged owing to refractive error was 17.13DS (HRT) and 15.75DS (GDx).

Discussion

The aim of this study was to reproduce a normal clinical scenario as closely as possible. In the literature, most research uses strict inclusion criteria and, despite sound scientific method, the results cannot be transferred to a typical hospital-based glaucoma population. The concept of such weaknesses within research has been described previously.[166] Of the many studies investigating sensitivity and specificity of the HRT and GDx, selection criteria may influence the outcomes of the investigations. Also, none of the research has considered the aspect of how widely the techniques can be applied. This study aimed to address this issue and is essentially different from most glaucoma research, in that there were no restrictions on inclusion and patients with poor visual acuity, high refractive error and coexisting ocular pathology were not excluded.

The proportion of these patients who could successfully undergo assessment with the 24-2 full threshold program of the HFA and the capture of five good-quality images with both the HRT and GDx was 98.3, 80.4 and 88.3%, respectively. Kremmer and colleagues[167] found that 4.7% of their population could not undergo SLP. However, the authors did not state the characteristics of their population with regard to acuity, refractive error and visual field status. Another study states that 92.6% could be imaged with the GDx,[119] although this study did not examine patients with poor acuity (>0.5), which would explain the higher success rate.

The applicability of polarimetry has also been examined by Vitale and colleagues:[168] from a large sample of 497 eyes of 249 patients, they determined the proportion of eyes that could undergo successful examination and analysis with the GDx and found a very similar result to the current study (88%). This finding is likely to be a manifestation of similarities in the experimental protocols, where a large number of individuals were examined and constraints of ocular parameters (e.g. refractive error, visual acuity, extent of visual field loss) were not placed on the experimental groups.

The apparently high applicability of perimetry from the current study only describes the number of patients who were able to perform the examination and does not taken into account reliability estimates. Current thresholds[169] are not based on any careful scientific study; infrequent sampling of FLs, FPs and FNs means that the estimates have very wide confidence limits.[133] Additionally, catch trial estimates have been shown to be poor predictors of visual field repeatability in glaucoma patients.[134]

When fields that exceed the standard thresholds for reliability are excluded, the proportion drops to 81.8% (similar to the proportion of patients who could be examined with the HRT). However, in considering the usefulness of the data, there is a fundamental difference between an unsuccessful attempt at imaging and a perimetric examination that exceeds the recommended reliability thresholds: if a patient is not successfully imaged with an instrument, there is no resulting clinical data on which a management decision could be made; in the scenario of a patient not reaching the recommended perimetric reliability criteria, some data are still available to aid the clinician. These facts suggest that, although an 'unreliable' field may not provide the clinician with the 'quality' of data he or she requires, it still may be sufficient to discriminate between a normal and abnormal result, and subsequent tests may yield more reliable data, as the patient becomes more familiar with the test. This situation contrasts with the scenario of a failed attempt at imaging from this study, where all reasonable efforts (e.g. mydriasis) were made to ensure a successful result and, therefore, additional examination would not bring further benefits – the data presented show the proportion of cases where there would be no data at all to aid the clinician.

Within this study, patients were categorised as having unreliable fields due to 'failing' a proportion of catch trials and incurring ≥ 20% FLs (17.6% of cases), ≥ 33% FPs (0.6%) or ≥ 33% FNs (6.3%). These rates compare fairly well to those from a large-scale study at Moorfields Eye Hospital (n = 62,784),[170] which found 17.4, 1.2 and 6.9%, respectively. Their overall rate of unreliability was higher than in this study (23.4% versus 18.8%). However, in comparing results, it

must be remembered that reliability is dependent on the severity of field defects in the experimental group, with more damaged fields associated with higher numbers of FN responses.[171]

The incidence of FLs was much higher than FP or FN responses. Patients who are accurate fixators can sometimes incur FLs if the blind spot is incorrectly located at the outset of the examination.[172] However, this occurrence is unlikely to be a causative factor in this study, since the perimetrist always replotted the blind spot if the patient failed the fixation checks. The re-establishment of the blind spot was always performed sufficiently early to prevent the patient 'failing' the critical 20% of trials if they were an otherwise steady fixator.

A follow-up study to that by Vitale and colleagues[168] found that perimetry could be successfully performed on 94% of eyes. However, there are several confounding factors in comparing research from the two studies. Vitale and colleagues used the Dicon LD 400 Autoperimeter 40-point suprathreshold program, which differs significantly from the HFA 24-2 full threshold examination; for example, the Dicon examination duration typically lies between 2 and 4 minutes,[173] which is far shorter than with the HFA (see Chapter 4); for patients who are unable to perform the examination for reasons such as postural disability, the length of the examination is of fundamental importance, since some patients may be able to tolerate the far shorter examination times. This factor, however, does not explain the lower applicability of perimetry by Vitale and colleagues, but further comparisons cannot be made because the authors do not give reasons for unsuccessful examinations. A comparison of the number of patients able to perform a 'reliable' perimetric test is also not possible since, although Vitale and colleagues give the range and mean of the frequency of failed catch trials, they do not state the number of individuals who exceeded recommended reliability criteria.

A reason for unsuccessful examination that affected all three instruments was patients who could not be adequately positioned at the instrument. In the three cases in this study, the patients were suffering from spondylitis, arthritic back pain and reduced mobility following a cerebrovascular accident. It is to be expected, with an age-related condition such as glaucoma, that age-related orthopaedic and systemic diseases will have an impact on test applicability.

Cataract and vitreous floaters were common causes of unsuccessful imaging. Owing to the lack of exclusion criteria, the approach used by the study was to attempt imaging on all patients. Unclear media contributed to 9% (HRT) and 3% (GDx) of unsuccessful examinations. Kremmer and colleagues[167] state that cataract was a common cause of failed examination with the GDx, but do not qualify the statement numerically.

When imaging could not be adequately performed on a patient owing to cataract, this study used the strategy of repeating the imaging under mydriasis, with the aim of improving image quality.[86] Despite this approach, cataract was still a major cause of failed examination.

This study did not use a defined threshold for determining unacceptable image quality with the HRT and GDx; in both cases, the decision was at the sole discretion of the operator. Each instrument does possess a quantitative description of image quality: the HRT calculates the MSD of the pixels in a mean image (derived from the height measurements of all picture elements), and the GDx produces a numerical quality score for each image. A limit of MSD was not used for the HRT because, although a limit of 50 μm is suggested by the manufacturers,[149] there is no widely accepted limit in the literature and, at the time the data were collected, the PC was not able to compute mean topography images. The GDx image quality score was not used because the value produced is not always a reliable indicator of image quality [see the section 'Results' (p. 99)]: occasionally, the GDx gives a high numerical value to an image that is poorly focussed or shows evidence of eye movement during image capture. An argument in favour of the study's protocol of using the technician's judgement in determining adequate image quality is the fact that most failed images were not borderline; in the majority of cases, unacceptable image quality was easy to define.

Cataracts were recorded by describing their morphology, together with a rough estimate of their density, in a similar manner to that often used in clinical notes. This method (rather than a grading system) was believed to be adequate, since no correlation has been shown for the HRT between the degree of cataract and the extent of image degradation.[174] The effect of cataract on GDx images is thought to be small,[175,176] which is consistent with the lower number of patients for whom cataract resulted in an unsuccessful examination with the GDx, compared with the

HRT. Minimum levels of acuity have been proposed which are thought to allow GDx images to be captured through cataract,[109,114] although this study found no clear threshold for either instrument.

An inability to fixate with the contra-lateral eye was a common reason for failed examination with the HRT and/or GDx. Contributory reasons were reduced acuity or very restricted visual field precluding a view of the fixation target, which was a particular problem with the GDx, since centration of the disc within the image is achieved by altering the position of gaze of the contra-lateral eye. Therefore, the proportion of patients who could not undergo imaging owing to difficulty with fixation was higher for the GDx than the HRT (38.1% versus 22.9%). Of interest, the shorter image capture time may have a positive bearing on the GDx, since there were two patients (one whose fellow eye was completely blind and the other who had an artificial eye) who were examined successfully with the GDx. The successful outcome was because the examiner was able to establish excellent cooperation and the patients were able to maintain steady fixation; albeit neither could do this for the longer capture times of the HRT.

The successors to the HRT and GDx, the HRT II and GDx VCC, each use an ipsi-lateral fixation target,[149,177] which will obviously overcome the problems described above. However, it may lead to difficulty examining the eye with poor acuity, especially in cases of ocular co-morbidity, such as age-related macular degeneration.

Both imaging devices could not examine patients with very high refractive error. The HRT is capable of examining patients who lie in the range –11.75 to +11.75DS and the GDx –10.00 to +10.00DS [in reality, a wider range can possibly be examined (Rennie–Cassell JR, Laser Diagnostic Technologies, California, USA: personal communication, 2001) although the limits of this are not given], but the study group contained three patients who lay outside these ranges. For these patients, attempts were made to perform the test through the patient's spectacles. This was largely unsuccessful owing to reflections from the lenses precluding successful imaging.

The number of patients in the study with large degrees of ametropia was low, which prevents meaningful comparison between myopia and hypermetropia. None of the patients within the study group wore contact lenses. Further research is necessary to determine whether the use of such devices would provide an adequate solution. An additional issue for investigation is the validity of GDx assessment of patients with high refractive error, since all measurements are compared with a normative database, which was compiled from individuals with <5.00DS ametropia.

The exclusion of patients who had problems with communication or memory [see the section 'Cross-sectional patients' (p. 17)] may have led to selection bias and may be a source of error in that the experimental group does not truly represent a population of glaucoma patients, where such conditions are not uncommonly encountered.

It would be useful to look at the relationship between successful imaging and the grade of cataract, which would allow further analysis regarding the likelihood of HRT and GDx examination in the presence of cataract. Unfortunately, the relevant supplementary assessment cannot be performed retrospectively. Also, analysis of HRT MSD and GDx image quality score would have provided an interesting insight into cases of failed examination. However, the latter analysis cannot be performed from the population described in this chapter, since much of the data, which were collected in the early period of the study, have been lost.

Chapter 7

Cross-sectional study

The main purpose of the cross-sectional study is to determine the sensitivity and specificity of the HRT and GDx for glaucomatous visual field loss. The study used the presence of a glaucomatous visual field defect as the diagnostic indicator of glaucoma from which the diagnostic precision of the HRT and GDx was determined.

Introduction

The HRT and GDx provide rapid and objective examination of the ONH and RNFL, respectively, and consequently have the theoretical potential to provide accurate diagnosis of glaucoma. The diagnostic precision of both instruments has been widely investigated in previous research. However, there are several problems in comparing such studies. These issues, and the subsequent approach taken by the glaucoma imaging study, are discussed in the following four subsections.

Definitions

Prior to the determination of sensitivity and specificity, a fundamental problem is **how** actually to define glaucoma. In the clinical and research settings, cases of glaucoma are usually identified by the presence of visual field defects and/or the appearance of the ONH and RNFL. When performing an evaluation, it is important to have a 'gold standard' classifier that is independent of the test being evaluated.[94] This exclusion prevents artificial inflation of the apparent sensitivity of the technique. Consequently, since this study has been investigating systems that examine the ONH and RNFL, glaucoma has been defined on the basis of visual fields alone. This strategy may incur a disparity between a patient's research and clinical diagnostic classifications.

A further issue, which is also of great relevance to this study, is defining the boundary between normality and pathology. Moving this limit can profoundly affect the apparent diagnostic accuracy of an examination modality. The absence of a universally accepted definition of glaucoma contributes to this problem.[1] This research, therefore, took the approach of comparing sensitivity and specificity estimates derived following categorisation of the experimental group by three different perimetric boundary definitions.

Inclusion criteria

An additional feature that can significantly affect diagnostic precision is the level of disease in the experimental group: if all the patients within the cross-sectional group had advanced glaucoma, it would not pose a diagnostic challenge to the equipment and the apparent sensitivity would be artificially high. The issue of inclusion and stratification of the experimental group has been widely discussed in the literature and underlies the criticism of some widely quoted research (e.g. by Tjon-Fo-Sang and Lemij[117]) that has reported inflated levels of diagnostic precision. Therefore, some of the analysis described in this chapter was performed subsequent to subclassification of the experimental group by the level of perimetric loss.

Specificity

In evaluating previous research, an obstacle to making comparisons between studies is the widespread use of isolated sensitivity and specificity values. For the purposes of screening, it is generally accepted that high specificity is required when dealing with diseases of low prevalence (approximately 1% of the population ≥ 40 years of age has undetected glaucoma). Therefore, where appropriate, this study chose to set the cut-off criteria to give a specificity of 95%, which is considered to be appropriate for glaucoma.[178]

Supplementary analysis

Some studies that have reported high sensitivity and specificity have done so via sophisticated mathematical manipulation of HRT or GDx parameters.[107,117] These modifications are not available with the instruments' default software and were not, therefore, used in this research, whose main objective was to report on the utility of these instruments in the normal clinical environment.

Method

Patients were recruited and examined according to the procedures detailed in Chapter 3.

Patient groups

Data from one, randomly selected, eye of 250 subjects were included in this study. Of this total, 98 were normal control individuals and 152 were patients from the cross-sectional arm of the study. The latter group comprised patients who had been recruited from the general and glaucoma clinics at MREH – they were not patients who met any consistent definition of glaucoma. In most cases, glaucoma patients were receiving treatment.

Originally, 179 glaucoma patients had been examined in the cross-sectional category (owing to over-achievement of the initial recruitment target). However, as described in Chapter 6, the three examinations could not be successfully performed on some patients. From the initial cohort, 139 patients were amenable to examination with all three instruments. Unfortunately, following the data loss [see the section 'Data loss' (p. 24)], data remained for only 61 patients (see *Table 5*). (This experimental group was badly affected by the database malfunction that occurred in the early phases of the research project.)

Therefore, in order to improve the confidence limits of the sensitivity and specificity estimates, additional data were added to the original cross-sectional group to increase its size. The supplementary data came from 48 longitudinal patients (from both the high risk and glaucoma groups) who were unsuitable for inclusion in the longitudinal arms owing to lack of data or early discontinuation. In addition, randomly selected data sets ($n = 43$) from the longitudinal glaucoma category were added to increase the group size to 152.

Analysis

Statistical analysis was performed with Excel 2000 (Microsoft) and MedCalc version 7.3.0.1 (Frank Schoonjans). The area under the receiver operating characteristic (ROC) curve[179–182] [with 95% confidence intervals (CIs)] was established for all the global HRT parameters ($n = 22$) (the parameter rim:disc area ratio was not included in the analysis, since there is a constant relationship between this parameter and C:D area ratio, which was incorporated) and GDx parameters ($n = 28$). A rough guide for classifying the accuracy of the area underneath a ROC curve is shown in *Table 19*.[180,182–184] For the HRT DFAs[87,88] the area underneath the ROC curve was plotted for the DFA statistic. The threshold of the DFA statistic that yielded a specificity of 95% was calculated, as were the sensitivity and specificity when zero cut-offs were used.

TABLE 19 *Interpretation of area underneath ROC curve*

Area under ROC curve	Interpretation
0.97–1.00	Excellent
0.92–0.96	Very good
0.75–0.91	Good
0.60–0.74	Fair
0.50–0.59	Very poor

This analysis was performed following classification of the experimental group by each of the three perimetric definitions listed below.

Definition 1: The fifth centile of the global MPD of the normal group was calculated. Patients with an MPD lying below this level were categorised as abnormal.

Definition 2: Perimetric data were analysed by the method described in the section 'Analysis of visual field data' (p. 20), which gives a probability value for asymmetry between superior and inferior visual field sectors. The concept of comparing the sensitivity of zones above and below the horizontal mid-line is well established for glaucoma diagnosis[60] and is the basis of the GHT used by the HFA.

For each eye, the number of areas with $p \geq 0.95$ was counted. Any patient with one or more sectors lying outside this threshold was defined as abnormal.

Definition 3: The same principle as Definition 2, except a cut-off of $p \geq 0.99$ was used.

With each perimetric definition, subjects from the control group whose visual fields were categorised as normal were entered into the analysis to generate specificity estimates. Patients from the cross-sectional group who were found to have abnormal fields with each system were used to calculate sensitivity values. Individuals from the cross-sectional category who had normal visual fields (according to each criterion) were categorised as 'at risk'.

In addition to analysis of the HRT stereometric parameters, the output from MRA was also investigated. The MRA outputs three categories (normal, borderline and outside normal limits), as

described in the section 'Diagnostic algorithms' (p. 8). To determine fully the sensitivity and specificity of the MRA, two sub-classifications were employed: MRA_1, where borderline results were treated as abnormal, and MRA_2, where borderline results were treated as normal.

Where appropriate, levels of sensitivity at approximately 95% specificity were estimated. In addition, positive and negative likelihood ratios were calculated, using the formulae sensitivity/(1 – specificity) and (1 – sensitivity)/specificity, respectively.

Sensitivity of the MRA and GDx number were also recalculated after the experimental group had been stratified by the degree of perimetric loss. The categories used were: mean pattern deviation from –1.3 to –2.00 dB, –2.01 to –3.00 dB, –3.01 to –6.00 dB and <–6.01 dB.

TABLE 20 *Characteristics of experimental groups*

Characteristic	Normal group	Cross-sectional group
Number	98	152
Minimum age (years)	39.5	42.0
Maximum age (years)	87.8	95.6
Median age (years)	59.7	69.0
Right eyes (%)	44.4	49.3
Left eyes (%)	55.6	50.7
Male (%)	43.4	58.6
Female (%)	56.6	40.4

Results

Characteristics of the experimental groups are detailed in *Table 20* and *Figure 16*. The mean age of the normal controls and cross-sectional groups was significantly different ($p \ll 0.01$; two-tailed Student's *t*-test). The visual field MPD was –0.52 dB (95% CI: –0.60 to –0.43 dB) and –3.29 dB (95% CI: –3.77 to –2.80 dB) for the control and cross-sectional groups, respectively (see *Figure 17*); this difference is highly statistically significant ($p \lll 0.001$; two-tailed Student's *t*-test). The number of patients with abnormal visual field segments is shown in *Figure 18*.

The fifth centile of MPD for the normal group was –1.3 dB; this value was used as the threshold of normality for definition 1. *Table 21* shows the proportion of patients from each of the experimental groups who are defined as having normal and abnormal visual fields by each of the methods. The resulting number of subjects in each subgroup is shown in *Table 22*.

HRT parameters

For each visual field definition, the areas underneath the ROC curve for the HRT parameters are given in *Table 23*. In all cases, optimum discrimination between glaucoma patients and normal individuals was provided by the Bathija DFA.[88] Other parameters that performed favourably were C:D area ratio, vertical C:D ratio, mean RNFL thickness and the Mikelberg DFA; there was a large degree of

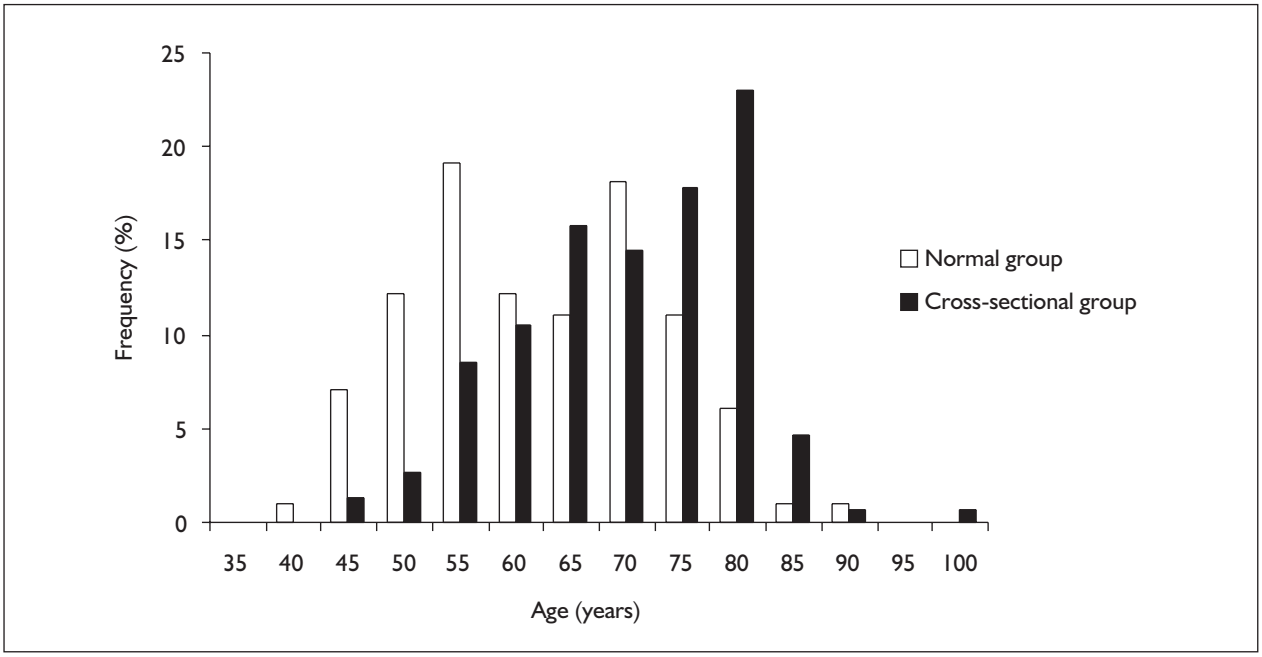


FIGURE 16 *Age distribution of subjects*

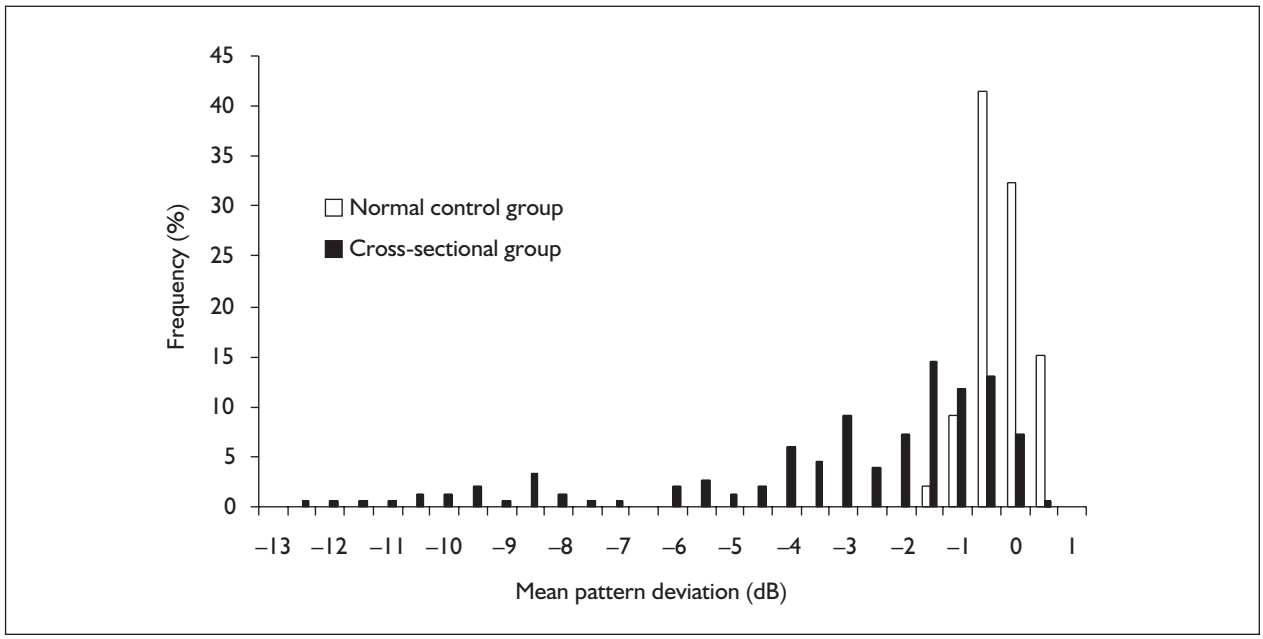


FIGURE 17 *Distribution of mean pattern deviation*

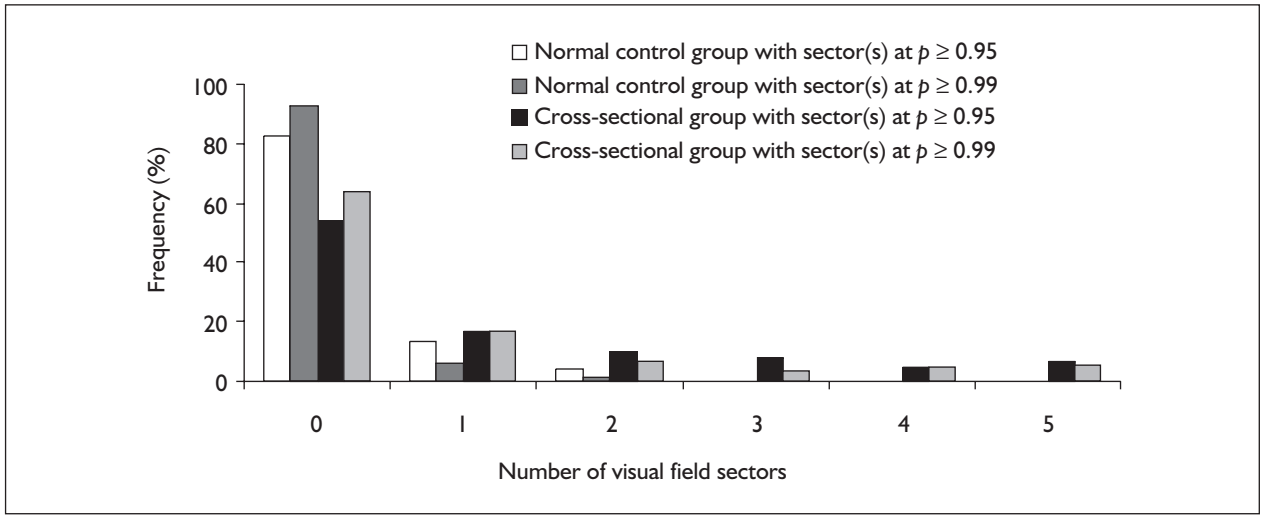


FIGURE 18 *Frequency of abnormal visual field segments. Minimum = 0; maximum = 5*

TABLE 21 *Outcome of stratification of experimental group by different criteria (95% CIs in parentheses)*

Experimental group	Outcome of visual field classification	Visual field definition 1 (%)	Visual field definition 2 (%)	Visual field definition 3 (%)
Normals	Normal fields	95.9 (90.0 to 98.4)	82.7 (74.0 to 88.9)	92.9 (86.0 to 96.5)
	Abnormal fields	4.1 (1.6 to 10.0)	17.4 (11.1 to 26.0)	7.14 (3.5 to 14.0)
Cross-sectional	Normal fields	28.3 (21.7 to 35.9)	54.6 (46.7 to 62.3)	63.8 (55.9 to 71.0)
	Abnormal fields	71.7 (64.1 to 78.3)	45.4 (37.7 to 53.3)	36.2 (29.0 to 44.1)

TABLE 22 *Number of subjects in experimental subpopulations*

	Number of subjects		
Category	**Visual field definition 1**	**Visual field definition 2**	**Visual field definition 3**
Normal	94	81	91
At risk	42	82	96
Glaucoma	109	69	55

commonality between the three visual field definitions. For Bathija DFA, the areas underneath the ROC curves were 0.831, 0.834 and 0.865 for visual field definitions 1, 2 and 3, respectively.

ROC graphs for the four 'best' HRT parameters are shown in *Figure 19*. In all cases, there was no significant difference in area under the ROC curve[182] between the parameters shown. To highlight the diagnostic accuracy at 95% specificity, *Table 24* shows the sensitivity (together with positive and negative likelihood ratios) of these parameters.

The DFA sensitivities shown above correspond to a specificity of approximately 95%. However, the cut-off of the DFA statistic to achieve this level was not zero. For each of the three categories, the cut-offs of the Bathija DFA were ≤0.059, ≤–0.087 and ≤0.059, respectively. The corresponding thresholds were ≤–0.973, ≤–0.951 and ≤–0.671 for the Mikelberg DFA. *Table 25* shows the sensitivities and specificities that were achieved when a level of zero was used. (The latter are consistent with the diagnostic accuracy of the HRT, since the software uses a cut-off of zero.)

Moorfields regression analysis

The results of the MRA are shown in *Table 26*. The highest sensitivity was 75% (with a corresponding specificity of 75%), which was for MRA_1 using visual field definition 1. The highest specificity (96% with sensitivity of 58%) was for visual field definition 3 determined with MRA_2. For all three visual field criteria, the group categorised as 'at risk' had a higher rate of MRA abnormality than the normal control group (for both MRA_1 and MRA_2).

GDx parameters

The areas under the ROC curves (together with 95% CIs) for the GDx are given in *Table 27*. The GDx parameters the number, maximum modulation, superior maximum/nasal median, ellipse modulation and inferior maximum/nasal median feature in the four 'best' for each visual field definition. ROC curves for these parameters are shown in *Figure 20*. From visual field definition 1, the number had a significantly larger area under the ROC curve[182] than ellipse modulation ($p = 0.011$), maximum modulation ($p = 0.002$) and superior maximum/nasal median ($p = 0.03$). In definition 2, a significant disparity in area was found between the number and ellipse modulation ($p = 0.027$) and maximum modulation ($p = 0.001$). From definition 3, a significant difference was found between the number and ellipse modulation ($p = 0.035$) and inferior maximum/nasal media ($p = 0.011$).

Table 28 gives the sensitivity, at approximately 95% specificity, of the four optimum parameters for each visual field definition; the positive and negative likelihood ratios are also given. Although the GDx number featured amongst the best parameters, the threshold of this parameter differed for definition 2. The optimum threshold for the three groups was ≥ 42, ≥ 36 and ≥ 42, respectively. The sensitivities and specificities for recommended thresholds[185] of the GDx number are shown in *Table 29*.

Stratification by level of damage

Table 30 shows the output of the MRA stratified by the level of perimetric loss. At the most severe level of damage (i.e. MPD <–6.01 dB), MRA_1 detected 83% of cases (at a specificity of 75%). Higher specificity (95%) was associated with a sensitivity of 79% for MRA_2. For patients with very early field loss, the sensitivities of MRA_1 and MRA_2 are 66 and 31%, respectively.

Similar analysis for the GDx number is shown in *Table 31*. Sensitivity is shown for three different cut-off values of the number: 42 (derived to give 95% specificity), and the manufacturer's recommended levels of 30 and 70. For the lowest threshold (i.e. number ≥ 31), the sensitivity is 75% for diagnosing patients in the category with the most advanced visual field loss (the specificity at this level is 86%).

TABLE 23 *Area underneath ROC curve for HRT parameters (95% CIs in parentheses)*

Global HRT parameter	Area under ROC curve		
	Visual field definition 1	**Visual field definition 2**	**Visual field definition 3**
Average variability	0.753 (0.688 to 0.81)	0.767 (0.691 to 0.832)	0.758 (0.680 to 0.825)
Bathija DFA	0.831 (0.773 to 0.879)	0.834 (0.765 to 0.89)	0.865 (0.799 to 0.916)
Cup area	0.782 (0.72 to 0.837)	0.788 (0.714 to 0.851)	0.801 (0.727 to 0.863)
Cup shape measure	0.750 (0.685 to 0.807)	0.782 (0.708 to 0.845)	0.796 (0.721 to 0.858)
Cup volume	0.757 (0.693 to 0.814)	0.748 (0.671 to 0.815)	0.773 (0.696 to 0.838)
C:D area ratio	0.806 (0.745 to 0.857)	0.802 (0.729 to 0.863)	0.827 (0.756 to 0.885)
Height variation contour	0.620 (0.550 to 0.684)	0.529 (0.446 to 0.611)	0.585 (0.500 to 0.665)
Horizontal cup:disk ratio	0.733 (0.668 to 0.792)	0.726 (0.648 to 0.796)	0.749 (0.671 to 0.817)
Infero-temporal contour line modulation	0.762 (0.698 to 0.818)	0.738 (0.659 to 0.806)	0.749 (0.671 to 0.817)
Maximum contour depression	0.637 (0.568 to 0.703)	0.691 (0.61 to 0.764)	0.678 (0.595 to 0.753)
Maximum contour elevation	0.772 (0.709 to 0.828)	0.725 (0.646 to 0.794)	0.761 (0.683 to 0.827)
Maximum cup depth	0.625 (0.555 to 0.691)	0.602 (0.519 to 0.681)	0.624 (0.541 to 0.703)
Mean cup depth	0.702 (0.635 to 0.763)	0.690 (0.609 to 0.763)	0.712 0.632 to 0.784)
Mean RNFL thickness	0.792 (0.730 to 0.845)	0.749 (0.672 to 0.816)	0.804 (0.730 to 0.865)
Mikelberg DFA	0.770 (0.707 to 0.826)	0.791 (0.717 to 0.853)	0.826 (0.755 to 0.884)
Reference height	0.499 (0.429 to 0.57)	0.523 (0.44 to 0.605)	0.514 (0.430 to 0.597)
Rim area	0.741 (0.675 to 0.799)	0.728 (0.649 to 0.797)	0.795 (0.721 to 0.858)
Rim volume	0.782 (0.72 to 0.836)	0.747 (0.669 to 0.814)	0.806 (0.733 to 0.867)
Rim:disc area ratio	0.806 (0.745 to 0.857)	0.802 (0.729 to 0.863)	0.827 (0.756 to 0.885)
RNFL cross-sectional area	0.755 (0.691 to 0.812)	0.710 (0.631 to 0.781)	0.783 (0.708 to 0.847)
Supero-temporal contour line modulation	0.719 (0.652 to 0.779)	0.722 (0.643 to 0.792)	0.786 (0.710 to 0.849)
Vertical C:D ratio	0.815 (0.755 to 0.865)	0.807 (0.734 to 0.867)	0.819 (0.747 to 0.878)

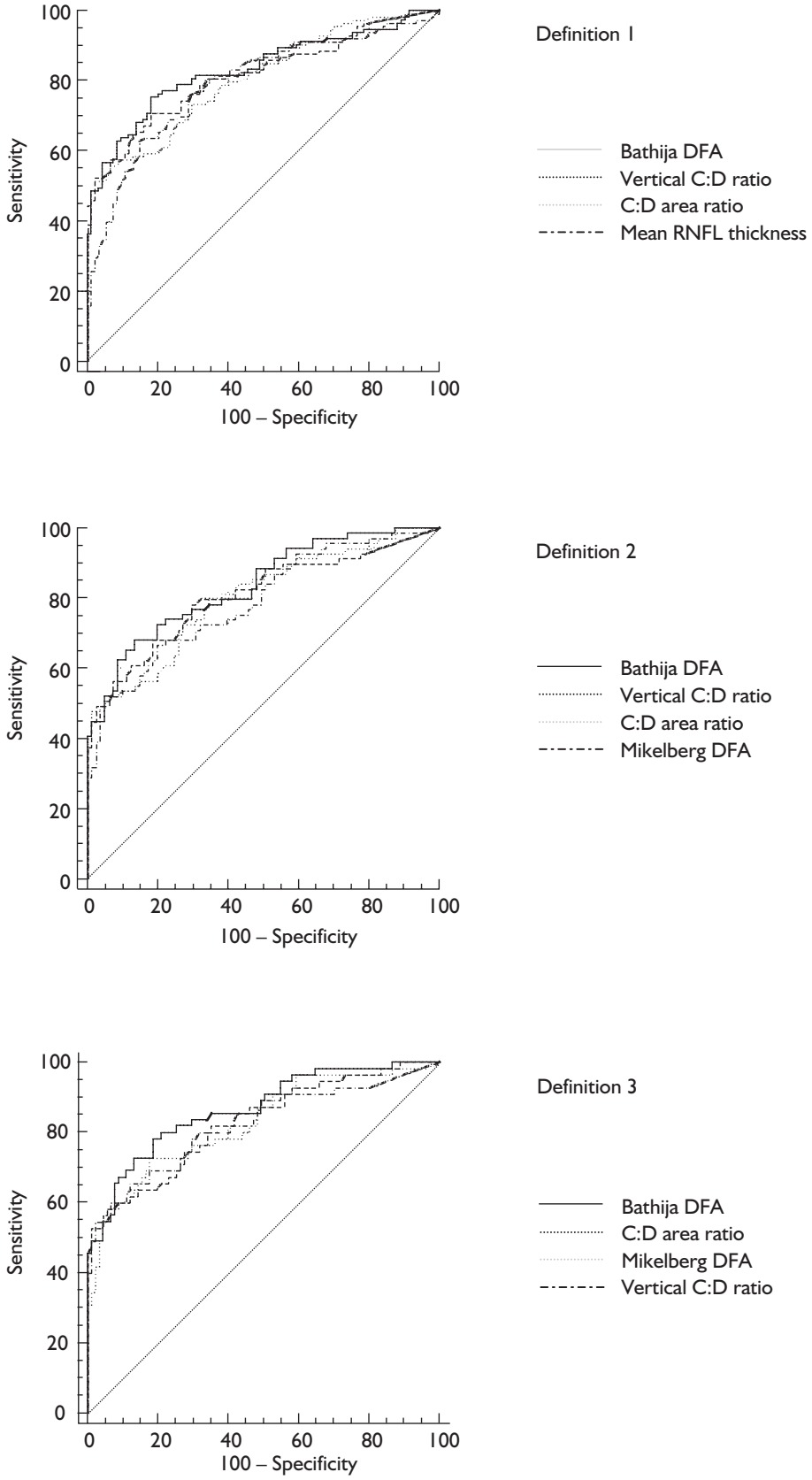


FIGURE 19 *ROC curves for HRT parameters*

TABLE 24 *Diagnostic precision of HRT parameters at 95% specificity (95% CIs in parentheses)*

	Sensitivity (%)	Positive likelihood ratio	Negative likelihood ratio
Visual field definition 1: specificity = 94.7% (88.0 to 98.2)			
Bathija DFA	56.6 (47.0 to 65.9)	10.65	0.46
Vertical C:D ratio	53.1 (43.5 to 62.5)	9.58	0.50
C:D area ratio	51.3 (41.7 to 60.8)	9.65	0.51
Mean RNFL thickness	34.5 (25.8 to 44.0)	6.49	0.69
Visual field definition 2: specificity = 95.1% (87.8 to 98.6)			
Bathija DFA	44.9 (32.9 to 57.4)	9.10	0.58
Vertical C:D ratio	49.3 (37.0 to 61.6)	9.98	0.53
C:D area ratio	47.8 (35.6 to 60.2)	9.68	0.55
Mikelberg DFA	49.3 (37.0 to 61.6)	9.98	0.53
Visual field definition 3: specificity = 94.5% (87.6 to 98.2)			
Bathija DFA	54.5 (40.6 to 68.0)	9.93	0.48
C:D area ratio	58.2 (44.1 to 71.3)	10.59	0.44
Mikelberg DFA	56.4 (42.3 to 69.7)	10.26	0.46
Vertical C:D ratio	54.5 (40.6 to 68.0)	9.93	0.48

TABLE 25 *Diagnostic characteristics of HRT discriminant function analyses (95% CIs in parentheses)*

	Visual field definition 1	Visual field definition 2	Visual field definition 3
Bathija DFA[88]			
Sensitivity at zero threshold of DFA statistic	53.1 (43.5 to 62.5)	49.3 (37.0 to 61.6)	52.7 (38.8 to 66.3)
Specificity at zero threshold of DFA statistic	95.7 (89.5 to 98.8)	95.1 (87.8 to 98.6)	95.6 (89.1 to 98.8)
Mikelberg DFA[87]			
Sensitivity at zero threshold of DFA statistic	63.7 (54.1 to 72.6)	65.2 (52.8 to 76.3)	70.9 (57.1 to 82.4)
Specificity at zero threshold of DFA statistic	79.8 (70.2 to 87.4)	80.2 (69.9 to 88.3)	82.4 (73.0 to 89.6)

TABLE 26 *Results from MRA of HRT data (95% CIs in parentheses)*

MRA Result[a]		Visual field definition 1		Visual field definition 2		Visual field definition 3	
		MRA_1	MRA_2	MRA_1	MRA_2	MRA_1	MRA_2
Normal	Negative	74.5 (64.8 to 82.2)	94.7 (88.2 to 97.7)	74.1 (63.6 to 82.4)	95.1 (88.0 to 98.1)	76.9 (67.3 to 84.4)	95.6 (89.2 to 98.3)
	Positive	25.5 (17.8 to 35.2)	5.3 (2.3 to 11.9)	25.9 (17.6 to 36.4)	4.9 (1.9 to 12.0)	23.1 (15.6 to 32.7)	4.4 (1.7 to 10.8)
At risk	Negative	53.5 (38.9 to 67.5)	79.1 (64.8 to 88.6)	31.3 (22.4 to 41.9)	31.3 (43.6 to 64.5)	34.0 (25.4 to 43.9)	57.7 (48.0 to 67.1)
	Positive	46.5 (32.5 to 61.1)	20.9 (11.4 to 35.2)	68.7 (58.1 to 77.6)	45.8 (35.5 to 56.5)	66.0 (56.1 to 74.6)	42.3 (32.9 to 52.2)
Glaucoma	Negative	24.8 (17.6 to 33.6)	41.3 (32.5 to 50.7)	34.8 (24.6 to 46.6)	49.3 (37.8 to 60.8)	30.9 (20.3 to 44.0)	41.8 (29.7 to 55.0)
	Positive	75.2 (66.4 to 82.4)	58.7 (49.3 to 67.5)	65.2 (53.5 to 75.4)	50.7 (39.2 to 62.2)	69.1 (56.0 to 79.7)	58.2 (45.0 to 70.3)

[a] Negative, MRA output 'normal'; positive, MRA output 'abnormal'.

TABLE 27 *Area under ROC curve for GDx parameters (95% CIs in parentheses)*

	Area under ROC curve		
GDx parameter	**Definition 1**	**Definition 2**	**Definition 3**
Average thickness	0.609 (0.539 to 0.677)	0.547 (0.464 to 0.629)	0.568 (0.483 to 0.649)
Ellipse modulation	0.763 (0.698 to 0.82)	0.753 (0.677 to 0.82)	0.747 (0.668 to 0.815)
Inferior integral	0.536 (0.464 to 0.606)	0.563 (0.479 to 0.643)	0.536 (0.452 to 0.619)
Inferior maximum	0.693 (0.624 to 0.755)	0.636 (0.553 to 0.713)	0.635 (0.551 to 0.713)
Inferior maximum/nasal median	0.741 (0.676 to 0.8)	0.719 (0.64 to 0.789)	0.719 (0.638 to 0.79)
Inferior ratio	0.693 (0.624 to 0.756)	0.633 (0.55 to 0.71)	0.629 (0.545 to 0.708)
Inferior/nasal integral	0.632 (0.561 to 0.698)	0.578 (0.495 to 0.659)	0.584 (0.499 to 0.665)
Inferior/temporal integral	0.654 (0.584 to 0.719)	0.588 (0.505 to 0.668)	0.580 (0.496 to 0.662)
Maximum modulation	0.759 (0.694 to 0.816)	0.733 (0.655 to 0.802)	0.719 (0.639 to 0.791)
Nasal maximum	0.601 (0.53 to 0.669)	0.600 (0.517 to 0.679)	0.588 (0.504 to 0.668)
Nasal integral	0.547 (0.476 to 0.617)	0.598 (0.515 to 0.677)	0.584 (0.500 to 0.665)
Nasal median	0.494 (0.423 to 0.565)	0.550 (0.467 to 0.631)	0.544 (0.46 to 0.627)

continued

TABLE 27 *Area under ROC curve for GDx parameters (95% CIs in parentheses) (cont'd)*

GDx parameter	Area under ROC curve		
	Definition 1	**Definition 2**	**Definition 3**
Number	0.84 (0.782 to 0.887)	0.827 (0.757 to 0.884)	0.820 (0.748 to 0.879)
Superior integral	0.576 (0.504 to 0.644)	0.511 (0.428 to 0.594)	0.546 (0.462 to 0.529)
Superior maximum	0.724 (0.657 to 0.784)	0.712 (0.632 to 0.783)	0.715 (0.634 to 0.786)
Superior maximum/nasal median	0.753 (0.687 to 0.81)	0.782 (0.707 to 0.845)	0.784 (0.708 to 0.847)
Superior ratio	0.734 (0.667 to 0.793)	0.718 (0.639 to 0.789)	0.712 (0.631 to 0.784)
Superior/inferior integral	0.523 (0.452 to 0.594)	0.610 (0.528 to 0.689)	0.632 (0.548 to 0.71)
Superior/nasal integral	0.645 (0.575 to 0.711)	0.637 (0.554 to 0.714)	0.658 (0.575 to 0.735)
Superior/temporal integral	0.668 (0.599 to 0.733)	0.633 (0.551 to 0.71)	0.642 (0.558 to 0.72)
Symmetry	0.518 (0.447 to 0.589)	0.615 (0.532 to 0.694)	0.611 (0.527 to 0.691)
Temporal integral	0.599 (0.528 to 0.667)	0.619 (0.536 to 0.697)	0.596 (0.512 to 0.677)
Temporal maximum	0.512 (0.441 to 0.583)	0.516 (0.433 to 0.598)	0.507 (0.423 to 0.591)
Temporal median	0.521 (0.45 to 0.591)	0.523 (0.440 to 0.605)	0.512 (0.428 to 0.596)
Temporal median/nasal median	0.498 (0.427 to 0.569)	0.529 (0.446 to 0.611)	0.543 (0.459 to 0.626)
Temporal/nasal integral	0.573 (0.502 to 0.642)	0.554 (0.471 to 0.636)	0.533 (0.448 to 0.616)
Total polar average	0.663 (0.594 to 0.728)	0.596 (0.513 to 0.675)	0.615 (0.531 to 0.694)
Total polar integral	0.527 (0.456 to 0.597)	0.554 (0.471 to 0.635)	0.521 (0.437 to 0.605)

Discussion

This study has examined the ability of two instruments, the HRT and GDx, to identify patients with glaucomatous visual field loss. It has the methodological advantage of comparing the diagnostic accuracy of the two instruments on one population, whereas many other studies review only one device.[90,92,108,186,187] Overall, the results show that both techniques fail to detect a significant number of cases of glaucoma. This finding, which is in line with many other researchers, is likely to be associated with the large overlap in the appearance of the ONH in normal and glaucomatous eyes.[90,188,189]

The Glaucoma Imaging Study used three different perimetric methods to define patients with glaucomatous visual field loss. Patients were defined on the basis of their visual field status alone so as not to induce an apparent increase in the sensitivity of the HRT and GDx by using ONH and RNFL characteristics as diagnostic criteria.[94] However, the use of a perimetric definition is, itself, associated with difficulties since, currently, there is no universally agreed definition of glaucoma. Although several systems have been described in the literature,[132,190,191] none has become accepted as definitive.[192] Therefore, this study took the approach of using three different methods of defining glaucoma. Visual field

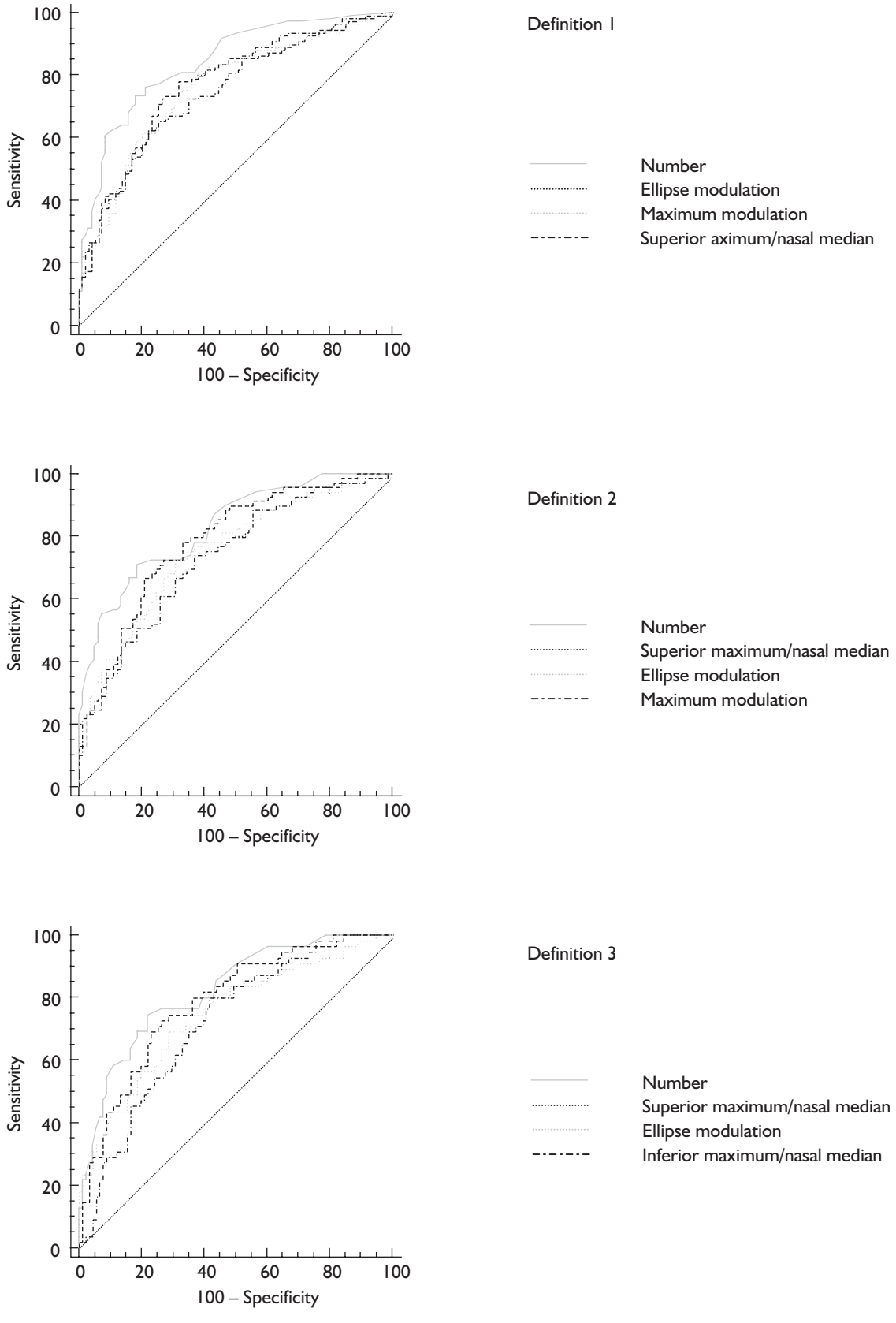


FIGURE 20 *ROC curves for GDx parameters*

TABLE 28 *Diagnostic precision of GDx parameters at 95% specificity (95% CIs in parentheses)*

	Sensitivity (%)	Positive likelihood ratio	Negative likelihood ratio
Visual field definition 1: specificity = 94.7% (88.0 to 98.2)			
Number	40.4 (31.1 to 50.2)	7.59	0.63
Ellipse modulation	26.6 (18.6 to 35.9)	5.00	0.78
Maximum modulation	25.7 (17.8 to 34.9)	4.83	0.78
Superior maximum/nasal median	26.6 (18.6 to 35.9)	5.00	0.78
Visual field definition 2: specificity = 95.1% (87.8 to 98.6)			
Number	44.9 (32.9 to 57.4)	9.10	0.58
Superior maximum/nasal median	24.6 (15.1 to 36.5)	4.99	0.79
Ellipse modulation	29.0 (18.7 to 41.2)	5.87	0.78
Maximum modulation	23.2 (13.9 to 34.9)	4.70	0.81
Visual field definition 3: specificity = 94.5% (87.6 to 98.2)			
Number	38.2 (25.4 to 52.3)	6.95	0.65
Superior maximum/nasal median	29.1 (17.6 to 42.9)	5.29	0.75
Ellipse modulation	30.9 (19.2 to 44.8)	5.63	0.73
Inferior maximum/nasal median	9.1 (3.1 to 20.0)	1.65	0.96

TABLE 29 *Diagnostic features of the GDx number (95% CIs in parentheses)*

	Visual field definition 1	Visual field definition 2	Visual field definition 3
Sensitivity for number ≥ 31	64.2 (54.5 to 73.2)	56.5 (44.0 to 68.4)	60.0 (45.9 to 73.0)
Specificity for number ≥ 31	86.2 (77.5 to 92.4)	88.9 (79.9 to 94.8)	85.7 (76.8 to 92.2)
Sensitivity for number ≥ 71	11.0 (5.8 to 18.4)	10.1 (4.2 to 19.8)	12.7 (5.3 to 24.5)
Specificity for number ≥ 71	100.0 (96.1 to 100.0)	100.0 (95.5 to 100.0)	98.9 (94.0 to 99.8)

definition 1 took the fifth centile of mean pattern deviation from the normal group and used this threshold as the upper limit of normality. Global indices, such as MPD, have the limitation that they take an average of data over all locations that have been tested. Consequently, the impact of small defects may be 'diluted' by considering a global measure. However, the study chose to utilise pattern (rather than total) deviation since, although the measurement gives one value for the whole visual field, the parameter expresses localised, rather than diffuse, loss. In real terms,

TABLE 30 *MRA analysis of HRT data following stratification by level of visual field damage (95% CIs in parentheses)*

MRA result	Mean pattern deviation (dB)	No. of subjects	Sensitivity of MRA1 (%)	Sensitivity of MRA2 (%)
Normal	–1.3 to –2.00	29	35 (20 to 53)	69 (51 to 83)
Glaucoma			66 (47 to 80)	31 (17 to 49)
Normal	–2.01 to –3.00	17	27 (13 to 53)	41 (22 to 64)
Glaucoma			71 (47 to 87)	59 (36 to 78)
Normal	–3.01 to –6.00	39	21 (11 to 36)	33 (21 to 49)
Glaucoma			80 (65 to 89)	67 (51 to 79)
Normal	<–6.01	24	17 (7 to 36)	21 (9 to 41)
Glaucoma			83 (64 to 93)	79 (60 to 91)

TABLE 31 *Analysis of GDx number following stratification by level of visual field damage (95% CIs in parentheses)*

Mean pattern deviation (dB)	No. of subjects	Sensitivity of GDx number ≥ 42(%)	Sensitivity of GDx number ≥ 31 (%)	Sensitivity of GDx number ≥ 71 (%)
–1.3 to –2.00	29	31 (17 to 49)	52 (34 to 69)	3 (1 to 17)
–2.01 to –3.00	17	35 (17 to 59)	65 (41 to 83)	6 (1 to 27)
–3.01 to –6.00	39	41 (27 to 57)	67 (51 to 79)	13 (6 to 27)
<–6.01	24	63 (43 to 79)	75 (55 to 88)	21 (9 to 40)

pattern deviation 'lifts off' any diffuse perimetric loss, for example due to cataract. This strategy does, however, incur the possibility of neglecting diffuse perimetric loss, which can occur in early glaucoma.[193] Definitions 2 and 3 aimed to address this issue by examining sectors of the visual field. This method is well established for determining localised loss and has been shown to perform well in the detection of glaucomatous visual field defects[60] and is the basis of the HFA's GHT [see the section 'Analysis of visual field data' (p. 20)], which is widely considered to be a useful tool for glaucoma diagnosis. There was little overall pattern of agreement between the detection of ONH and RNFL damage and any of the visual field definitions (although definitions 1 and 3 tended to perform slightly better than 2) – this is likely to be a manifestation of the wide range of patterns of damage in glaucoma and the complexity of the relationship between morphological change and alteration in visual function.

For determining diagnostic precision, two different methods were used: the sensitivity at 95% specificity was calculated for the optimum parameters – this level was used since it is thought to be appropriate for a condition such as glaucoma.[178] In addition, the areas underneath the ROC curve were computed for all HRT and GDx parameters – this technique has the advantage of being a criterion-free method of establishing overall diagnostic accuracy.

The majority of global HRT stereometric parameters had areas underneath the ROC curve that were interpreted as either **good** or **fair**, which was not the case for the GDx, where a few parameters fell into the **good** or **fair** categories, but most were categorised as **very poor**; many measurements yielded a result just over 0.5, which represents little better than pure chance.

With the HRT, the single best discriminator (of the parameters) was the Bathija DFA.[88] The superiority of combinations of stereometric parameters over single values has been documented previously.[194,195] The Bathija algorithm combines the parameters cup shape measure, height variation contour, rim area and RNFL thickness; of interest, these parameters also fare comparatively well as individual measurements. The maximum area under the ROC curve was 0.865 for the Bathija DFA, which means that a randomly selected glaucoma patient will exceed the normal value of the measurement 87% of the time. This result is very similar to that found by Zangwill and colleagues; for the Bathija DFA, they found an area under the ROC curve of 0.85.[36]

For the Mikelberg DFA, previous research has found sensitivity and specificity estimates of 74% and 88%[196], 87% and 84%[87] and 42% and 90%,[36] compared with 56% and 95% in this study. As already highlighted, making comparisons between studies is confounded by multiple factors which could easily account for the apparent disparities in diagnostic precision.

Reasonable diagnostic precision was also found with the Mikelberg DFA[87] with a maximum area under the ROC curve of 0.826. This formula combines cup shape measure, rim volume and height variation contour. Again, these measurements also perform well singly. The sensitivity of these indices at 95% specificity was 57% and 56% for Bathija and Mikelberg, respectively. It must be noted that these levels of diagnostic accuracy do not represent those produced by the HRT software; the reason for this is that the HRT uses a cut-off of zero of the DFA statistic. When the latter level was used, sensitivity and specificity of 53% and 96%, respectively, were achieved for the Bathija DFA; corresponding values were 71% and 82%, for the Mikelberg DFA. A potential source of error in our specificity estimates at non-zero thresholds is that, ideally, the analysis should be repeated with an independent data set, other than that on which the new cut-off was established. For all three visual field definitions, the Bathija cut-offs were far closer to zero than those for the Mikelberg algorithm, which is consistent with the superior diagnostic performance of the former.

Two definitions of the MRA were used to allow for alternative classifications of the 'borderline' category. MRA_1 treated borderline cases as abnormal, whereas MRA_2 categorised them as normal. Optimum diagnostic ability was found for MRA_1 with visual field definition 1 (sensitivity of 75% for 75% specificity); these estimates were 59% and 95%, respectively, for MRA_2. The highest specificity (96%) was found for MRA_1 with visual field definition 3 at 58% sensitivity. On account of the categorical output of the MRA, specificity could not be pegged at 95%.

A further way of describing diagnostic precision is in terms of positive and negative predictive values (positive and negative predictive values were calculated assuming a prevalence of 1.5%).[197] For MRA_1, these levels were 4.4% and 99.4%, respectively, for MRA_2, 16.8% and 99.3% and 15.4% and 99.3% for the Bathija DFA (using a cut-off of zero).

The MRA has the theoretical advantage for diagnosing glaucoma in that it takes into account optic disc size, which is especially important in the light of the close correlation between disc, neuroretinal rim and cup sizes,[198,199] and is of particular significance in small discs.[200] This factor is likely to explain the better discriminating ability (albeit slight) that the MRA has over the other HRT algorithms, which do not take account of overall disc size.

MRA data were also analysed for the group considered 'at risk' of developing glaucoma (these patients were originally included in the cross-sectional group, but had no evidence of visual field damage by each of the definitions). In all three cases, the MRA found a higher incidence of abnormality at the ONH in the 'at risk' group than in the control group. This finding is likely to be due to patients who have undergone morphological change at the ONH prior to the onset of functional damage (so-called pre-perimetric glaucoma).[33,93,201,202] For definitions 1 and 3, the levels of abnormality from the 'at risk' category were lower than for those with perimetric loss, although the relationship was reversed for definition 2, which is possibly a manifestation of the complex nature of the relationship between structure and function in glaucoma.

The issue of pre-perimetric changes to the ONH and RNFL may influence the diagnostic precision

of the HRT and GDx; if a significant number of patients developed structural change (prior to functional loss), it would reduce the specificity of the technique. In the situation where the specificity is pegged at 95%, it would then reduce the sensitivity.

Miglior and colleagues performed a large-scale study that evaluated the MRA and Mikelberg DFAs.[97] They classified 'borderline' results as 'normal' (i.e. equivalent to MRA_2 from this study) and found a sensitivity of 74% for 85% specificity, compared with 58% and 96% in the present study. However, although Miglior and colleagues used the commendable strategy of not defining glaucoma on the basis of retinal characteristics, they defined abnormality as the presence of a glaucomatous visual field defect plus a history of IOP >21 mmHg. This strategy will have excluded from their experimental group the proportion of glaucoma patients (between 25 and 50%) who have IOPs within the normal range at diagnosis,[7,17,27] so their findings are not necessarily applicable to a typical population of glaucoma patients where significant numbers have normal tension glaucoma.

Diagnostic precision with the MRA from this study is lower than that determined by the creators of the analysis, Wollstein and colleagues.[95] They found sensitivity and specificity of 78% and 81%, respectively. Our findings are in agreement with those found in another study that yielded values of 58% and 96%[96] (identical with our estimates from MRA_2 with visual field definition 3). The finding of lower levels of diagnostic precision from independent studies is not surprising, since an algorithm will perform optimally when tested on the original population from which it was derived. This explanation could also account for differences in performance of the two DFAs compared with the original research.[87,88] This study's levels of diagnostic accuracy were more in line with those found by other independent research.[96]

A potential limitation of the approach taken by this study is that we only considered global HRT parameters and did not examine results from sectoral values. However, it is highly likely that this will have no consequence on the final outcome. The basis for this assumption is that one of the definitive papers in this area did review both global and sectoral parameters and found that, when areas under the ROC curve were put in descending order, only one sectoral parameter was superior to the Bathija linear discriminant function (LDF), and this was by a margin of 0.01.[36] Therefore, sectoral analysis is unlikely to have yielded superior diagnostic accuracy to its global counterpart. A further weakness of the study is that the control subjects from this study were used to derive the normative database from which the visual field probability values were calculated. An optimal strategy would have been to use a separate data set, since characteristics of the normal database can have a profound effect on diagnostic precision.[156]

Overall, the performance of the GDx did not match that of the HRT, although, for both instruments, the highest area under the ROC curve for the GDx was for the number (0.840), with a sensitivity of 45% at 95% specificity. Consequently, the positive and negative predictive values were 12.3% and 99.1%, respectively. The other GDx parameters that performed well were superior maximum/nasal median, maximum modulation and ellipse modulation. The reason why these parameters achieved reasonable diagnostic precision is that (with the exception of the number) they do not express straightforward thickness measurements. This will render them less likely to be affected by anomalous values of corneal polarisation [see the section 'Corneal polarisation' (p. 13)]. The number is also thought to be relatively insensitive to anterior segment birefringence, which may contribute to its favourable performance. Individual compensation has been shown to improve the discriminating ability of the GDx.[125,203] However, such modifications could not be incorporated in this study as they are only available with the GDx VCC, which is the successor to the model used in our laboratory. Since the developments incorporated in the new instrument are both hardware and software based, our instrument cannot be upgraded to allow us to determine the effect of customised corneal compensation on our data set. Therefore, the distribution of diagnostic abilities of the GDx parameters found from this study may well differ from the GDx VCC.

Four new GDx parameters have recently been described.[104] However, they are only available via a special version of the software and are not computed by the default version on our instrument. Of interest, in the study by Colen and colleagues[104] each of the new parameters did not perform as well as the number, so the results of our study still represent the optimal GDx parameters. From their research, they found the number to have an area under the ROC curve of 0.90 (compared with 0.84 from our results).

However, the former figure was derived from the entire experimental group who had an average mean deviation of –10.4 dB, that is, far more severe damage than in this study, which would account for the higher diagnostic precision. When they stratified their experimental group by the degree of damage, the subset they classified as having 'early' loss (mean MD = –3.5 dB) had an area under the ROC curve of 0.84 for the number (identical with our estimate).

Similar performance of the GDx number was found in the study of Zangwill and colleagues, who found it to have an area under the ROC curve of 0.81.[36] They also examined an LDF[204] (not currently available with the basic GDx software), which had previously been found to have good discriminating ability. However, this function was not included in our research, since this study did not include parameters that require supplementary analysis for their derivation.

The strategy of subclassifying the experimental group according to the level of damage highlights the issue of how critical the degree of perimetric loss is to sensitivity estimates. This study used very narrow increments to subdivide the glaucoma population, but still yielded very different sensitivity estimates for each category. For MRA_1, the sensitivity for patients with mean pattern deviation of –1.3 to –2.00 dB was 66%, whereas for those with loss in the –3.01 to –6.00 dB range it was 80% (the specificity for both of these categories was 75%). A similar pattern was found for the GDx number. Although this form of analysis has the advantage of allowing enhanced exploration of the experimental group, the adverse consequence of examining small sample sizes is that it significantly increases the 95% CIs of the estimates.

Overall, this study has found that, although the HRT and GDx do appear to have reasonable diagnostic precision, they still fail to detect a significant number of cases with glaucomatous visual field loss. This finding is likely to be a manifestation of the complex relationship between structure and function in glaucoma. Previous work has shown that significant damage can occur at the ONH and RNFL prior to the onset of perimetric loss.[201,205] However, more recent research suggests that this is not necessarily always the case,[35] where some patients may show evidence of functional damage prior to structural change. Greater understanding of the association between structure and function is necessary before the shortcomings of the HRT and GDx in glaucoma diagnosis will be fully apparent.

Chapter 8

Longitudinal high risk study

This chapter reports the results from the longitudinal high risk category of patients.

Introduction

In the literature, there is a paucity of studies that describe the performance of imaging techniques in longitudinal follow-up of glaucoma suspects; of those that exist, most only include a relatively short period of follow-up.

The fundamental issue to be addressed in this section of the study concerns which of the two methods of assessment [topographic analysis of the ONH (HRT) or polarimetric examination of the RNFL (GDx)] is best able to identify patients developing glaucomatous optic neuropathy (known as conversion).

As already described in Chapter 7, the issue of definitions is pivotal in determining diagnostic precision. The problematic issue with glaucoma is that there is no gold standard definition and alterations in the threshold criteria can profoundly affect the outcome.

Consistent with the theme of this report, a perimetric definition of glaucoma is used. This strategy is taken in order to prevent bias of the results towards the imaging devices.[94]

Method

Patients from the longitudinal high risk category were examined according to the procedures detailed in the section 'Data collection' (p. 18). Two different types of analysis were performed on visual field data to identify patients who showed perimetric conversion: the rate of change in pattern deviation values derived from linear regression analysis (described in the following subsection) and comparison of results with data collected from the control population (see the subsequent subsection).

Rate of change of pattern deviation

Longitudinal visual field data were analysed by three different methods:

1. Global
Linear regression analysis was performed on longitudinal global MPD values. The number of patients whose rate of change exceeded the p levels of ≤ 0.05, ≤ 0.02, ≤ 0.01 and ≤ 0.005 was calculated. In order to estimate the amount of experimental noise, the number of patients whose rate of change exceeded the p levels of ≥ 0.95, ≥ 0.98, ≥ 0.99 and ≥ 0.995 was also calculated.

2. Sectoral
Linear regression analysis was performed on sectoral MPD, for 10 visual field sectors, using the GHT locations [see the section 'Analysis of visual field data' (p. 20)]. The computation was repeated for the same levels of probability as described above.

3. Pointwise
Pointwise linear regression analysis was performed on the 54 points in each visual field. The computation was repeated for the same levels of probability as described above.

The following definitions were used to identify converters from each category:

1. Global
Patients who showed a rate of change with a $p \leq 0.05$ on linear regression.

2. Sectoral
Patients who had ≥ 1 sector with a rate of change with a probability of $p \leq 0.05$.

3. Pointwise
Patients who had ≥ 3 points with rates of change with $p \leq 0.02$.

A Venn diagram was constructed to describe the extent of agreement between the three methods described above.

Patients who showed conversion were subsequently subclassified:

- **Definite converters:** patients who fulfilled all three criteria for conversion.
- **Probable converters:** patients who fulfilled any two criteria for conversion.

- **Suspect converters:** patients who fulfilled only one criterion for conversion.
- **Stable fields:** patients who showed no significant change in any of the three criteria.

Each of the above categories is mutually exclusive; for example, the classification of probable converters does not also include those who showed definite conversion.

The category of patients with 'stable fields' was included for those cases that showed evidence of conversion on ONH and/or RNFL characteristics, but not on the basis of their perimetric data.

From the cross-sectional study (Chapter 7), the four parameters with the largest areas underneath the ROC curves were selected for analysis. For the HRT, these were C:D area ratio, mean RNFL thickness, vertical C:D ratio and the Bathija DFA, and for the GDx they were, the number, ellipse modulation, maximum modulation and superior maximum:nasal median. Linear regression analysis was performed on each parameter.

For each perimetric conversion category (see above), the number of patients showing significant change (at the 5% level) in their HRT and GDx parameters was calculated; this analysis was repeated for both 'improvement' and 'deterioration'. The $p \geq 0.95$ or $p \leq 0.05$ threshold was chosen for each parameter depending on whether the parameter value increased or decreased with abnormality, respectively. Venn diagrams were constructed to show the extent of overlap between the detection of conversion between the three techniques; the analysis was repeated, taking into account the number of parameters showing evidence of change.

Pattern deviation outside normal limits

Global MPD at baseline and the final examination was compared with the fifth centile from the normal control population ($n = 98$) described in Chapter 7. The lower limit of normality was –1.3 dB [see the section Results (p. 49)]: global MPDs lying above and below this threshold were classified as normal and abnormal, respectively. The number of patients who showed an alteration in classification between their initial and final examinations was calculated.

For patients who showed evidence of conversion, the number of HRT and GDx parameters that showed significant change (at the 5% or 95% level, as appropriate) was documented and agreement between the three techniques ascertained.

For this section of the analysis, the HRT parameters global and vertical C:D ratio were not used; instead, each parameter was divided by disc diameter, which was calculated as 2 (disc area/π)^{½}. This strategy was adopted to account for the well-established relationship between neuro-retinal rim area and disc size.[198,199,206–208]

Results

Patient demographics, together with duration of follow-up, the number of examinations and data required by standards established by the QUOROM[152] are given in the section 'Longitudinal high risk and longitudinal glaucoma groups' (p. 25); additional details are given in *Table 32*.

Figure 21 shows the frequency distribution of global MPD at baseline and completion of the study. To quantify the overall amount of perimetric change that was observed during the period of the study, the frequency distribution of the gradients from linear regression of global MPD is shown in *Figure 22*. *Figure 23* shows the relationship between the gradient of the regression line and the associated probability value.

TABLE 32 *Details of longitudinal high risk population*

Characteristic	Longitudinal high risk patients
Number of subjects	240
Minimum age at baseline (years)	38.2
Maximum age at baseline (years)	86.5
Median age (years)	63.9
Right eyes (%)	51.3
Left eyes (%)	48.7
Male (%)	60.8
Female (%)	39.2

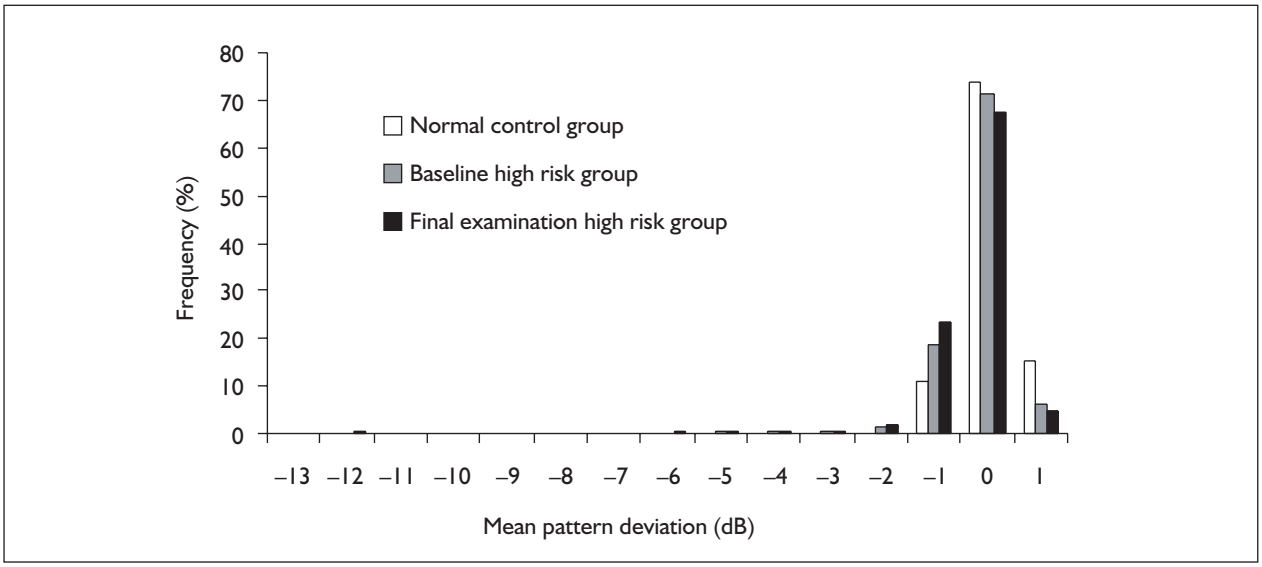


FIGURE 21 *Frequency distribution of global MPD at baseline and final examination. Data from the normal control population (Chapter 7) have been incorporated for comparison.*

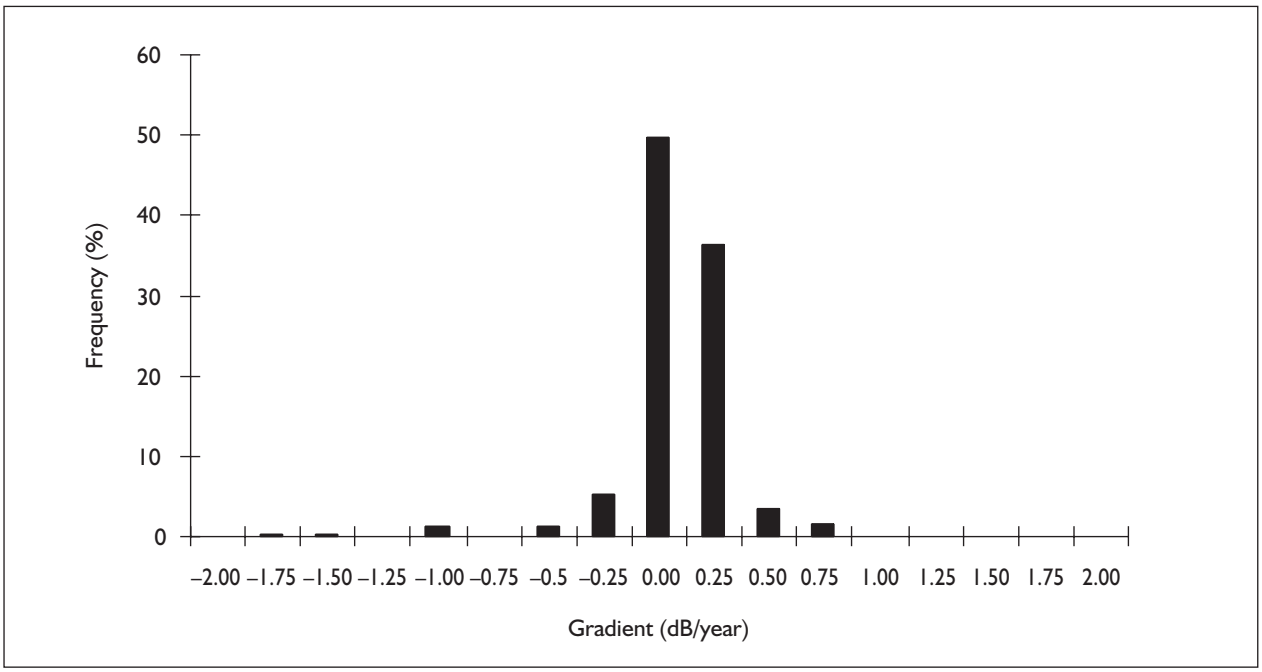


FIGURE 22 *Frequency distribution of gradient of global MPD*

Table 33 illustrates the outcome of linear regression analysis of visual field data; the results from global, sectoral and pointwise analysis are given. In addition to cases showing significant deterioration (i.e. with negative gradient of regression line and probability values ≤ 0.05, ≤ 0.02, ≤ 0.01 and ≤ 0.005), the proportion of patients who showed apparent 'improvement' at the corresponding levels of significance are also given.

In total, 72 patients showed some evidence of perimetric conversion by one, or more, of the three definitions. The number of patients who converted according to the global, sectoral and pointwise criteria were 26, 30 and 45, respectively. Clearly, the criteria agreed on some cases, but disagreed on others. *Figure 24* shows the frequency distribution of high risk patients with respect to the number of perimetric criteria in which they showed perimetric conversion. The extent of

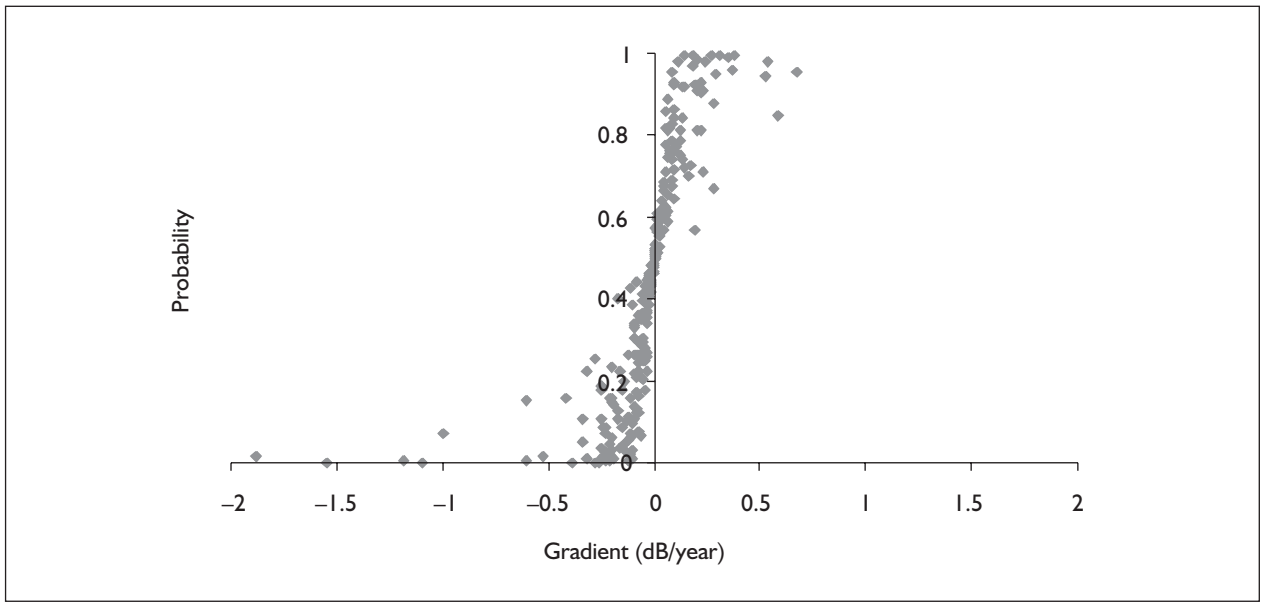


FIGURE 23 *Results from linear regression of global MPD*

TABLE 33 *Outcome of linear regression analysis of visual field data [data shown as a percentage of total group (n = 240)]*

		Sectoral (%)				Pointwise (%)			
***p*-Value**	**Global (%)**	**≥ 1 sector**	**≥ 2 sectors**	**≥ 3 sectors**	**≥ 4 sectors**	**≥ 1 point**	**≥ 2 points**	**≥ 3 points**	**≥ 4 points**
Negative gradient									
≤ 0.05	10.8	49.2	22.9	6.3	3.3	93.8	78.3	59.6	43.3
≤ 0.02	7.5	31.3	7.1	2.1	1.3	72.5	41.7	18.8	8.3
≤ 0.01	5.4	19.2	3.3	0.8	0.4	47.5	20.0	7.1	2.5
≤ 0.005	2.5	12.5	1.3	0.4	0.0	29.2	9.6	2.9	1.7
Positive gradient									
≥ 0.95	5.8	40.4	11.3	4.2	1.7	92.1	75.4	48.3	26.7
≥ 0.98	3.3	20.8	3.8	0.4	0.0	61.3	29.2	13.8	4.6
≥ 0.99	2.5	12.9	1.7	0.0	0.0	39.2	13.3	3.3	1.7
≥ 0.995	0.8	8.8	1.3	0.0	0.0	26.3	4.2	1.3	0.0

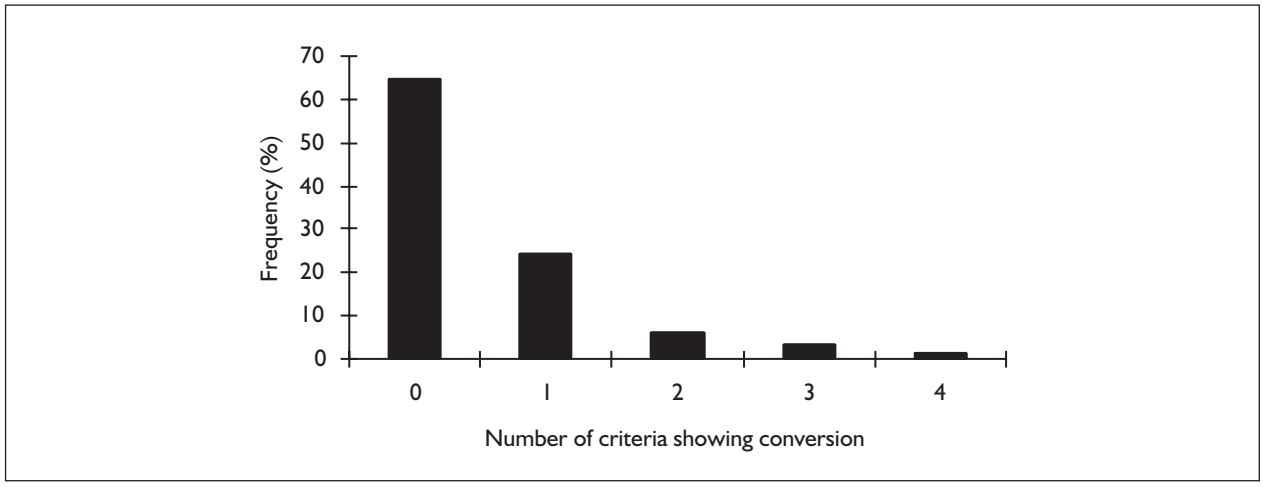


FIGURE 24 *Number of perimetric criteria showing conversion*

agreement between the three methods derived from linear regression is shown in *Figure 25*.

From the entire high risk group, the number of patients showing significant deterioration and improvement of their HRT and GDx parameters is shown in *Table 34*.

The extent of agreement between the HFA, HRT and GDx for glaucomatous conversion when global, sectoral and pointwise definitions of perimetric conversion were used is shown in *Figure 26*. In this case, significant deterioration with the HRT and GDx was defined as any patient who had linear regression that was significant at the ≤ 5% level (or ≥ 95% level, as appropriate) for one or more of the four parameters considered.

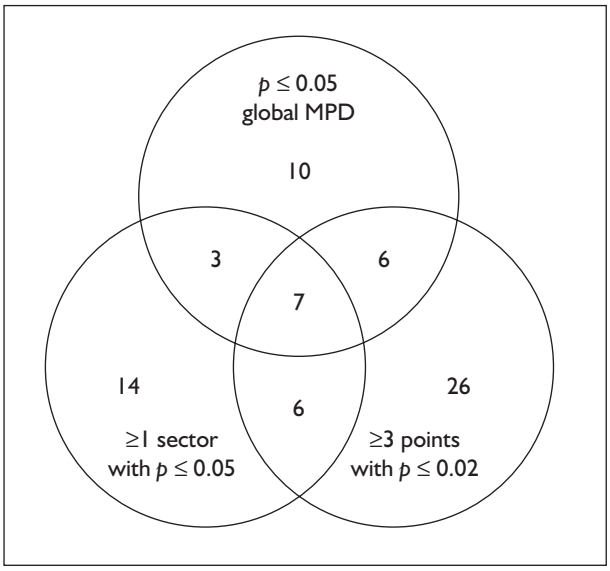


FIGURE 25 *Venn diagram showing agreement between different criteria for detecting perimetric conversion (total number = 240)*

From the total of 240 high risk eyes, the numbers showing definite conversion, probable conversion, suspect conversion and stable visual fields were 7, 15, 50 and 168, respectively.

Definite converters

Seven patients were defined as showing 'definite perimetric conversion', that is, they fulfilled all three criteria for conversion derived from linear regression of their perimetric data. From this group, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 35*. The agreement between the visual field analysis, HRT and GDx is shown in *Figure 27*.

Probable converters

Fifteen patients were defined as showing 'probable conversion', that is, they fulfilled any two criteria for conversion derived from linear regression of their perimetric data (patients categorised as definite converters were **not** included in this group). From the probable converters, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 36*. The agreement between the visual field analysis, HRT and GDx is shown in *Figure 28*.

Suspect converters

Fifty patients were defined as showing 'suspect conversion', that is, they fulfilled any one criterion for conversion derived from linear regression of their perimetric data (patients described in the previous two sections were **not** included in this group). From this population, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 37*. The agreement between visual field analysis, HRT and GDx is shown in *Figure 29*.

TABLE 34 *Number of patients with significant deterioration in linear regression of HRT and GDx parameters (maximum = 240)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	34	28	48	26
No. of eyes with significant improvement	18	20	15	14
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	46	35	32	36
No. of eyes with significant improvement	7	11	15	17

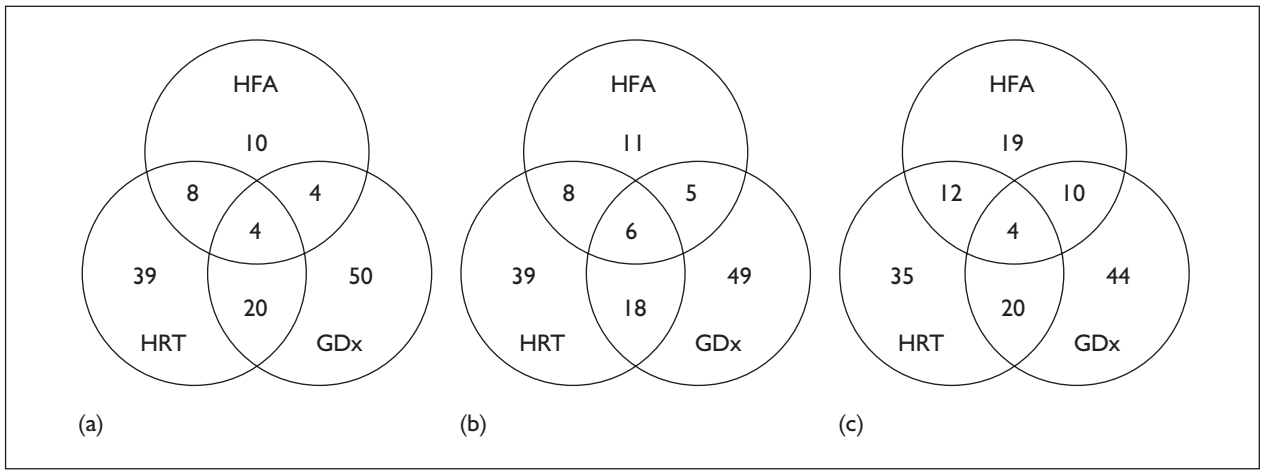


FIGURE 26 *Venn diagrams showing the extent of agreement between change in HRT and GDx parameters for different definitions of glaucomatous perimetric conversion. Data from patients who showed conversion according to (a) the global criterion, (b) the sectoral criterion and (c) the pointwise criterion*

TABLE 35 *Number of definite converters showing significant change in HRT and GDx parameters (maximum = 7)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	4	4	6	4
No. of eyes with significant improvement	2	2	0	2
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	4	3	5	4
No. of eyes with significant improvement	1	1	1	1

Stable visual fields

No significant perimetric change by any of the three methods was found in 168 eyes. From this population, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 38*. The agreement between visual field analysis, HRT and GDx is shown in *Figure 30*.

Comparison with normative data

From Chapter 7, the 95th centile of the global mean pattern deviation from the normal group was –1.3 dB. Twenty-three high risk eyes changed from above to below this threshold during their period of follow-up. Fourteen subjects showed an apparent 'improvement', that is, they went from below to above the threshold during the study.

From the 23 eyes that showed conversion according to this criterion, the numbers of patients whose HRT and GDx parameters lay outside the fifth or 95th centile of normality (as appropriate) are shown in *Table 39*. These figures give the sensitivity of the technique to detect early glaucoma at 95% specificity. For example, the sensitivity of the GDx number is 26%.

Table 40 shows the number of patients who showed significant change in their HRT and GDx parameters. *Figure 31* shows the agreement between visual field analysis, HRT and GDx.

Discussion

This study has the methodological advantage of comparing two imaging techniques (HRT and GDx) in one population of 'high risk' patients. In the literature, there is a definite lack of large-scale longitudinal research on the use of the HRT and

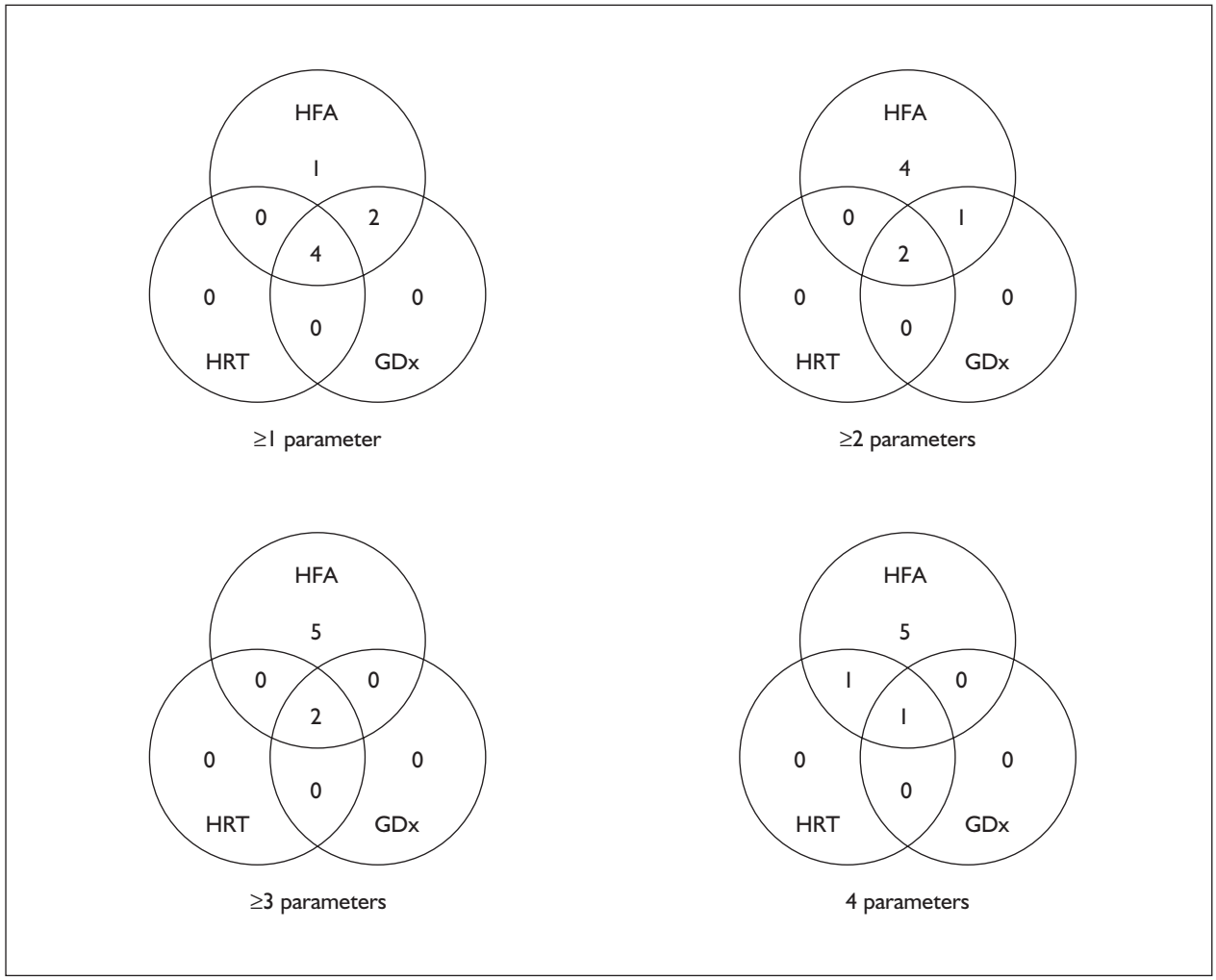


FIGURE 27 *Agreement between visual field analysis, HRT and GDx in patients showing definite conversion (maximum = 7)*

TABLE 36 *Number of probable converters showing significant change in HRT and GDx parameters (maximum = 15)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	5	5	5	2
No. of eyes with significant improvement	2	2	5	3
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	6	3	2	4
No. of eyes with significant improvement	1	1	0	3

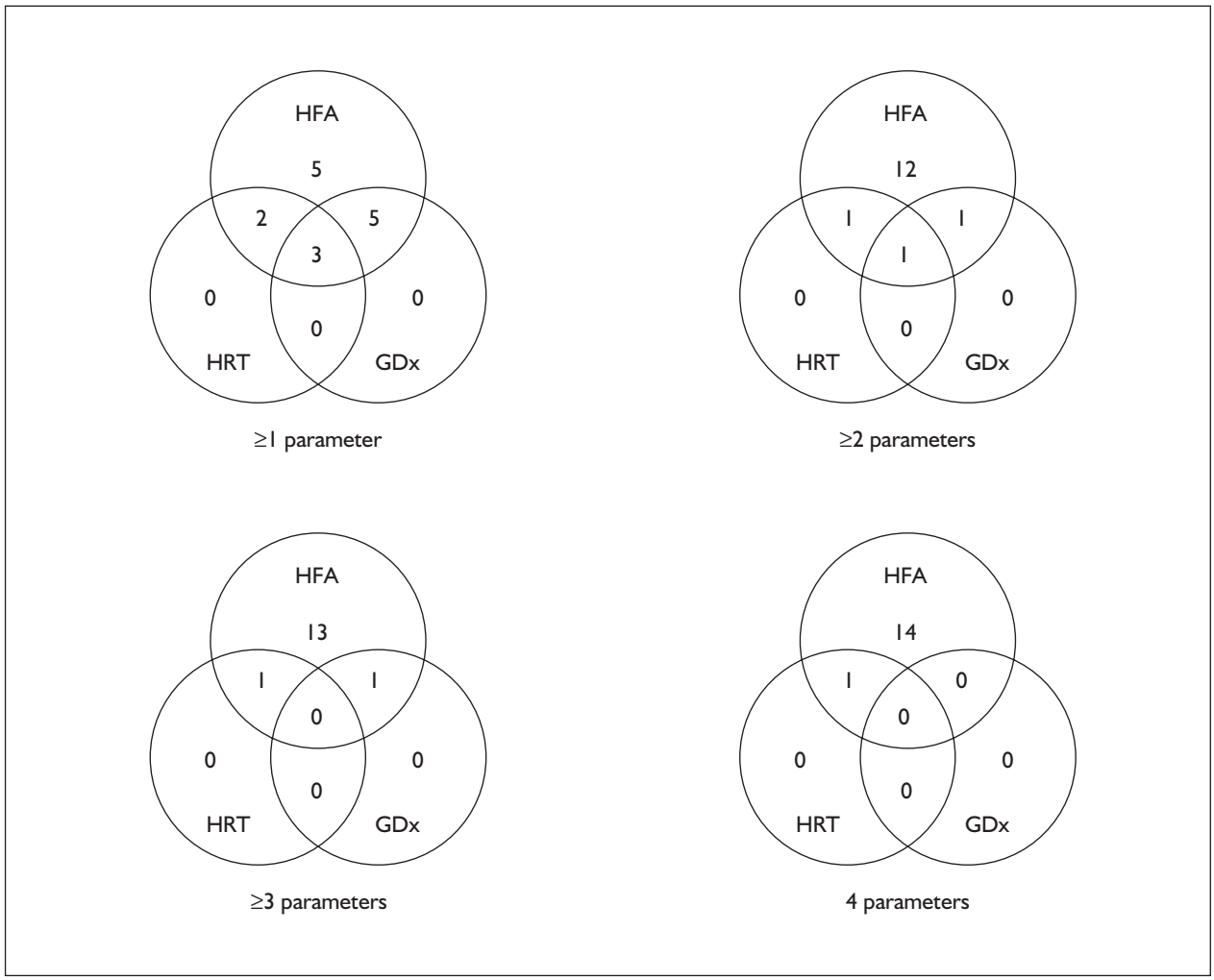


FIGURE 28 *Agreement between visual field analysis, HRT and GDx in patients showing probable conversion (maximum = 15)*

TABLE 37 *Number of suspect converters showing significant change in HRT and GDx parameters (maximum = 50)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	10	3	18	6
No. of eyes with significant improvement	3	7	2	5
No eyes with significant deterioration	**GDx parameter**			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	7	4	6	6
No. of eyes with significant improvement	1	4	6	5

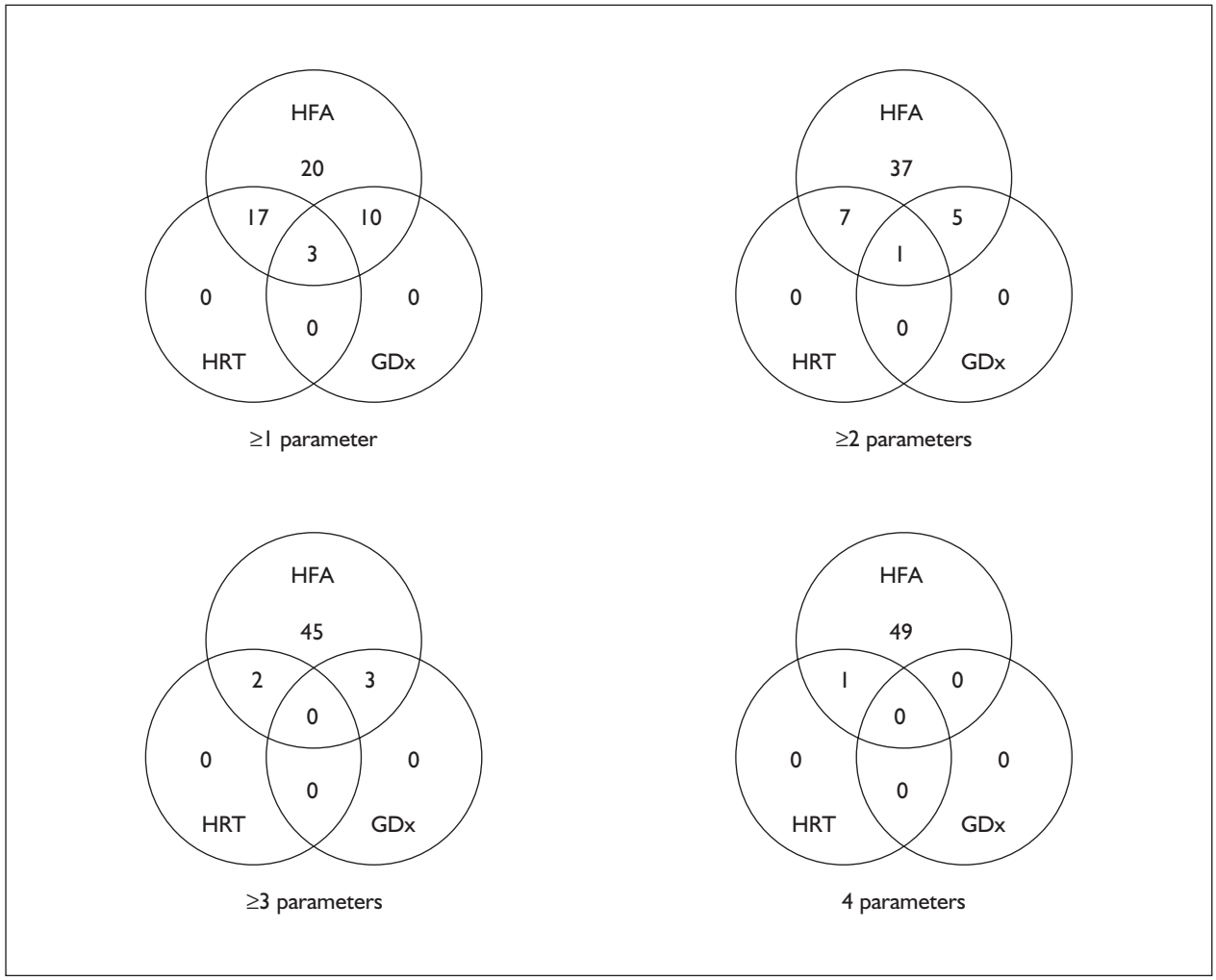


FIGURE 29 *Agreement between visual field analysis, HRT and GDx in patients showing suspect conversion (maximum = 50)*

TABLE 38 *Number of suspect converters showing significant change in HRT and GDx parameters (maximum = 168)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	15	16	19	14
No. of eyes with significant improvement	11	9	6	4
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	29	25	19	22
No. of eyes with significant improvement	4	5	8	8

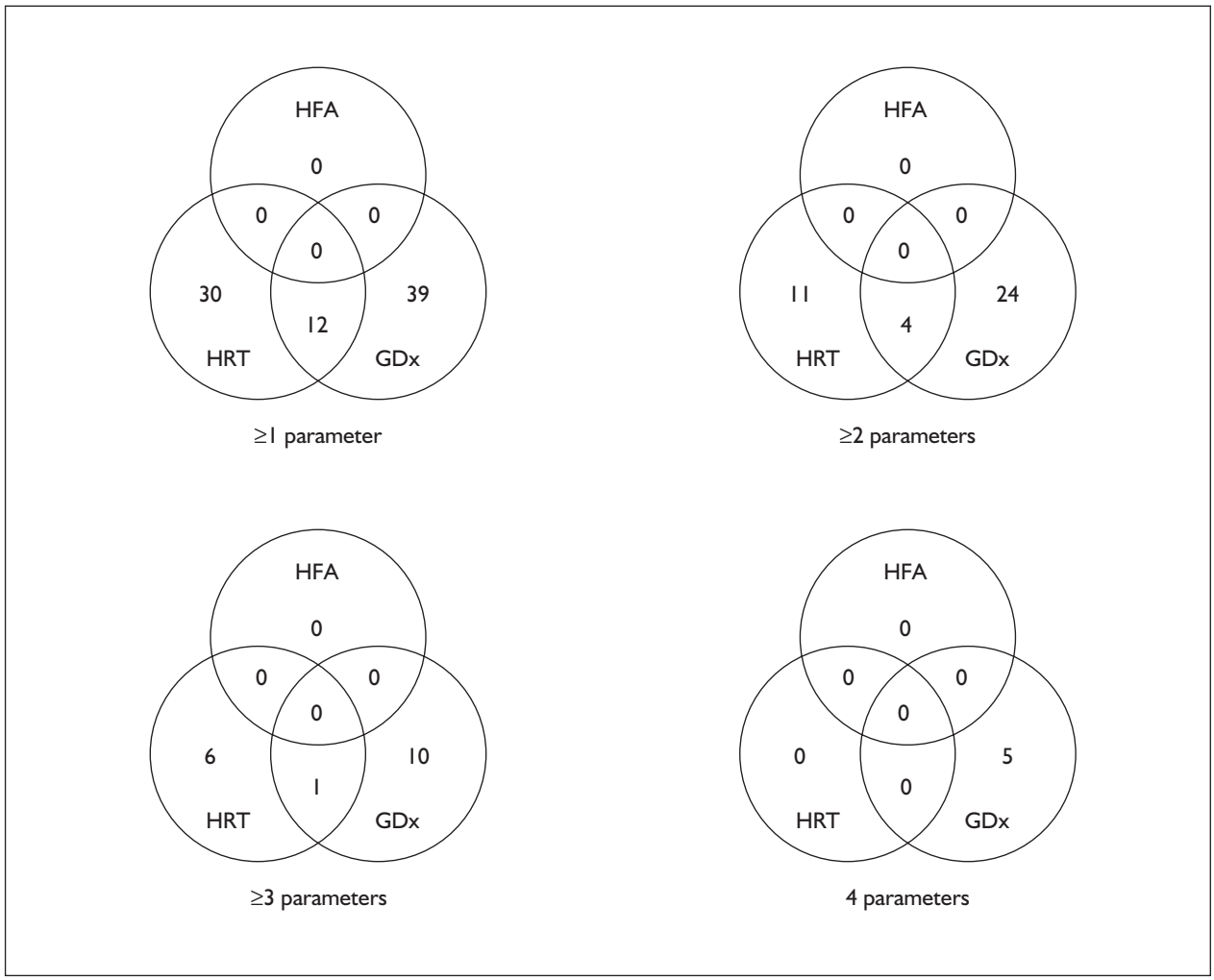


FIGURE 30 *Agreement between visual field analysis, HRT and GDx in patients with stable visual fields (maximum = 168)*

TABLE 39 *Number of converters with abnormal HRT and GDx results (maximum = 23)*

	HRT parameter			
	C:D area ratio/ disc diameter	**Mean RNFL thickness**	**Vertical C:D ratio/disc diameter**	**Bathija DFA**
Outside normal limits at final examination	9	3	6	8
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
Outside normal limits at final examination	6	3	6	7

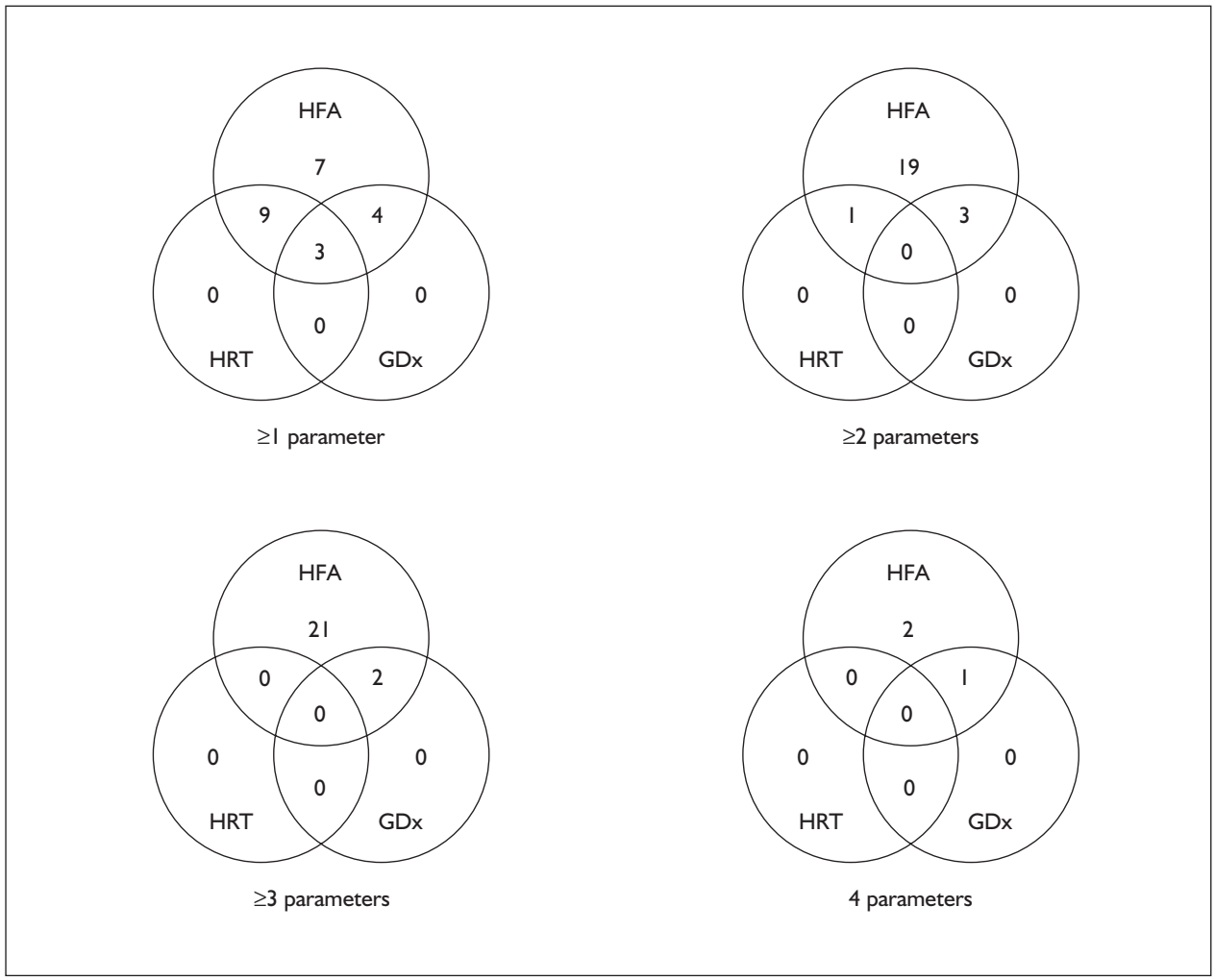


FIGURE 31 *Agreement between visual field analysis, HRT and GDx (maximum = 23)*

TABLE 40 *Number of converters who showed change in HRT and GDx parameters (maximum = 23)*

	HRT parameter			
	C:D area ratio/ disc diameter	**Mean RNFL thickness**	**Vertical C:D ratio/disc diameter**	**Bathija DFA**
No. of eyes with deterioration	6	2	2	3
No. of eyes with improvement	6	1	0	5
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with deterioration	3	3	4	3
No. of eyes with improvement	4	7	3	3

GDx in such patients. This paucity was highlighted at the annual meeting of the ARVO and Ophthalmology in April 2004, where few such longitudinal studies were described.

In this study, the designation of eyes as high risk was conferred either due to ocular hypertension and/or a fellow eye with glaucomatous visual field loss. Consistent with the theme of this report, glaucomatous conversion was defined in terms of visual field changes. In all, 72 patients showed some evidence of the development of a visual field defect (otherwise known as perimetric conversion); however, the level of certainty of this status differed between individuals. The total number of converters exceeded the study's original target of 55.

When the converters were stratified according to the degree of certainty that they had developed a visual field defect, 3, 6 and 21% of the initial high risk population showed definite, probable and suspect conversion, respectively. The different proportions in each of these categories highlights the susceptibility of conversion rates to the definition of perimetric conversion. This vulnerability, along with selection criteria, confounds the comparison of conversion rates from this study with those found by other research, where alternative definitions have been employed. From a study with a similar population size, Kamal and colleagues found a conversion rate of 5.5% for ocular hypertensives reviewed over a period of 5 years.[202] However, the authors used a different tool for identifying conversion.[132] Johnson and colleagues[209] highlighted the susceptibility of conversion rates to the defining criteria: from their large-scale study (479 eyes), they found that between 5.3 and 17.5% of subjects showed evidence of conversion.

The Longitudinal Diagnostics Innovations in Glaucoma Study found a conversion rate of 7.5%.[210] However, this figure was derived using short-wavelength perimetry, the results of which are not necessarily comparable to those derived from this study. Using an alternative definition of conversion with conventional perimetry, the Ocular Hypertension Treatment Study found that 10% of patients converted over a period of 5 years.[211]

A factor influencing conversion rate is the degree of risk of each patient; an important risk factor is IOP.[6,27,211–215] Consequently, a greater number of untreated ocular hypertensive patients within the study population would increase the rate of perimetric deterioration. The majority of the patients within the Glaucoma Imaging Study were receiving treatment for their ocular hypertension. A possible further factor contributing to the proportion of converters is that in recruiting for a longitudinal study, where a significant commitment is required from participants, there is the potential of selection bias: on the whole, patients who volunteer for such research are conscientious, reliable individuals who are compliant with their medication regime and regularly attend for clinic appointments. Also, any patient participating in the study in whom any deterioration was detected (e.g. visual acuity, visual field status, appearance of the ONH, IOP, compliance with treatment) was immediately referred back to their clinician, with the intention of minimising any decline. Consequently, the conversion rate from the study population may not reflect that found in an average glaucoma clinic. Such disparities will form the basis of further research within the University of Manchester over the next 2 years.

For detection of change in visual field data, HRT and GDx parameters, linear regression analysis was used. This technique has the advantage that it will identify a small magnitude of change in patients who show low intraindividual variability, whereas it requires a greater amount of change to achieve significance in cases where measurement noise is larger. Individualised regression was considered an optimum strategy compared with averaging values across the entire population.

In addition to documenting the numbers of patients in whom statistically significant deterioration in visual field, optic disc or RNFL characteristics were detected, the numbers of individuals in whom an apparent improvement occurred were also documented. The latter could be considered a surrogate for the number of FP converters. Alternatively, in a few cases, the apparent improvement in HRT and/or GDx parameters may be attributable to pharmacologically or surgically decreased IOP, where the optic disc may 'rebound'.[216,217] Supplementary investigation of the relationship between intervention and change in HRT and GDx parameters lies outside the scope of this study.

This study took the approach of comparing conversion rates with three different definitions of perimetric conversion: global, sectoral and pointwise. There was a specific rationale to each of these criteria. The global value was used because it

is known that diffuse visual field loss can occur in early glaucoma.[44] The study used the visual field measure 'pattern deviation', which has the benefit of negating any change due to conditions such as cataract (which may well be a factor in a longitudinal study where many of the patients are in their seventh and eighth decades). However, on account of attempting to 'remove' any diffuse change, pattern deviation may negate diffuse glaucomatous loss.

Sectoral visual field loss was investigated using the same sectors as the GHT [see the section 'Analysis of visual field data' (p. 20)]. This form of analysis was employed because it has been acknowledged to be ideal for detecting localised glaucomatous visual field loss,[60] which is characteristic of the condition.

Pointwise analysis was also used in order to detect very small areas of visual field loss, whose impact may be lost when values are averaged over a larger area. In order to minimise the number of FPs, the threshold was defined at ≥ 3 points with probability values of ≤ 0.02. The agreement between patients detected by the different definitions was poor and is a manifestation of the different patterns of visual field loss encountered in glaucoma.[218]

In the group showing definite conversion, analysis of the extent of agreement between the three techniques is adversely affected by the small number of subjects in this category ($n = 7$). When considering patients who showed significant deterioration in one or more parameter, agreement between the three instruments was good. However, there was little overlap when the threshold was increased to four parameters.

The general trend in the other subpopulations is that there was a relatively low degree of commonality between the different tests. Some perimetric converters appeared to be detected by the HRT and others by the GDx, with only a small extent of agreement between the two. In the group of patients with stable visual fields (who could feasibly be undergoing pre-perimetric changes) there is, again, comparatively little agreement between cases identified by the HRT and GDx.

The finding that visual field and ONH analysis detect different groups of patients is consistent with other recent research. Over the last 10 years, scientific opinion dictated that morphological change at the ONH and RNFL preceded visual field damage detected with conventional perimetry.[201,205] However, recently, other researchers have concluded that there are two disparate groups of patients: those who initially develop perimetric damage and others who initially develop optic disc changes.[34,35] The findings of this study coincide with the more contemporary conclusions.

The section of our analysis that compared results from the longitudinal high risk group with data from our own normal population had the methodological advantage that we used control individuals with a similar demographic profile to the patients and also the same equipment had been used to examine all individuals. However, since linear regression analysis was not appropriate in this situation, the strategy of only utilising the initial and final examinations does not take full advantage of the longitudinal data and renders the results more vulnerable to measurement error.

A further limitation of the analysis, which also holds for our sensitivity and specificity estimates (see Chapter 7), is that using a normal population of around 100 individuals with a definition of abnormality at the fifth centile makes the threshold very vulnerable to a small number of anomalous values.

The study was compromised by the fact that longitudinal data had not been collected on the normal control individuals; these data would have been useful for calculating an estimate of measurement variability: changes outside this level could have been defined as conversion.

From the cross-sectional analysis, relatively good discrimination between glaucoma patients and normal individuals was achieved by the MRA. However, data from this algorithm could not be incorporated in this chapter, as the MRA output is categorical, which is not amenable to longitudinal analysis. Consequently, MRA results have not been analysed for the high risk eyes.

The study took the approach of considering the four HRT and GDx parameters that had been identified by the cross-sectional study as having optimum diagnostic accuracy (largest area under the ROC curve). A strength of this strategy is that it concentrated on the parameters that were most likely to detect abnormality from our experimental population. A disadvantage is that it may have neglected other parameters that were capable of providing detection of perimetric conversion. An argument in favour of our approach is that it has

been documented that some of the GDx parameters are of minimal utility and may give clinically misleading information when they are flagged as being abnormal.[121]

This report has already discussed the issue that the version of the GDx that was used during the study was not capable of making individualised adjustments for patients' corneal polarisation. This factor may have had significant consequences for the cross-sectional study, where sensitivity and specificity estimates may have been adversely affected. However, for analysis of longitudinal data, there should be no negative consequence for our results, as temporal fluctuations of corneal polarisation are unlikely.[219]

This study examined the number of eyes that had undergone statistically significant deterioration during the period of follow-up. However, in cases where abnormality was defined by linear regression of pattern deviation values, it has not considered results from the baseline examination, as to whether the HRT and GDx categorised patients as normal or abnormal. There is the potential that, at the outset of the study, a patient had normal visual fields, but clinically detectable optic disc and/or RNFL damage (i.e. a temporal shift between structural and functional change). If the patient showed no further alteration in the last two variables during the study, they would not have been identified as having a detectable abnormality. This issue was addressed for patients whose data were analysed with respect to the normal control population. From the former group, at 95% specificity, sensitivity estimates of the HRT and GDx parameters ranged from 13 to 39%. However, these levels of diagnostic accuracy do not necessarily relate to patients who fulfilled the other criteria for conversion.

Overall, analysis of the high risk eyes has shown that there is poor agreement for identification of converters between visual fields, ONH and RNFL examination. Further work investigating the relationship between structure and function in glaucoma is necessary to understand the relationship between these three variables. From the data presented in this chapter, the conclusion is that examinations with the HRT and GDx are a useful adjunct to visual field assessment. However, on account of the fact that the HRT and GDx fail to detect a significant number of cases of conversion, they cannot provide a replacement for visual field examination.

Chapter 9

Longitudinal glaucoma study

This chapter reports the results from the longitudinal glaucoma category of patients.

Introduction

As with the high risk category, there are very few published data concerning the use of the HRT and GDx in longitudinal examination of glaucoma patients.

The aim of this section of the study was to ascertain whether the HRT or GDx can detect deterioration in patients who already have glaucomatous visual field defects. Patients showing such change have been termed 'progressors'; in line with the rest of this report, a perimetric definition of progression has been used.

This issue of definitions is, yet again, pivotal in describing progression. As already highlighted in Chapters 7 and 8, where it has been identified that there are no well-established criteria for defining glaucoma and glaucomatous conversion, respectively, there is also no widely accepted gold-standard definition for determining the occurrence of glaucomatous progression. Therefore, a variety of perimetric criteria were used.

Method

Patients from the longitudinal glaucoma group were examined according to the criteria described in the section 'Data collection' (p. 18).

Longitudinal visual field data were analysed by three different methods:

1. Global
Linear regression analysis was performed on longitudinal global MPD values. The number of patients whose rate of change exceeded p levels of ≤ 0.05, ≤ 0.02, ≤ 0.01 and ≤ 0.005 was calculated. In order to estimate the amount of experimental noise, the number of patients whose rate of change exceeded p levels of ≥ 0.95, ≥ 0.98, ≥ 0.99 and ≥ 0.995 was also calculated.

2. Sectoral
Linear regression analysis was performed on sectoral MPD values, using the GHT sectors [see the section 'Analysis of visual field data' (p. 20)]. The computation was repeated for the same levels of probability as described above.

3. Pointwise
Pointwise linear regression analysis was performed on the 54 points in each visual field. The computation was repeated for the same levels of probability as described above.

The following definitions were used to identify progressors from each category:

1. Global
Patients who showed a rate of change with a probability value of ≤ 0.05 on linear regression.

2. Sectoral
Patients who had ≥ 1 sector with a rate of change with $p \leq 0.05$

3. Pointwise
Patients who had ≥ 3 points with rates of change with $p \leq 0.02$.

A Venn diagram was constructed to describe the extent of agreement between the three methods described above.

Patients who showed progression were subsequently subclassified:

- **Definite progressors:** patients who fulfilled all three criteria for progression.
- **Probable progressors:** patients who fulfilled any two criteria for progression.
- **Suspect progressors:** patients who fulfilled only one criterion for progression.
- **Stable fields:** patients who showed no significant change in any of the three criteria.

Each of the above categories is mutually exclusive; for example, the classification of probable progressors does not also include those who showed definite progression.

TABLE 41 *Details of longitudinal glaucoma patients*

Characteristic	Longitudinal glaucoma patients
Number of subjects	75
Minimum age at baseline (years)	38.2
Maximum age at baseline (years)	84.1
Median age (years)	67.8
Right eyes (%)	51.4
Left eyes (%)	48.6
Male (%)	58.1
Female (%)	41.9

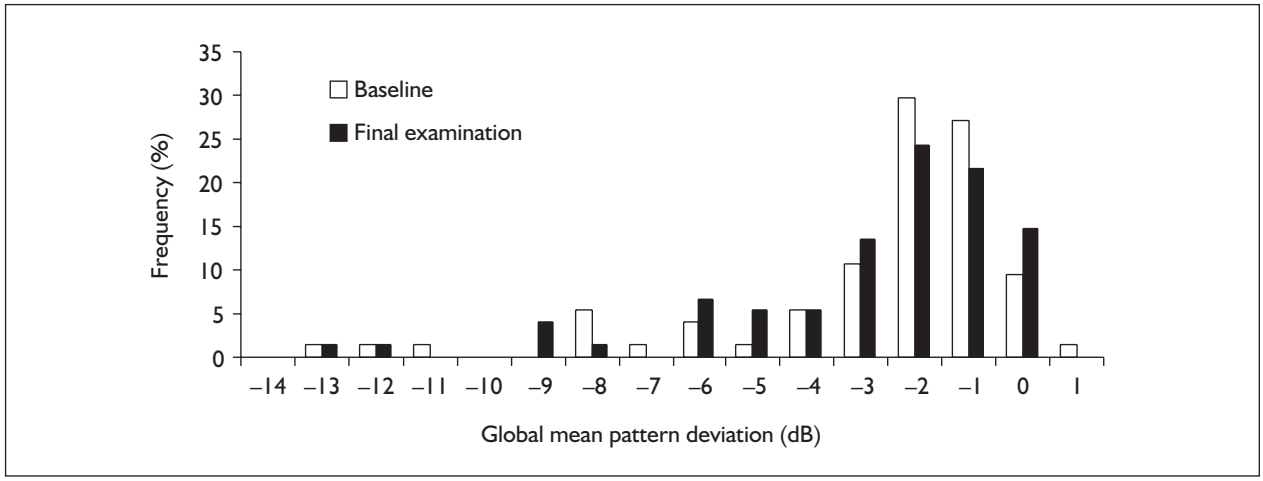


FIGURE 32 *Frequency distribution of global MPD at baseline and final examination*

The category of stable visual fields was incorporated for those cases that showed evidence of structural (ONH and/or RNFL) progression but no evidence of deterioration according to their perimetric data.

From the cross-sectional study (Chapter 7), the four parameters with the largest areas underneath the ROC curves were selected for analysis. For the HRT these were C:D area ratio, mean RNFL thickness, vertical C:D ratio and the Bathija DFA, and for the GDx they were the number, ellipse modulation, maximum modulation and superior maximum:nasal median. Linear regression analysis was performed on each parameter.

For each perimetric progression category (see above), the number of patients showing significant change (at the 5% level) in their HRT and GDx parameters was calculated; this analysis was repeated for both 'improvement' and 'deterioration'. The $p \geq 0.95$ or $p \leq 0.05$ threshold was chosen for each parameter depending on whether the parameter value increased or decreased with abnormality, respectively. Venn diagrams were constructed to show the extent of overlap between the detection of progression between the three techniques; the analysis was repeated, taking into account the number of parameters showing evidence of change.

Results

Patient demographics, together with duration of follow-up, the number of examinations and data required by standards established by QUOROM[152] are given in the section 'Longitudinal high risk and longitudinal glaucoma groups' (p. 25); additional details are given in *Table 41*.

Figure 32 shows the frequency distribution of global MPD at baseline and completion of the study. The baseline and final global MPDs were –3.20 dB (95% CI: –3.84 to –2.56) and –3.27 dB (95% CI: –3.92 to –2.63), respectively; these two levels were not significantly different ($p = 0.52$; Wilcoxon matched pairs test). To quantify the overall amount of perimetric change that was observed during the period of the study, the frequency distribution of the gradients from linear regression of global MPD is shown in *Figure 33*.

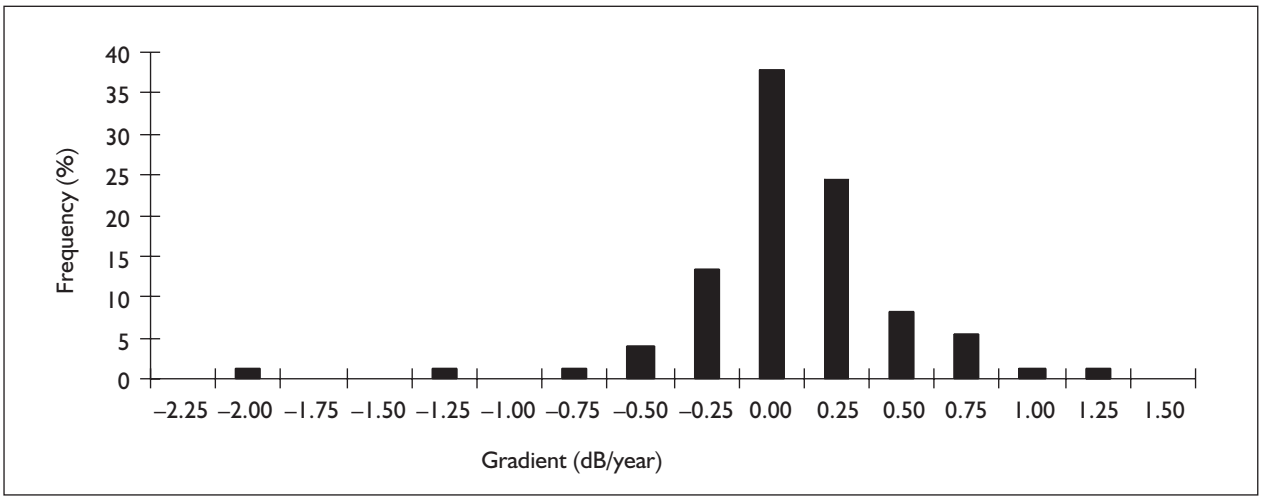


FIGURE 33 *Frequency distribution of gradient of global MPD*

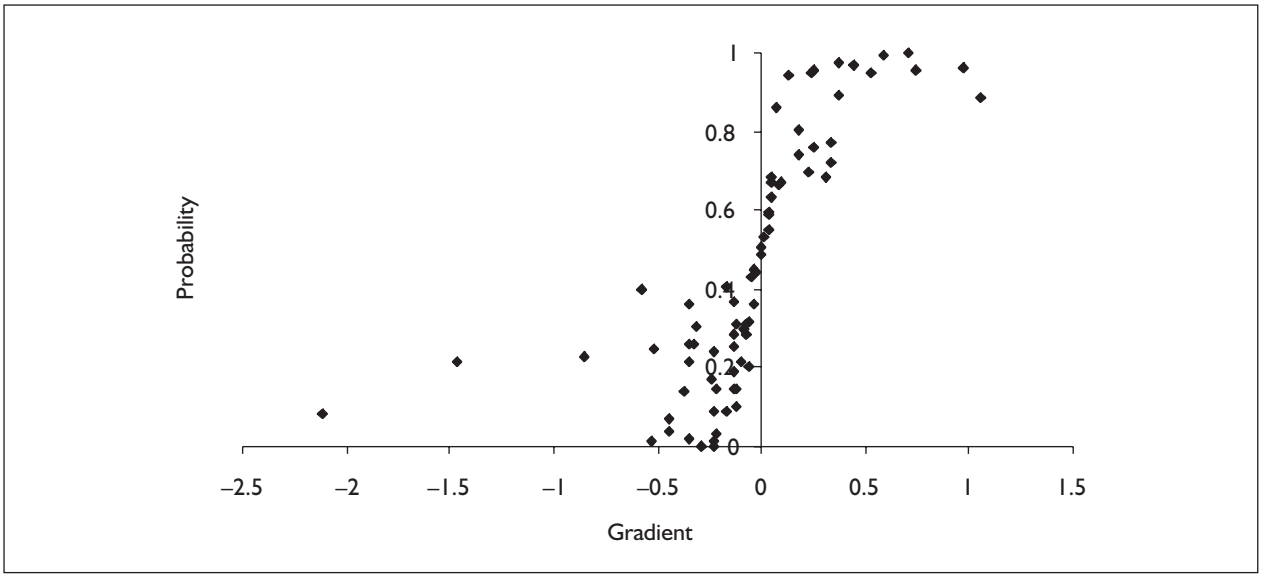


FIGURE 34 *Results from linear regression of global MPD*

Figure 34 shows the relationship between the gradient of the regression line and the associated probability value.

Table 42 illustrates the outcome of linear regression analysis of visual field data; the results from global, sectoral and pointwise analysis are given. In addition to cases showing significant deterioration (i.e. with negative gradient of regression line and probability values ≤0.05, ≤0.02, ≤0.01 and ≤0.005), the proportion of patients who showed apparent 'improvement' at the corresponding levels of significance is also given.

In total, 44 (59%) patients showed some evidence of perimetric progression by one or more of the three definitions. The number of patients who progressed according to the global, sectoral and pointwise criteria were 7 (9%), 42 (56%) and 14 (19%), respectively. Clearly, the criteria agreed on some cases, but disagreed on others. *Figure 35* shows the frequency distribution of glaucoma patients with respect to the number of perimetric criteria in which they showed visual field progression. The extent of agreement between the three methods is shown in *Figure 36*.

Figure 37 shows the relationship between baseline global MPD and the number of perimetric criteria that showed progression. There is no significant difference in the mean global MPD between the four categories.

TABLE 42 *Outcome of linear regression of visual field data [data shown as percentage of total group (n = 75)]*

		Sectoral (%)				Pointwise (%)			
***p*-Value**	**Global (%)**	**≥ 1 sector**	**≥ 2 sectors**	**≥ 3 sectors**	**≥ 4 sectors**	**≥ 1 point**	**≥ 2 points**	**≥ 3 points**	**≥ 4 points**
Negative gradient									
≤ 0.05	9	56	23	8	4	97	82	57	39
≤ 0.02	7	31	9	3	1	69	41	19	8
≤ 0.01	3	16	5	1	0	45	18	9	3
≤ 0.005	3	14	1	0	0	38	8	4	1
Positive gradient									
≥ 0.95	11	38	15	4	0	89	73	54	38
≥ 0.98	3	22	3	1	0	74	35	12	11
≥ 0.99	3	14	3	1	0	46	14	7	3
≥ 0.995	1	8	0	0	0	34	5	1	1

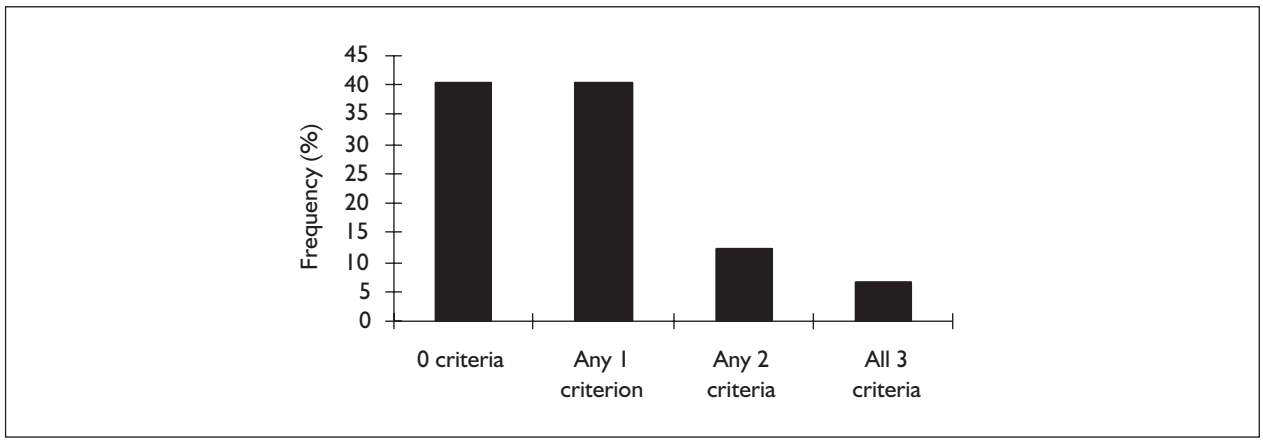


FIGURE 35 *Number of perimetric criteria showing progression (note that the groups are mutually exclusive)*

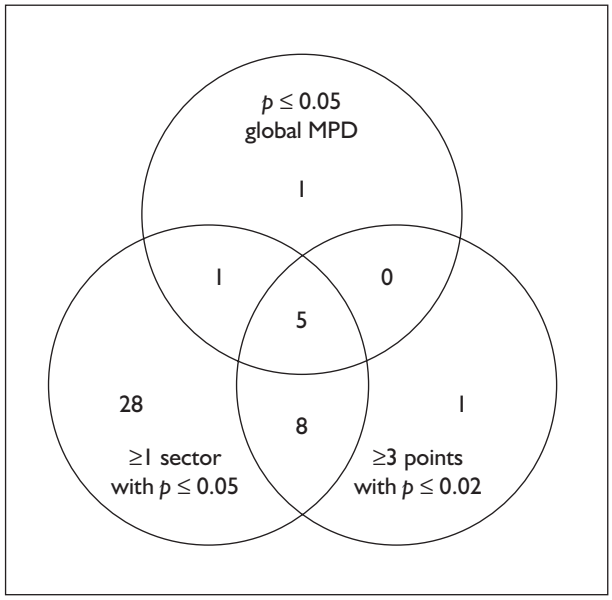


FIGURE 36 *Venn diagram showing agreement between different criteria for detecting perimetric progression (maximum = 75)*

From the entire longitudinal glaucoma group, the number of patients showing significant deterioration and improvement of their HRT and GDx parameters is shown in *Table 43*.

The extent of agreement between the HFA, HRT and GDx for glaucomatous progression when global, sectoral and pointwise definitions of perimetric progression were used is shown in *Figure 38*. In this case, significant deterioration with the HRT and GDx was defined as any patient who had linear regression that was significant at the ≤ 5% level (or ≥ 95% level, as appropriate) for one or more of the four parameters considered.

From the total of 75 eyes, the numbers showing definite progression, probable progression, suspect progression and stable visual fields were 5, 9, 30 and 31, respectively.

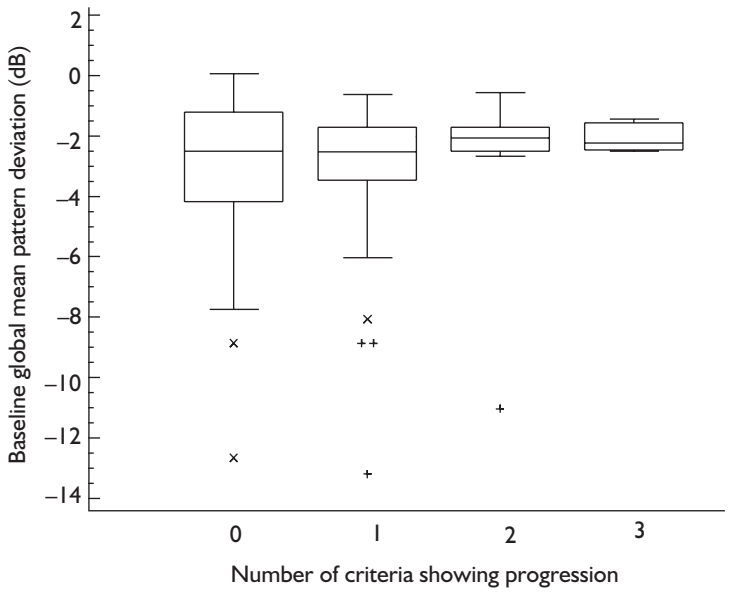


FIGURE 37 *Box and whisker plot showing relationship between baseline global MPD and number of criteria showing progression. Note that the number of subjects in each group was not equal. Outside value (×) = value smaller than the lower quartile minus 1.5 times the interquartile range. Far out value (+) = value smaller than the lower quartile minus 3 times the interquartile range.*

TABLE 43 *Number of patients with significant deterioration in linear regression of HRT and GDx parameters. (maximum = 75)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	8	9	6	8
No. of eyes with significant improvement	3	10	4	5
	GDx Parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	12	11	8	9
No. of eyes with significant improvement	1	2	3	3

Definite progressors

Five patients were defined as showing 'definite perimetric progression', that is, they fulfilled all three criteria for progression derived from linear regression of their perimetric data. From this group, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 44*. The agreement between the visual field analysis, HRT and GDx is shown in *Figure 39*.

Probable progressors

Nine patients were defined as showing 'probable progression', that is, they fulfilled any two criteria for progression derived from linear regression of their perimetric data (patients categorised as definite progressors were **not** included in this group). From the probable progressors, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 45*. The agreement between the visual field analysis, HRT and GDx is shown in *Figure 40*.

Suspect progressors

Thirty patients were defined as showing 'suspect progression', that is, they fulfilled any one criterion for progression derived from linear regression of their perimetric data (patients

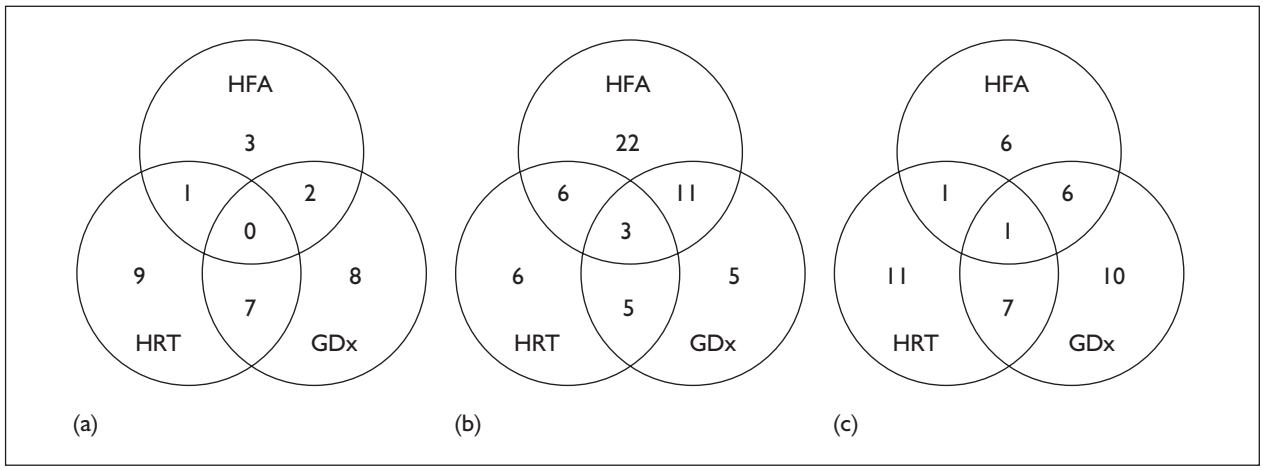


FIGURE 38 *Venn diagrams showing the extent of agreement between change in HRT and GDx parameters for different definitions of glaucomatous perimetric progression. Data from patients who showed progression according to (a) the global criterion, (b) the sectoral criterion and (c) the pointwise criterion.*

TABLE 44 *Number of definite progressors showing significant change in HRT and GDx parameters (maximum = 5)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	0	0	0	0
No. of eyes with significant improvement	1	0	1	1
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	2	2	1	0
No. of eyes with significant improvement	0	2	0	0

described in the previous two sections were **not** included in this group). From this population, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 46*. The agreement between visual field analysis, HRT and GDx is shown in *Figure 41*.

Stable visual fields

No significant perimetric change by any of the three methods was found in 31 eyes. From this population, the number of patients showing significant change in their HRT and GDx parameters is shown in *Table 47*. The agreement between visual field analysis, HRT and GDx is shown in *Figure 42*.

Discussion

This chapter has presented data collected on a group of glaucoma patients who were reviewed over a period of approximately 3.5 years. It has the methodological advantage of comparing two imaging techniques (HRT and GDx) in one population of patients. The scarcity in the literature of studies that review the longitudinal examination of perimetric, ONH and RNFL characteristics in glaucoma patients has been identified.[220]

The longitudinal glaucoma category of patients contained 75 eyes from 75 patients; this figure represents an over-achievement of the original target by 50%. As in previous chapters, a

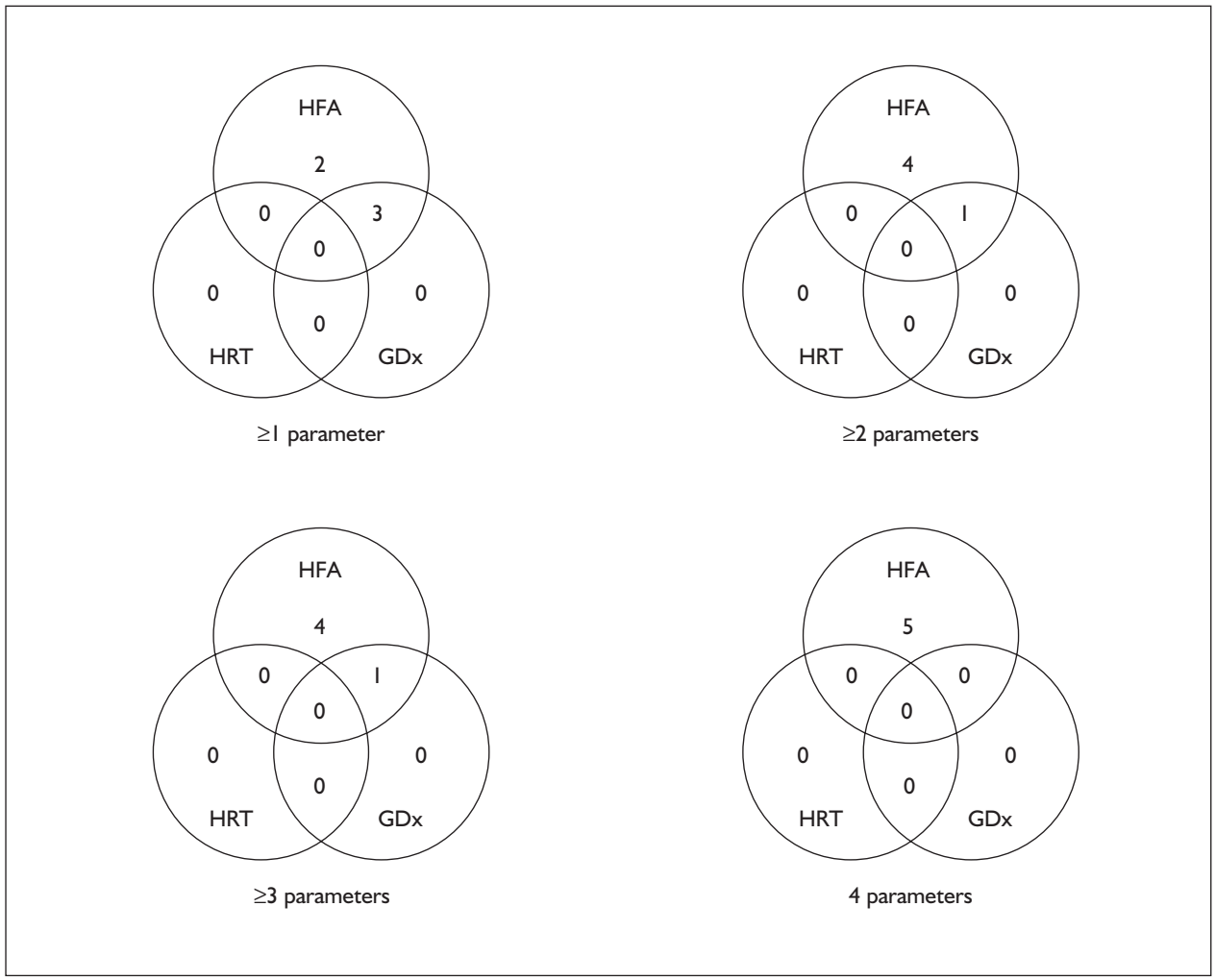


FIGURE 39 *Agreement between visual field analysis, HRT and GDx in patients showing definite progression (total = 5)*

TABLE 45 *Number of probable progressors showing significant change in HRT and GDx parameters (maximum = 9)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:P ratio**	**Bathija DFA**
No. of eyes with significant deterioration	1	2	1	1
No. of eyes with significant improvement	1	2	1	1
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	2	2	2	1
No. of eyes with significant improvement	0	0	0	0

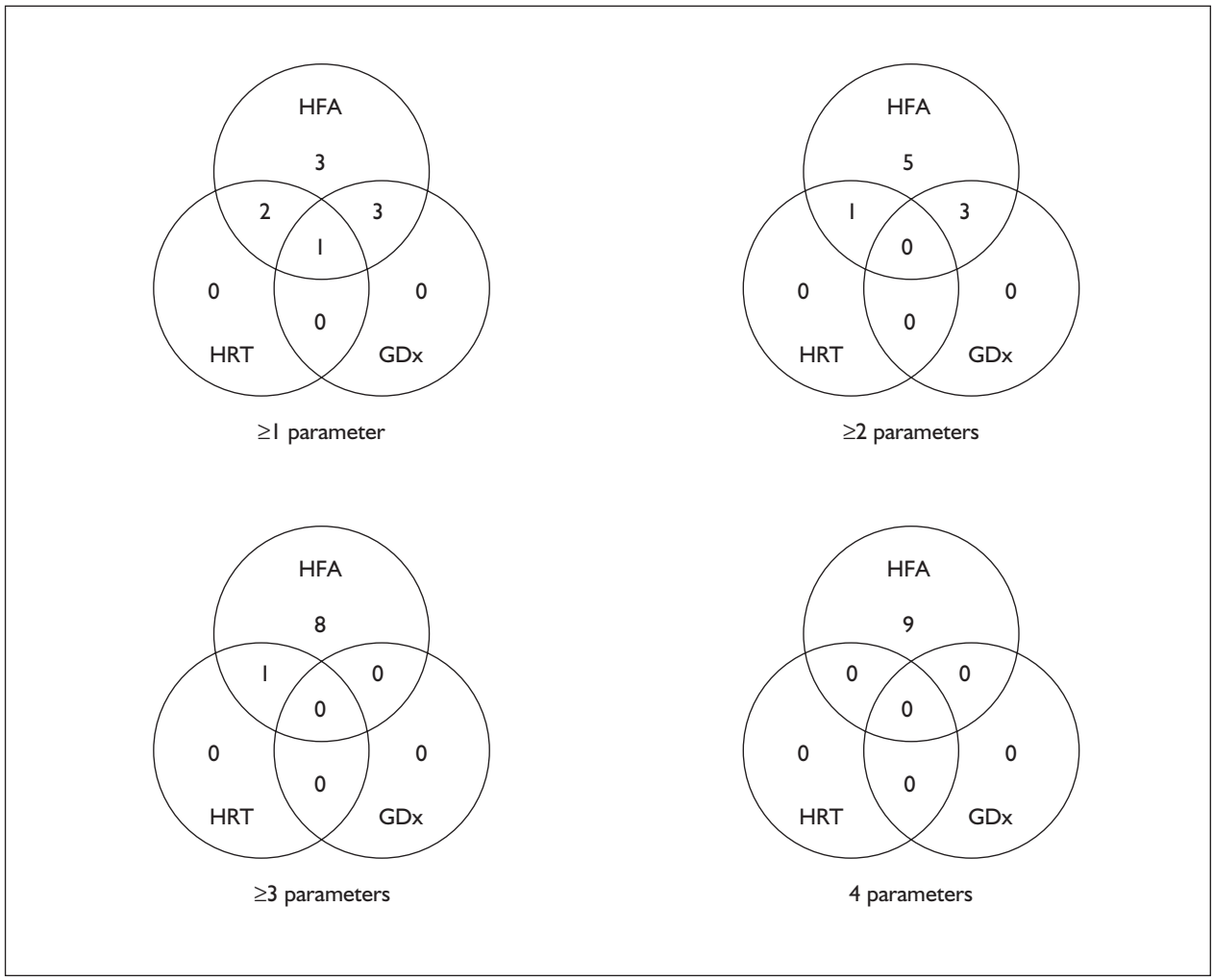


FIGURE 40 *Agreement between visual field analysis, HRT and GDx in patients showing probable progression (total = 9)*

TABLE 46 *Number of suspect progressors showing significant change in HRT and GDx parameters (maximum = 30)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	2	4	2	2
No. of eyes with significant improvement	1	5	1	2
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	2	4	1	3
No. of eyes with significant improvement	0	0	2	1

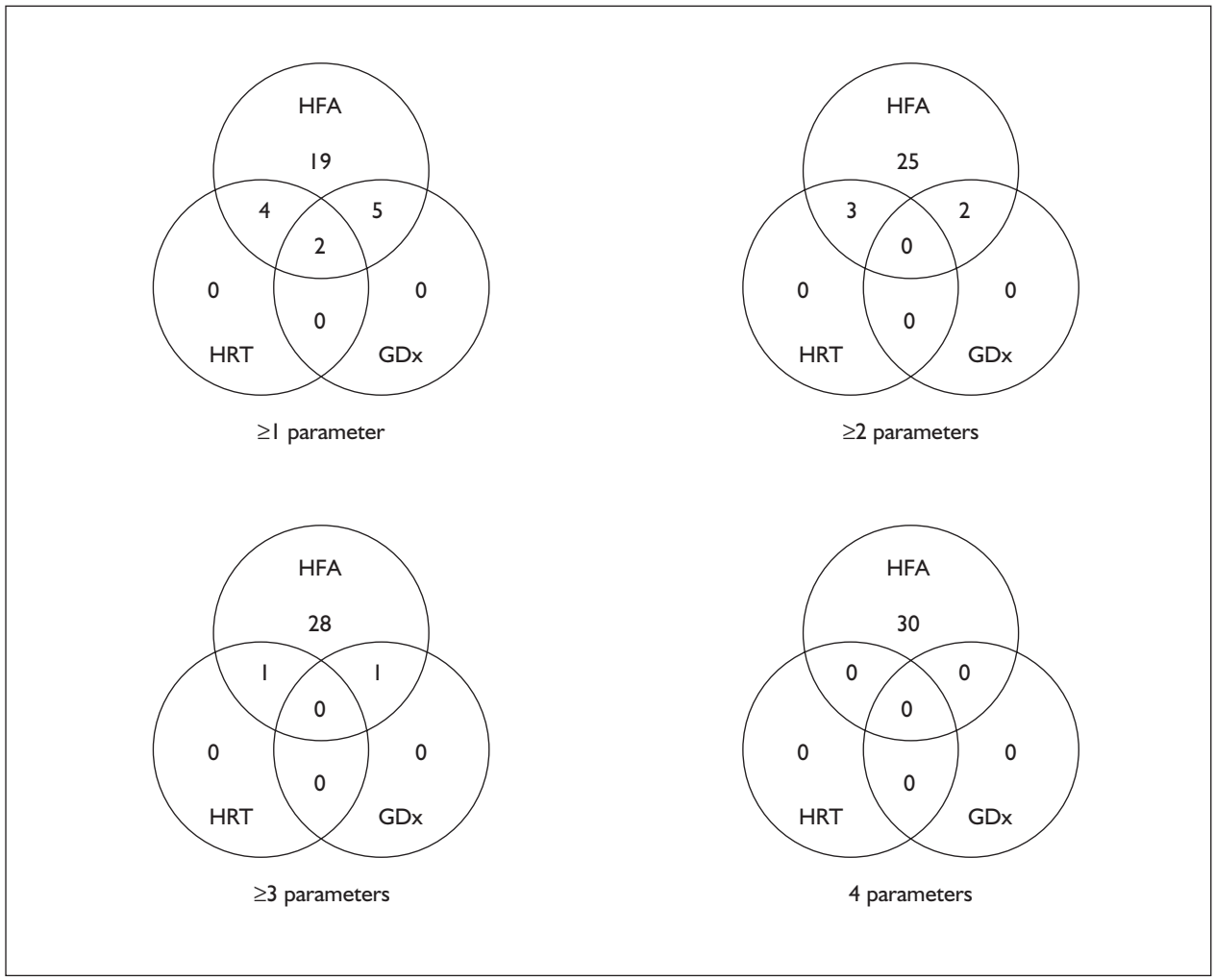


FIGURE 41 *Agreement between visual field analysis, HRT and GDx in patients showing suspect progression (total = 30)*

TABLE 47 *Number of suspect progressors showing significant change in HRT and GDx parameters (maximum = 31)*

	HRT parameter			
	C:D area ratio	**Mean RNFL thickness**	**Vertical C:D ratio**	**Bathija DFA**
No. of eyes with significant deterioration	5	3	3	5
No. of eyes with significant improvement	0	3	1	1
	GDx parameter			
	Number	**Ellipse modulation**	**Maximum modulation**	**Superior maximum: nasal median**
No. of eyes with significant deterioration	6	3	4	5
No. of eyes with significant improvement	1	0	1	2

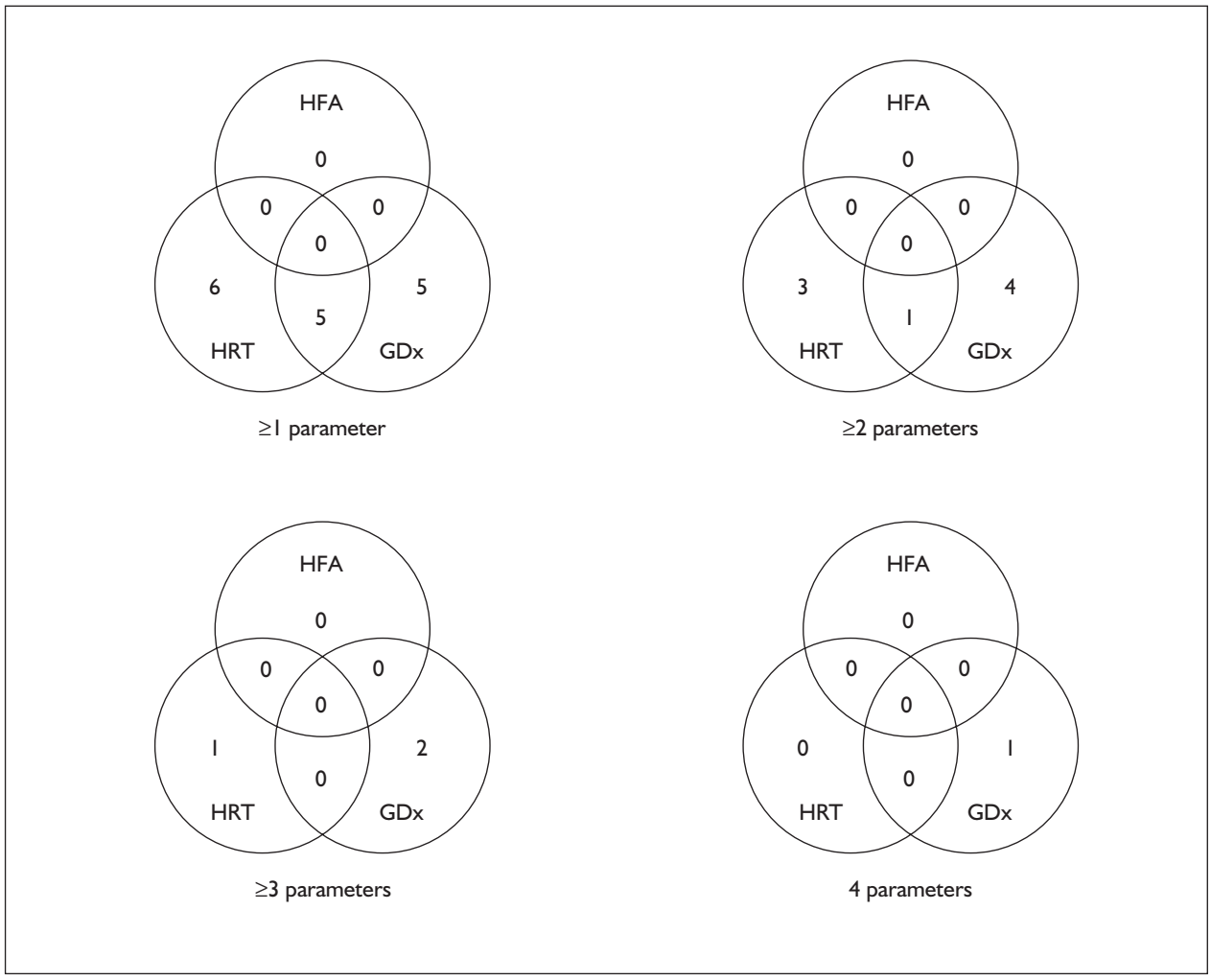


FIGURE 42 *Agreement between visual field analysis, HRT and GDx in patients with stable visual fields (total = 31)*

perimetric criterion for establishing progression was adopted. Three different definitions were used – the rationale for each is given in the section 'Discussion' (p. 68).

Overall, there was very little deterioration in the visual field status of most patients; indeed, there was no statistically significant difference between the global MPD of the whole group between baseline and final examinations. A variety of factors may contribute towards this finding:

- Patients were receiving medication (or had previously undergone glaucoma filtration surgery) to lower their IOP. These interventions have been shown to lower the risk of progression.[215,221–224]
- The study protocols excluded patients with visual acuity worse than 6/12. This strategy restricted the level of disease within the experimental group and may have lessened the proportion of progressors, since patients with more advanced glaucoma are thought to be more likely to undergo progression.[225]
- Patients who had previously undergone complicated glaucoma filtration surgery were excluded from the study, as were individuals who had secondary glaucoma (e.g. due to pseudoexfoliation or pigment dispersion).[225]
- There is a likely selection bias in recruiting patients for a longitudinal study such as the Glaucoma Imaging Study, which requires a significant commitment from participants. Patients who enrol are likely to be compliant conscientious individuals who use their medication as directed and regularly attend clinic appointments.
- A further form of selection bias might have occurred: if an individual was invited to join the study, but was undergoing a period of rapid decline in their condition, they will have been undergoing frequent review at clinic (e.g. every

3 months) and, consequently, may have felt reluctant to attend for supplementary research appointments.
- The fact that patients were participating in the study may have led to a decreased rate of progression – at each appointment, if there was any evidence of clinical deterioration (e.g. decreased visual acuity, ONH and/or RNFL change, worsening of the visual field), poor concordance with medication regimen or adverse drug response, then the patient was referred back to their consultant ophthalmologist. This strategy may have resulted in a more rapid response to a small decline than if the patient waited until their routine clinic appointment.
- The relatively short length of follow-up in this study will have reduced the number of progressors. Review periods ranged from 1.1 to 5.1 years (median = 3.5 years). In another study, with much longer follow-up (8.5–24.1 years; median = 15.0 years), 60% of eyes showed a significant rate of change.[226]

The factors stated above might have all contributed, in some way, towards a low rate of perimetric deterioration within the experimental group. Findings from this study may, therefore, not relate to an average glaucoma population, where a larger proportion of patients may show a decline.

When the progressors were categorised according to the level of certainty that they had shown evidence of perimetric change, the proportions of patients who demonstrated definite, probable and suspect progression were 7, 12, and 40%, respectively (these groups were mutually exclusive). In common with the previous chapter, which reviewed glaucomatous conversion, data from the glaucoma patients also highlight the vulnerability of the number of progressors to the definition employed. Using global, sectoral and pointwise definitions, 9, 56 and 19% of patients, respectively, showed evidence of perimetric progression. There was only a relatively small extent of agreement between the three criteria – 11% of cases. The susceptibility of the number of patients showing perimetric deterioration to the criteria used has been previously identified by Johnson and colleagues.[209]

The Early Manifest Glaucoma Trial (EMGT) found that 51% of untreated and 41% of treated glaucoma patients showed evidence of progression.[215] However, different criteria for establishing perimetric and morphological change were used compared with this study. An interesting similarity in the findings between the two studies is that the majority of progressors from the EMGT were detected by evidence of perimetric deterioration, rather than from observation of alterations at the ONH. This result, together with data presented in this chapter, conflicts with previous scientific thinking that structural change precedes functional change.[33,201] However, other research groups have made similar findings,[34,35] which also challenge the previously held assertion of the temporal relationship between structure and function.

In common with the previous chapter, linear regression analysis was used to detect change in perimetric, ONH and RNFL characteristics. This technique has the advantage that it will identify a small magnitude of change in patients who show low intrasubject variability, but requires a larger magnitude of change in cases of higher variability. Detecting change by utilising regression is a superior technique to a 'change analysis' approach, where individual values are compared with group values. Additionally, linear regression of perimetric data has been shown to be a useful tool for predicting future loss.[227,228]

In addition to calculating the number of patients who showed 'deterioration' in their perimetric, HRT and GDx indices, the proportion of patients who showed 'improvement' was also calculated; the latter could be considered a surrogate for the FP rate. Following linear regression of the global MPD, 9 and 11% of subjects showed significant deterioration and improvement, respectively, at the 5% level. When sectoral analysis was employed, the proportions were 56 and 38% (at the same level of significance). From the pointwise data, 57 and 54% showed significant negative and positive change when the threshold was set at ≥ 3 points. These data are a manifestation of the high level of noise encountered in the glaucomatous visual field.[229–231]

In contrast to the number of patients showing apparent 'improvement' in perimetric characteristics, some of the HRT and GDx parameters appeared to be less immune to intrasubject variability. In particular, linear regression analysis of the GDx number identified very few apparent cases of 'improvement'. This finding may indicate that the number may have significant promise for detection of longitudinal change, with the inclusion of very few FPs. The opposite scenario is found in the HRT parameter mean RNFL thickness. In most cases, the number

of significantly positive and negative gradients was approximately equal, which suggests that this parameter is particularly vulnerable to measurement variability. Owing to the small numbers of cases showing significant change, further analysis of this data is not possible.

With regard to the extent of commonality between the detection of change between the three different forms of examination, the outcome is in line with the previous chapter, namely that there is little overlap between perimetry, HRT and GDx. When the threshold for abnormality was set at ≥ 1 parameter, perimetry and the GDx agreed on the presence of progression in three out of five cases of definite progression. As the threshold for the number of parameters was increased, the level of agreement decreased – this trend was consistent for all the other subgroups. Unfortunately, further investigation of the proportion of cases detected by the different techniques is undermined by the small number of progressors.

The lack of agreement between the three modalities of examination could be a manifestation of incorrect identification of true progressors. It has already been identified that a relatively high proportion of cases showed significant perimetric 'improvement'; this finding gives an indication that a high proportion of 'converters' may be FPs. Consequently, a lack of concordance between the presence of deterioration detected by perimetry, HRT and GDx would be expected.

This study found that there was no relationship between baseline global MPD and number of criteria showing progression. This finding differs from EMTG data, where it has been shown that eyes with more significant perimetric damage are more likely to undergo further deterioration.[232] The result from the Glaucoma Imaging Study may have been affected by the small number of cases showing definite progression ($n = 5$). A study by Schwartz and colleagues[233] conflicts with the data from the EMTG: the former reports a more rapid rate of visual field loss in less damaged fields. The disparate conclusions may be explained by differences in the populations used by the two studies: the EMTG population had baseline mean deviation of –5 dB, whereas the subjects in Schwartz and colleagues' study had far more advanced glaucoma (baseline mean threshold = 18 dB). The combined results of the two studies may suggest different phases of perimetric decline at different stages of the condition.

A limitation of the study is that it has not performed any form of 'event analysis'. During the period of review, some patients might have undergone an episode of relatively rapid decline that was halted owing to alteration in medication regimen and/or surgical intervention. Associated changes in visual field, optic disc and/or RNFL characteristics may not have been detected by linear regression analysis, and so the case not identified as one that showed significant deterioration.

The wider findings of the study are limited by the early stage of the condition in the experimental group. Ideally, the glaucoma population could have been stratified according to the level of perimetric damage and the analysis repeated to determine the ability of the HRT and GDx to determine progression at different stages of the disease. However, owing to the constraints of the cohort, namely that the patients showed mainly early perimetric loss and few demonstrated a large extent of deterioration, such analysis was not possible.

Overall, longitudinal follow-up of our cohort of glaucoma patients showed that most of the group remained stable, with only a small proportion showing definite progression (five out of 75 eyes). There was little commonality in patients identified by perimetry, HRT and GDx as progressing. In common with the findings from the high risk group, data from the glaucoma cohort suggest that examination with the HRT and/or GDx cannot provide a replacement for visual field examination. Instead, it is more likely that the instruments would provide a useful clinical adjunct in facilitating rapid and objective examination of the ONH and RNFL. Reasons underlying the lack of agreement between perimetry, HRT and GDx may lie in the complex structure–function relationship in glaucoma. Future research to elucidate fully the nature of this relationship may explain why the HRT and GDx fail to detect cases of glaucomatous progression.

Chapter 10

Interoperator variability in Heidelberg Retina Tomograph and GDx parameters

The aim of this chapter is to determine the magnitude of interoperator variability of HRT and GDx parameters. There are two main components to interoperator variability: image capture and the placement of the peri-papillary contour line. This chapter does not discriminate between the two sources of variability, and considers them as one entity; however, Chapter 11 reviews, in detail, the effects of variability in placement of the peri-papillary contour line.

Introduction

Individual studies have previously considered reproducibility for the HRT[78,234–237] and GDx.[110,238,239] However, there has been no study that compares the reproducibilities of the two instruments on the same group of patients. Such a study is necessary to provide vital information in the overall clinical evaluation of the equipment.

Method

Data presented in this chapter were collected during supplementary examination sessions, and are not presented elsewhere in the report.

Three images of the ONH and peri-papillary RNFL were taken with the HRT and GDx, respectively, on five subjects by five operators. Two glaucoma patients were chosen randomly from the study database and three control individuals volunteered to participate. The chin-rests were adjusted after examination of each patient. During examination, all study criteria were maintained.

A mean image was computed from the three single images for both the HRT and the GDx. Each observer identified the limits of the ONH, according to the procedure described in the section 'Data collection' (p. 18). Operators were blinded to their colleagues' contour lines.

The five operators all had strong histories in clinical glaucoma and/or glaucoma imaging. One (SQ) is an ophthalmologist with a special interest in glaucoma. Three (AKJ, IPG and AC) are present (or previous) data collection technicians for the Glaucoma Imaging Study. One (RAH) is an optometrist with an extensive track record in both clinical and academic glaucoma and is also a co-author of this report.

The five HRT parameters that were entered into the analysis were selected on the basis of having optimum diagnostic precision, as determined within the literature. The same criterion was applied to the GDx. The parameter disc area was included in the analysis of HRT parameters, since it provides a useful measure of variability due to contour line alignment; it is also of clinical significance because of the role it plays in the MRA [see the section 'Diagnostic algorithms'. (p. 8)].[95]

Agreement between the five observers was tested for each parameter using the intraclass correlation coefficient (ICC). Agreement was preferred to simple correlation since ICC measures whether observations are identical, whereas correlation ascertains the presence of a linear relationship between observations. Also, ICC allows for variation between subjects. ICC outputs values that lie between –1.00 and 1.00; guidelines for suggested interpretation are given in *Table 48*.[240] It is widely accepted that ICC ≥ 0.90 indicated high reproducibility of measurements.

To identify if there was a certain trend in observer agreement, the coefficient of determination (r^2) was calculated for each possible combination of

TABLE 48 *Interpretation of intraclass correlation coefficient*

Intraclass correlation coefficient	Interpretation of agreement
–1.0 to 0.0	Poor agreement
0.01 to 0.20	Slight agreement
0.21 to 0.40	Fair agreement
0.41 to 0.60	Moderate agreement
0.61 to 0.80	Substantial agreement
0.81 to 0.99	Almost perfect agreement
1.0	Perfect agreement

TABLE 49 *HRT parameter values for each operator: values are presented as mean (standard deviation)*

Operator initials	HRT parameter				
	Disc area (mm^2)	Rim area (mm^2)	Cup shape measure	Mikelberg DFA	Bathija DFA
AC	2.04 (0.51)	1.33 (0.25)	–0.10 (0.02)	0.42 (0.79)	0.88 (0.94)
AKJ	1.98 (0.44)	1.29 (0.17)	–0.10 (0.04)	0.53 (1.65)	0.70 (0.71)
IPG	1.92 (0.54)	1.31 (0.23)	–0.09 (0.02)	0.35 (1.15)	1.12 (0.66)
RAH	1.80 (0.44)	1.18 (0.08)	–0.11 (0.04)	0.02 (0.81)	1.25 (0.76)
SQ	2.12 (0.49)	1.30 (0.27)	–0.11 (0.04)	0.50 (2.22)	1.09 (1.33)

TABLE 50 *GDx parameter values for each operator: values are presented as mean (standard deviation)*

Operator initials	GDx parameters				
	Number	Superior maximum: nasal median	Superior ratio	Ellipse modulation	Maximum modulation
AC	21.6 (18.1)	1.93 (0.20)	2.39 (0.27)	2.06 (0.37)	1.41 (0.28)
AKJ	24.0 (18.3)	1.88 (0.16)	2.35 (0.29)	2.05 (0.38)	1.39 (0.27)
IPG	21.0 (17.2)	1.92 (0.24)	2.51 (0.33)	2.23 (0.45)	1.51 (0.33)
RAH	21.2 (20.1)	1.88 (0.20)	2.53 (0.37)	2.18 (0.28)	1.60 (0.39)
SQ	19.0 (15.1)	1.94 (0.22)	2.49 (0.29)	2.07 (0.30)	1.51 (0.28)

two observers. Bland and Altman plots[241] were, subsequently, constructed for pairs of observers.

Results

Mean HRT and GDx parameter values for each observer are shown in *Tables 49* and *50*, respectively.

Tables 51 and *52* give the ICC for each HRT and GDx parameter, respectively. Interobserver agreement for the HRT parameters was almost perfect for all parameters except rim area, which had substantial agreement; ICCs of the five selected parameters ranged from 0.66 to 0.99. Interobserver agreement for the GDx was almost perfect for all five parameters (ICCs from 0.86 to 0.99).

Agreement between operators is shown in *Tables 53* and *54*. The general pattern was of considerable variability between the extent of agreement for the different pairs of examiners: a particular pair may have had good agreement for one parameter, but poor concurrence for another. Overall, the highest level of agreement was found between observers RAH and SQ, although for the parameter rim area, this pair showed poor agreement ($r^2 = 0.15$).

For the GDx parameters, the highest level of agreement was between operators AC and SQ. These two operators had very high levels of concordance for all parameters, except ellipse modulation.

The Bland and Altman plots (*Figures 43–46*) show examples of high and low variability for HRT and

TABLE 51 *Interobserver agreement of HRT parameters*

HRT parameter	ICC
Disc area	0.99
Rim area	0.66
Cup shape measure	0.92
Mikelberg DFA	0.87
Bathija DFA	0.95

TABLE 52 *Interobserver agreement of GDx parameters*

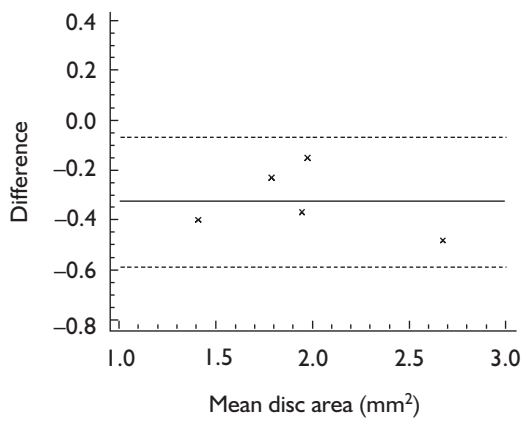
GDx parameter	ICC
Number	0.99
Superior maximum:nasal median	0.92
Superior ratio	0.93
Ellipse modulation	0.86
Maximum modulation	0.94

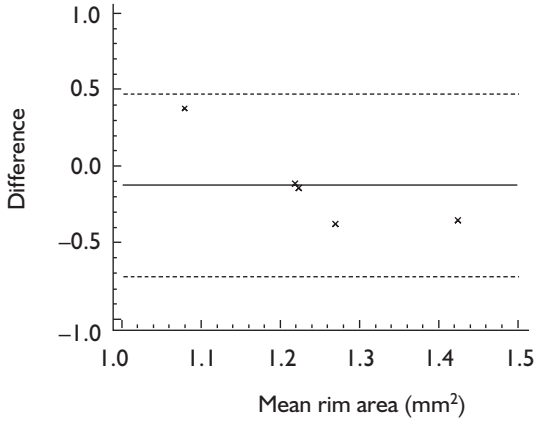
TABLE 53 *Agreement between observers with HRT parameters*

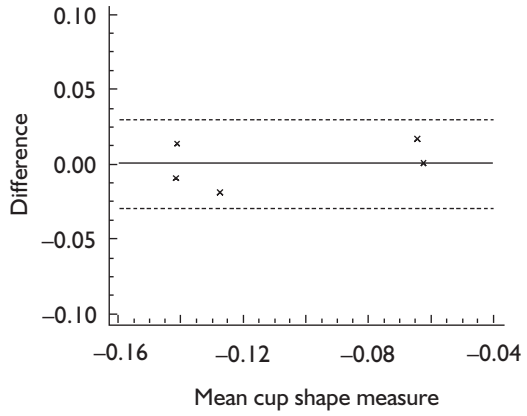
Pairs of operators	r^2 values for HRT parameters				
	Disc area	Rim area	Cup shape measure	Mikelberg DFA	Bathija DFA
AC & AKJ	0.96	0.44	0.10	0.23	0.71
AC & IPG	0.99	0.61	0.05	0.37	0.82
AC & RAH	0.92	0.85	0.13	0.31	0.92
AC & SQ	0.89	0.00	0.08	0.20	0.81
AKJ & IPG	0.95	0.20	0.94	0.39	0.71
AKJ & RAH	0.96	0.19	0.98	0.77	0.52
AKJ & SQ	0.97	0.13	0.80	0.88	0.61
IPG & RAH	0.92	0.72	0.95	0.14	0.61
IPG & SQ	0.86	0.02	0.97	0.38	0.78
RAH & SQ	0.93	0.15	0.89	0.85	0.81

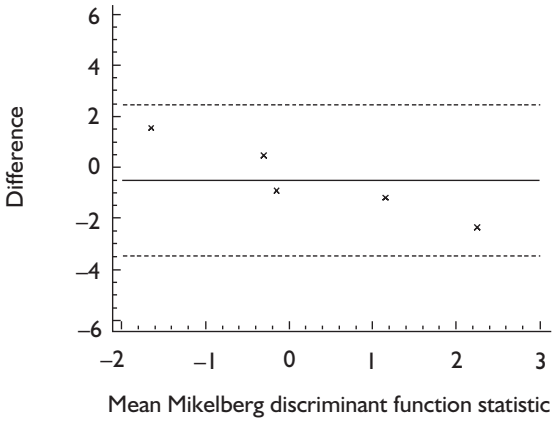
TABLE 54 *Agreement between operators with GDx parameters*

Pairs of operators	r^2 values for GDx parameters				
	Number	Superior maximum: nasal median	Superior ratio	Ellipse modulation	Maximum modulation
AC & AKJ	0.91	0.59	0.27	0.99	0.41
AC & IPG	0.97	0.44	0.67	0.90	0.79
AC & RAH	1.00	0.47	0.44	0.01	0.43
AC & SQ	1.00	0.92	0.95	0.77	0.98
AKJ & IPG	0.93	0.25	0.74	0.95	0.70
AKJ & RAH	0.92	0.64	0.46	0.01	0.70
AKJ & SQ	0.91	0.83	0.44	0.70	0.50
IPG & RAH	0.98	0.27	0.33	0.04	0.40
IPG & SQ	0.98	0.45	0.82	0.51	0.85
RAH & SQ	1.00	0.54	0.49	0.06	0.51









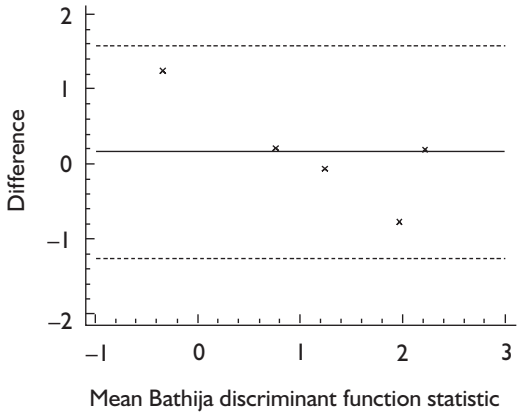


FIGURE 43 *Bland and Altman plots for HRT parameters from observers RAH and SQ. Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

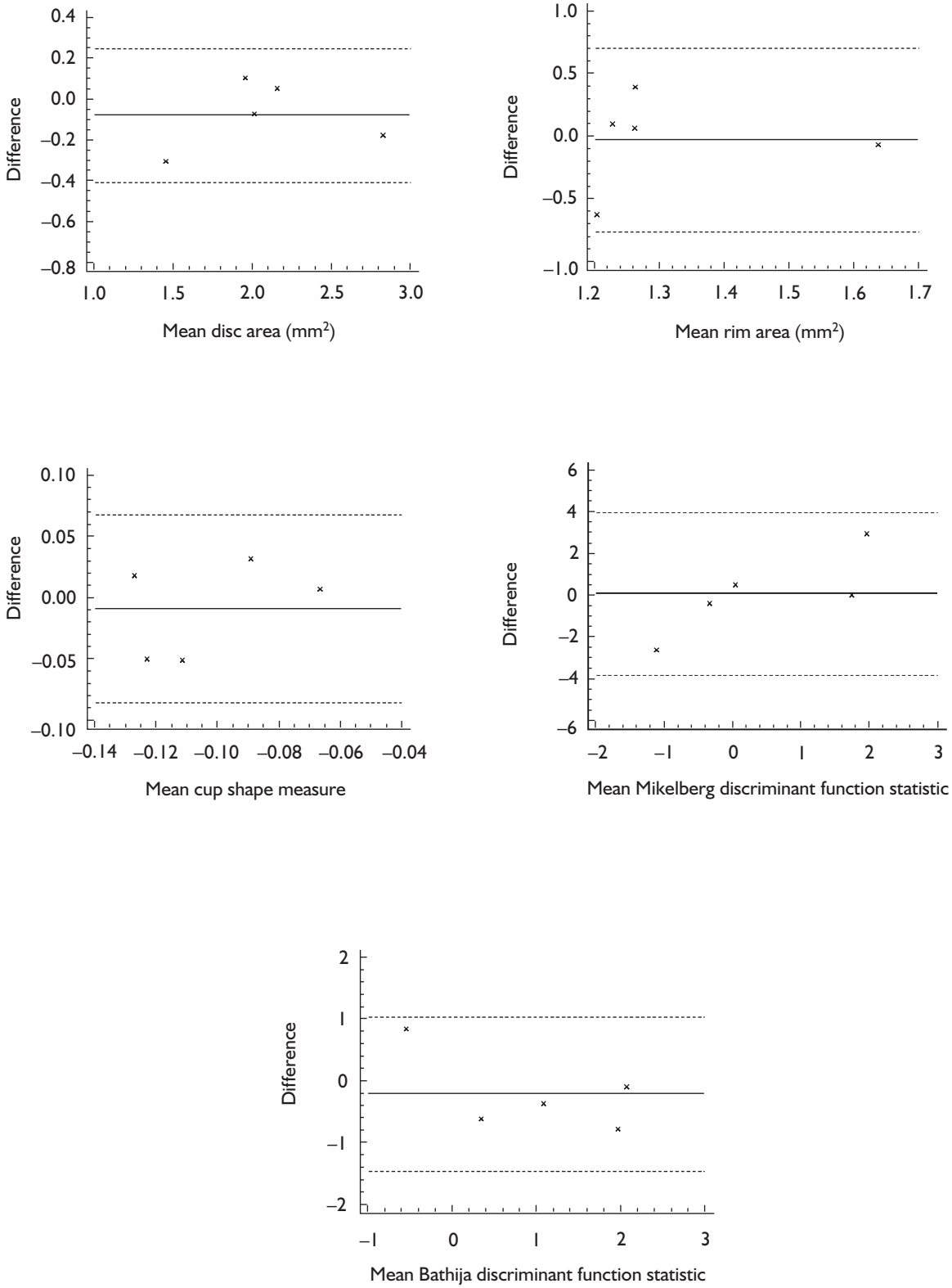


FIGURE 44 *Bland and Altman plots for HRT parameters from observers AC and SQ. Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

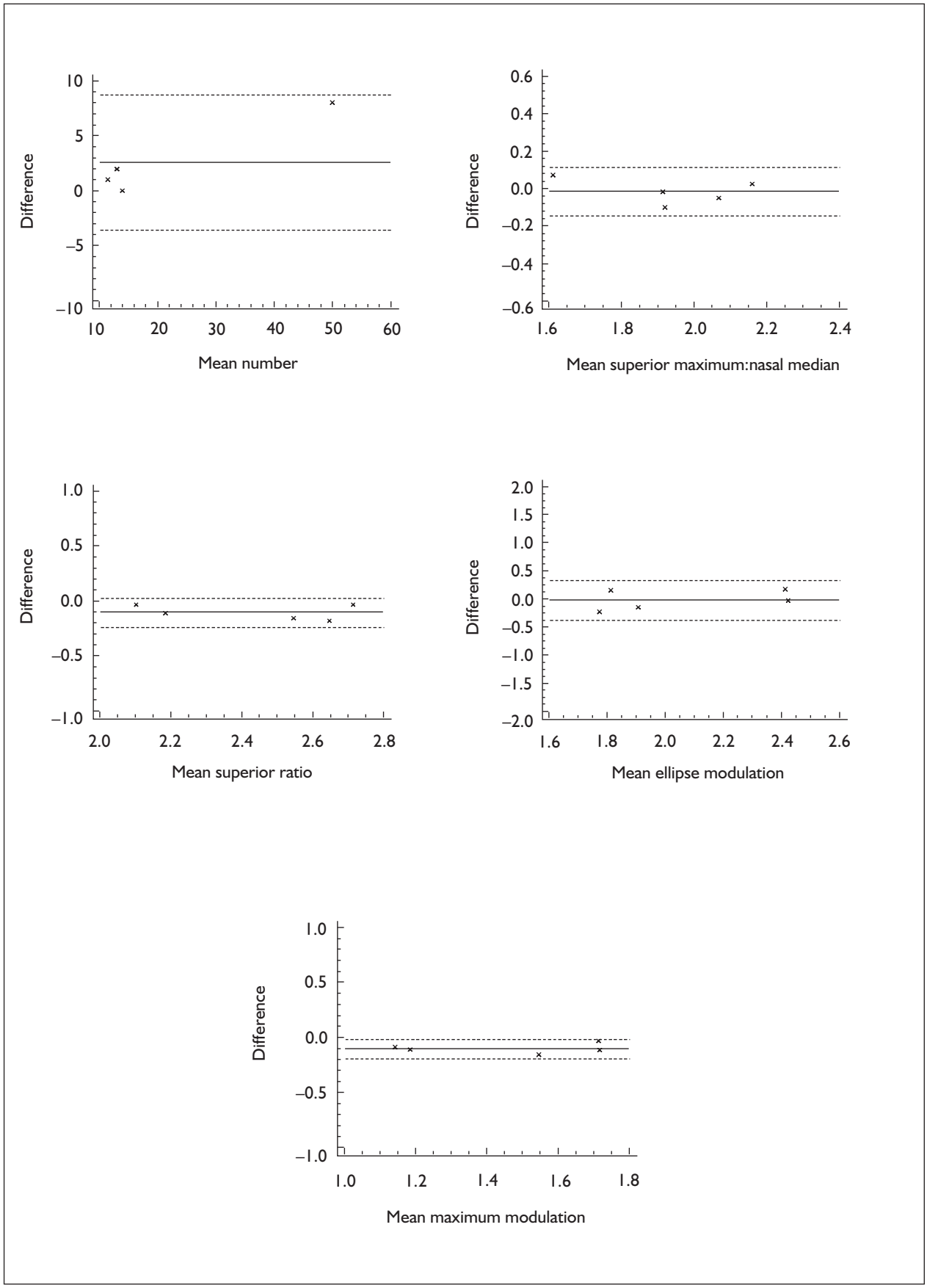


FIGURE 45 *Bland and Altman plots for GDx parameters from observers AC and SQ. Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

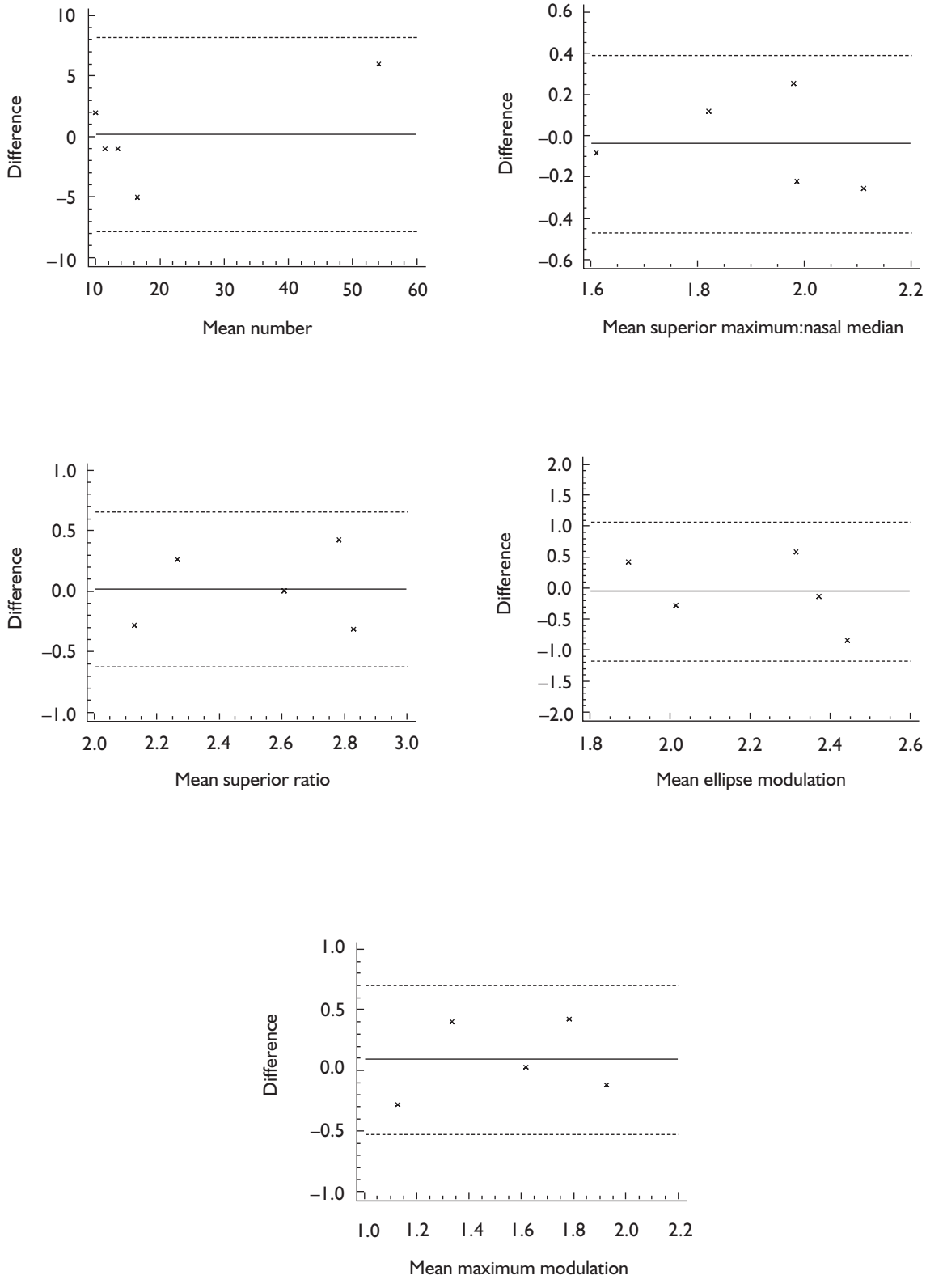


FIGURE 46 *Bland and Altman plots for GDx parameters from observers IPG and RAH. Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

GDx parameters. HRT parameters shown in *Figure 43* are from operators RAH and SQ who, generally, showed good agreement. These graphs are contrasted in *Figure 44* with data from AC and SQ, who showed poor agreement. Of interest, the latter pair of operators had high concurrence for GDx parameters (*Figure 45*), compared with IPG and RAH, for whom higher variability was found (*Figure 46*). (For each parameter, the ordinate scale has been matched between graphs for different pairs of operators to facilitate comparison.) The low sample size ($n = 5$) does not allow for any meaningful analysis of the variability with parameter value.

Discussion

This study determined interobserver agreement for five parameters from the HRT and five from the GDx. Except for the HRT parameter disc area, the parameters were selected because they had previously been shown to provide optimal discrimination between glaucoma patients and normal individuals. For all the parameters, interobserver agreement was very good for both the HRT and GDx. This study is of interest, since no study where repeatability of the two instruments has been compared in the same group of patients has been reported.

For the HRT parameters, agreement was substantial to almost perfect and for GDx almost perfect for all parameters. This finding has important implications for the utilisation of the equipment in the normal clinical environment, where multiple users may operate the same equipment.

There were, however, some differences between operators. RAH and SQ had the highest level of agreement for the HRT, although one parameter showed a very low coefficient of determination ($r^2 = 0.15$). For the GDx parameters, operators AC and SQ had the highest level of agreement for all parameters, except ellipse modulation.

Of the HRT parameters, disc area provides the most direct measurement of the precision of placement of the contour line. The parameter is also of clinical significance because of the role it plays in the MRA.[95] This parameter had the highest agreement between the operators. This finding is slightly surprising, since it would be thought that the disc area would be subject to more variability between operators because of the influence of the contour line placement. The conflicting argument is that disc area may be less affected by sources of variability in image capture, such as head position, eye position and angle of incidence of the laser beam into the eye, on account of the fact that it does not describe aspects of the three-dimensional characteristics of the disc (such as cupping).

Of the five GDx parameters, ellipse modulation is probably the most sensitive to variation in contour line placement, since the calculated area contains the pixels covered by the measurement ellipse surrounding the ONH. Subsequently, this parameter had the lowest ICC and, therefore, the least agreement between operators. Despite these factors, agreement was almost perfect. The highest ICC value was found for the number. This might be explained by the fact that the number is calculated by a neural network, rather than a direct measurement of RNFL characteristics. Therefore, the effect of interobserver variability may have been smoothed out by values that are insensitive to such variability.

Although the experimental group comprised a mixture of control individuals and glaucoma patients, the small numbers of each (three and two, respectively) do not allow meaningful analysis of the effect of diagnosis on variability. However, the literature is unequivocal that variability is higher in glaucomatous than normal eyes.[77–79]

The issue of repeatability must be considered within the context of diagnostic precision (Chapter 7). Bailey and colleagues[242] highlighted the issue that, if concordance between examinations is high, it may indicate that the scale of measurements is too coarse, which would reduce diagnostic precision.

Our method of capturing three images on each eye with each instrument was to simulate the use of the equipment in normal clinical practice. This strategy differs from the study's rigorous criteria of capturing five images of each eye and the operator then selecting the three best from the original five images. It was felt that, for the purposes of this study, the selection process would add an undesirable component of variability.

Overall, interoperator agreement between five observers for five HRT and GDx parameters was very good, with almost all parameters demonstrating almost perfect agreement.

Chapter 11

Effect of inter- and intraobserver variability in the definition of the peri-papillary contour line on Heidelberg Retina Tomograph and GDx parameters

Chapter 10 determined interobserver variability in image capture with the HRT and GDx. This chapter examines, in detail, one specific component of this variability – definition of the optic disc margin.

Introduction

In order to perform their calculations, both the HRT and GDx require the operator to define the margin of the optic disc. This study was carried out in response to an observation that the Mikelberg DFA[87] could yield a different diagnostic category following a minor modification to the disc outline. An example of this occurrence is shown in *Figure 47*, where the 'classification' changes following minor modification to the contour line.

A positive attribute of a clinical instrument is its independence from operator effects. Our

FIGURE 47 *HRT diagnostic classification. Parts (A) and (B) show conflicting diagnostic categorisation following a minor modification to the contour line. Results are shown for subject 56 (EC), who showed an early perimetric defect.*

experience has shown that this is not necessarily the case with the HRT. This study aimed to establish the magnitude of the variability and to investigate whether the GDx shows similar susceptibility.[185]

This feature, *per se*, has not been widely explored. Variability has been investigated at the optic disc[75,76] and macula,[80,243,244] but these studies (as with Chapter 10 of this report) included image capture as a variable. A recent study[245] examined the effect of intraobserver variability in the placement of the contour line on nine HRT stereometric parameters. Each disc was outlined seven times, producing ICCs of 0.861–0.997 (mean 0.934) for 10 normal individuals and 0.946–0.999 (mean 0.957) for 10 glaucoma patients. However, the sample size was small and only one operator's variability was analysed. Another study that examined the effects of inter- and intraobserver variability in contour line placement on HRT parameters was performed by Garway-Heath and colleagues.[246] They found that the interobserver coefficients of variation were 4.4% (disc area), 8.2% (neuro-retinal rim area), 13.6% (rim volume) and 7.2% (reference height). For within-observer variability, the coefficients of variation were 2.4, 4.5, 8.4 and 5.7% for the same parameters.[246]

Spencer and colleagues noted high interobserver reproducibility (ICC = 0.969)[236] for HRT disc height using the 'circle' facility rather than drawing around the disc margin. They did not report on any other HRT parameters.

As with the HRT, studies reviewing GDx variability include image capture.[110,238,239] At the time of writing, there is no specific mention in the literature of the effect on GDx parameters of variability in locating the peri-papillary contour line.

Also, no study exists which compares HRT and GDx parameters from the same group of patients, which have been outlined by the same group of operators.

Method

One examination was randomly selected for 60 people (20 normal individuals, 20 high risk and 20 glaucoma patients) from the Glaucoma Imaging Study database. Normal individuals were taken from the database of control patients, whereas high risk patients were individuals selected from the longitudinal high risk category, who had no evidence of visual field loss. Patients with glaucoma were taken from members of the cross-sectional and longitudinal glaucoma groups who had repeatable visual field loss.

For each examination, one image with the HRT and GDx was randomly chosen. Images had already been considered to be of adequate quality at the point of image capture; no further quality-control measures were established for this part of the study. The visual field MD for that visit was noted.

Five observers were chosen. Three (AJK, DBH and RAH) are optometrists with clinical vision science backgrounds, one (JK) is a consultant ophthalmologist and the other (STP) is an optometrist with an academic background and many years' clinical experience.

To investigate interobserver variability, all five operators outlined the 60 discs on both the HRT and GDx; no time constraints were set for the task. Three of the experimenters (AJK, DBH and JK) repeated the exercise in a second session >1 week after the first, in order to determine their intraobserver variability.

All observers were masked to the diagnosis of the patients and also to the outlines drawn by their colleagues. They were all given appropriate training and shown an identical set of 'sample' images.

A sixth observer (AKJ) graded all the images on the HRT and GDx. Each was graded for image quality (which combined centration, focus, evenness of illumination and eye movement during image capture) and ease of identification of the disc margin (the ease with which the disc margin could be differentiated, as seen on the VDU). Each was categorised according to a three-level score of good, moderate or poor for image quality and whether it was easy, intermediate or difficult to determine the disc margin.

Discs were also assessed for the presence of significant PPA. PPA is divided into two zones: central zone β and peripheral zone α. An eye was considered to have significant PPA if the combined extent of zones α and β was greater than half of the disc diameter.

From each instrument, six parameters were considered for analysis. These were selected on account of their previously documented

TABLE 55 *Interpretation of ICC*

ICC	Interpretation
–1.00 to 0.00	Poor agreement
0.01 to 0.20	Slight agreement
0.21 to 0.40	Fair agreement
0.41 to 0.60	Moderate agreement
0.61 to 0.80	Substantial agreement
0.81 to 0.99	Almost perfect agreement
1.00	Perfect agreement

superiority in the diagnosis and monitoring of glaucoma.

HRT

This study was performed prior to the upgrading of the HRT to the new software with its improved method of identifying the disc margin. All researchers participating in the study were given the option of 'drawing' the disc outlines with either a 'conventional' computer mouse (IntelliMouse Explorer, Microsoft) or a pen-style mouse (Tablet, Wacom). Instructions were given to draw along the inner aspect of Elschnig's ring and also not to follow automatically the 'dark–light' boundary, but to consider the overall shape of the disc.

The study used images whose 'aligned' component had been stored. This allowed a facility in the HRT software to be used whereby any of the 32 'sections' can be selected to aid identification of the disc margin. This is especially helpful in small, healthy discs, which often have indistinct margins. Most previous studies only used HRT mean images, which do not offer this option.

The use of stereophotographs has been proposed in order to enhance differentiation of the disc margin.[247,248] However, this aid was not incorporated in this study, the aim of which was to reproduce a realistic clinical environment, where such photographs are routinely not available.

The six parameters selected for analysis were C:D area ratio, cup area, cup volume, rim area, rim volume and cup shape measure. The Mikelberg DFA[87] was also recorded.

GDx

Operators were instructed to adjust the preformed ellipse such that it overlaid the inner aspect of Elschnig's ring. When outlining the discs of patients with PPA, operators were instructed to disobey this rule in order to prevent the measurement ellipse overlapping the atrophic area. This strategy is in accordance with established guidelines.[102]

The six parameters selected were the number, symmetry, superior:nasal ratio, average thickness, ellipse modulation and superior average. Diagnostic categorisation for the number was taken from the manufacturer's guidelines of normal (≤30), borderline (31–70) and glaucoma (71–100).

Analysis

Following identification of the disc margins, parameters were exported from the instruments, read into spreadsheets (Microsoft Excel 2000) and analysed with SPSS (version 10.0.7). Agreement for each parameter within and between operators was calculated using ICCS.[249] This statistic was selected in preference to correlation because it measures agreement, whereas correlation identifies the presence of a linear relationship. Also, ICC allows for variation between subjects. It produces a result that lies between –1.00 and 1.00; guidelines for suggested interpretation are shown in *Table 55*.[240] It is widely accepted that ICC ≥ 0.90 indicates high reproducibility of measurements.

For both the HRT and GDx, the consistency of diagnostic classification (produced by the Mikelberg DFA and interpretation of the number, respectively) was determined. The proportion of subjects for whom there was not perfect agreement between the five observers was calculated.

The study aimed to ascertain whether any other factors have a bearing on variability in reference ring placement. The areas of interest were patient's diagnosis, the presence of PPA, image quality and ease of identification of the disc margin. To explore each of these, the experimental group was stratified into the relevant categories and the analysis repeated.

Bland and Altman plots were constructed to display intra-observer repeatability for each operator.

Results

The mean age of the experimental group was 63.42 ± 9.91 years. The visual field MD for each group was normal individuals –0.66 ± 1.56 dB; high risk –0.47 ± 1.42 dB; and glaucoma –5.24 ± 6.09 dB. The results of the independent observer's assessment of the images are detailed in

TABLE 56 *Independent observer's assessment of images*

Category	Number of subjects							
	Quality of HRT images[a]		HRT distinction of disc margin[b]		Quality of GDx images[c]		GDx distinction of disc margin[d]	
Normal	Poor	3	Difficult	15	Poor	2	Difficult	6
	Moderate	12	Intermediate	4	Moderate	7	Intermediate	9
	Good	5	Easy	1	Good	11	Easy	5
High risk	Poor	3	Difficult	4	Poor	3	Difficult	0
	Moderate	9	Intermediate	6	Moderate	8	Intermediate	5
	Good	8	Easy	10	Good	9	Easy	15
Glaucoma	Poor	4	Difficult	2	Poor	4	Difficult	4
	Moderate	8	Intermediate	5	Moderate	7	Intermediate	6
	Good	8	Easy	13	Good	9	Easy	10
Total	Poor	10	Difficult	21	Poor	9	Difficult	10
	Moderate	29	Intermediate	15	Moderate	22	Intermediate	20
	Good	21	Easy	24	Good	29	Easy	30

[a] $p = 0.022$
[b] $p \ll 0.001$
[c] $p = 0.025$
[d] $p = 0.0002$ (χ^2 test)

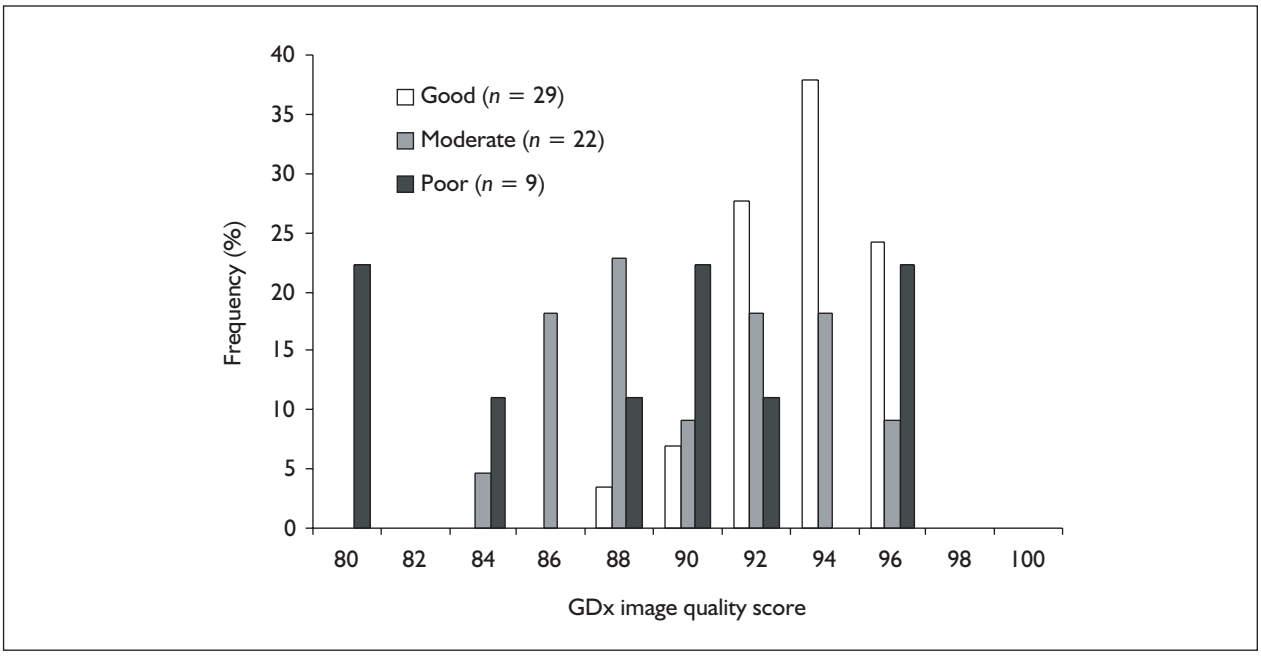


FIGURE 48 *Distribution of GDx image quality score. Data are grouped by the independent observer's ranking of image quality.*

Table 56. Thirteen patients were noted to have significant PPA; of these, six were normal individuals, one was a high risk eye and five were glaucoma patients – this distribution was not statistically significant ($p = 0.191$; χ^2 test). There was exact agreement between the two instruments regarding the presence of PPA.

Figure 48 shows the relationship between the GDx image quality score (maximum 100) and AKJ's subjective assessment of image quality. It is apparent that some images were judged as 'poor' quality, but achieved a high numerical value, which is a not uncommon occurrence with the GDx. No similar comparison can be made for the HRT, as the instrument does not produce a quantitative assessment for single images.

With both instruments, there is a statistically significant relationship between the ease with

TABLE 57 *HRT parameter values: data shown as mean (standard deviation)*

Observer (initials)	HRT parameter					
	C:D area ratio	Cup area (mm^2)	Cup volume (mm^3)	Rim area (mm^2)	Rim volume (mm^3)	Cup shape measure
1 (AJK)	0.357 (0.224)	0.768 (0.600)	0.220 (0.255)	1.223 (0.389)	0.287 (0.155)	–0.166 (0.091)
2 (DBH)	0.362 (0.223)	0.762 (0.603)	0.219 (0.267)	1.158 (0.348)	0.270 (0.141)	–0.163 (0.091)
3 (JK)	0.779 (0.623)	1.888 (0.536)	0.334 (0.100)	0.376 (0.238)	0.230 (0.301)	0.623 (0.257)
4 (RAH)	0.365 (0.221)	0.752 (0.582)	0.213 (0.258)	1.135 (0.335)	0.263 (0.126)	–0.161 (0.092)
5 (STP)	0.374 (0.228)	0.786 (0.609)	0.230 (0.282)	1.167 (0.410)	0.274 (0.151)	–0.156 (0.090)

TABLE 58 *GDx parameter values. data shown as mean (standard deviation)*

Observer (Initials)	GDx parameter					
	The number	Symmetry	Superior nasal ratio	Average thickness (μm)	Ellipse modulation (μm)	Superior average (μm)
1 (AJK)	26.783 (19.168)	1.002 (0.138)	1.884 (0.355)	67.917 (11.913)	2.122 (0.691)	76.029 (13.998)
2 (DBH)	27.317 (18.72)	1.006 (0.137)	1.872 (0.329)	67.950 (11.942)	2.153 (0.715)	76.185 (14.505)
3 (JK)	31.383 (20.620)	0.998 (0.137)	1.825 (0.319)	67.783 (11.888)	2.126 (0.685)	75.689 (14.241)
4 (RAH)	26.650 (18.742)	1.002 (0.138)	1.875 (0.336)	67.917 (11.814)	2.126 (0.662)	76.065 (14.250)
5 (STP)	27.483 (19.613)	1.004 (0.134)	1.871 (0.333)	68.000 (11.876)	2.146 (0.781)	75.973 (14.157)

TABLE 59 *Interobserver variability for HRT and GDx*

HRT		GDx	
Parameter	**ICC**	**Parameter**	**ICC**
Cup to disc area ratio	0.814	Number	0.847
Cup area	0.840	Symmetry	0.902
Cup volume	0.839	Superior:nasal ratio	0.914
Rim area	0.761	Average thickness	0.988
Rim volume	0.731	Ellipse modulation	0.820
Cup shape measure	0.826	Superior average	0.981
MEAN	0.802	MEAN	0.909

which the disc margin can be identified and diagnosis, with glaucomatous eyes having discs with margins that were rated easier to define.

Interobserver variability

The mean and standard deviation for HRT and GDx parameters for each observer are shown in *Tables 57* and *58* and the ICCs in *Table 59*. All parameters show at least 'substantial agreement' (i.e. ICC ≥ 0.61). Four HRT parameters and all GDx parameters have 'almost perfect' agreement (ICC ≥ 0.81).

This study is particularly interested in the effect of variability in contour line placement on diagnostic categorisation by the two instruments. There was inconsistency in the diagnostic category from the HRT for 20 patients (33%) and from the GDx for 15 patients (25%), which occurred to different patients for each instrument. There was no tendency for a group of patients with a particular diagnosis to be misclassified (as shown on p. 103). No particular observer was responsible for the diagnostic discrepancies with either the HRT or GDx.

Diagnostic category for the HRT was taken from the Mikelberg DFA. The mean (and standard deviation) for each group was 1.87 (1.55), 1.18 (1.76) and –1.29 (1.89) for normal, high risk and

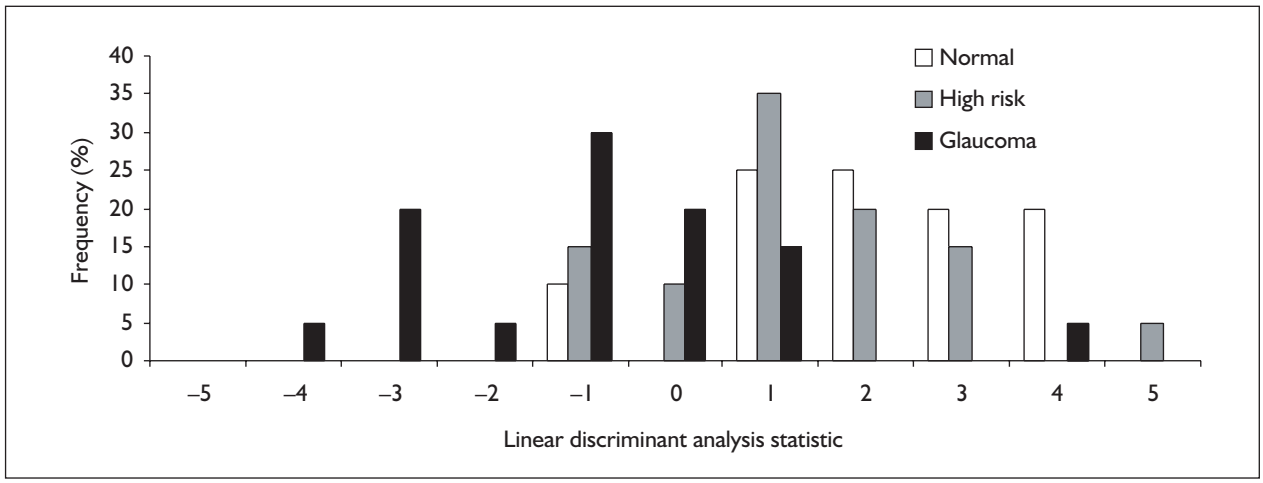


FIGURE 49 *Distribution of mean Mikelberg discriminant function statistic. Values shown are the mean from all five observers.*

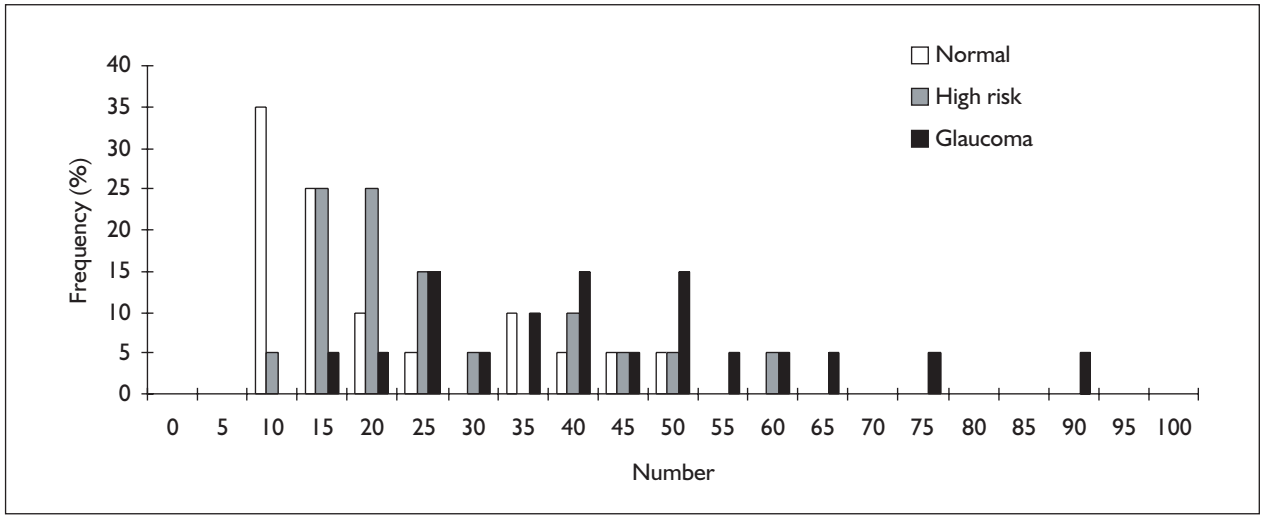


FIGURE 50 *Distribution of Mean GDx number. Values shown are the mean from all five observers*

TABLE 60 *Effect of patient's diagnosis on interobserver variability with HRT*

	ICC		
Parameter	**Normal**	**High risk**	**Glaucoma**
C:D area ratio	0.661	0.728	0.858
Cup area	0.661	0.827	0.813
Cup volume	0.445	0.636	0.883
Rim area	0.461	0.852	0.798
Rim volume	0.791	0.772	0.653
Cup shape measure	0.721	0.820	0.718
Mean	0.623	0.773	0.787

TABLE 61 *Effect of patient's diagnosis on interobserver variability with GDx*

	ICC		
Parameter	**Normal**	**High risk**	**Glaucoma**
Number	0.763	0.717	0.924
Symmetry	0.804	0.979	0.888
Superior:nasal ratio	0.918	0.887	0.861
Average thickness	0.980	0.990	0.991
Ellipse modulation	0.718	0.980	0.806
Superior average	0.947	0.992	0.992
Mean	0.855	0.924	0.910

TABLE 62 *Effect of patient's diagnosis on inconsistent diagnostic categorisation with HRT and GDx*

	Inconsistent diagnostic category		
Instrument	**Normal**	**High risk**	**Glaucoma**
HRT[a]	7	6	7
GDx[b]	4	3	8

[a] $p = 0.407$ (χ^2 test).
[b] $p = 0.247$ (χ^2 test).

glaucoma, respectively. These values are the mean of results from all five observers and are shown graphically in *Figure 49*. Interpretation of these results categorises 37 patients as 'normal' and 23 as having glaucoma.

Diagnostic results for the GDx were taken from interpretation of the GDx number. The mean (and standard deviation) for each group was 18.70 (12.64), 23.93 (13.77) and 41.14 (19.52) for normal, high risk and glaucoma, respectively. These data are illustrated in *Figure 50*, where patients are stratified by the diagnostic category used within the study. From the mean value of the number from all five observers, 36 patients were categorised as normal, 22 as borderline and two as glaucomatous. The relatively high frequency of numbers <50 is reflected in the distribution of those who received inconsistent diagnostic category: of the total 15, only four fluctuated between 'borderline' and 'glaucoma', whereas 11 fluctuated between 'normal' and 'borderline'.

Effect of clinical diagnosis

The results for interoperator variability after the experimental group had been stratified by diagnosis (i.e. as normal, high risk or glaucoma) are shown in *Tables 60–62*. Agreement is lowest for normal eyes with both the instruments and highest for glaucoma eyes (HRT) and high risk eyes (GDx). With both instruments, the largest difference in mean ICC for the six parameters was between normal and high risk eyes.

A patient's diagnosis had no effect on the likelihood of receiving diagnostic misclassification with either instrument (*Table 62*).

Effect of PPA

Agreement of HRT and GDx parameters for patients with PPA ($n = 13$) and without PPA ($n = 47$) is given in *Tables 63* and *64*. The presence of significant PPA had opposite effects on variability from the two instruments: for the HRT, all parameters examined had higher ICCs in eyes with PPA, whereas the converse was true for all GDx parameters, except one. There was no relationship between PPA and a tendency to show inconsistent diagnostic categorisation with either instrument (*Table 65*).

Effect of image quality

ICCs for each parameter within the three categories of image quality (good, moderate and poor) for each instrument are shown in *Tables 66* and *67*.

TABLE 63 *Effect of PPA on interobserver variability with HRT*

Parameter	ICC		
	With PPA (*n* = 13)	**Without PPA (*n* = 47)**	**Difference (= with PPA – without PPA)**
C:D area ratio	0.885	0.794	0.091
Cup area	0.891	0.833	0.058
Cup volume	0.915	0.832	0.083
Rim area	0.822	0.735	0.087
Rim volume	0.836	0.673	0.163
Cup shape measure	0.943	0.820	0.123
Mean	0.882	0.781	0.101

TABLE 64 *Effect of PPA on interobserver variability with GDx*

Parameter	ICC		
	With PPA (*n* = 13)	**Without PPA (*n* = 47)**	**Difference (= with PPA – without PPA)**
Number	0.649	0.867	–0.218
Symmetry	0.863	0.931	–0.068
Superior:nasal ratio	0.945	0.900	0.045
Average thickness	0.979	0.991	–0.012
Ellipse modulation	0.799	0.833	–0.034
Superior average	0.967	0.990	–0.023
Mean	0.867	0.919	–0.052

TABLE 65 *Effect of PPA on diagnostic categorisation with HRT and GDx*

Instrument	ICC	
	With PPA (*n* = 13)	**Without PPA (*n* = 47)**
HRT[a]	4	16
GDx[b]	5	10

[a] $p = 0.856$ (χ^2 test).
[b] $p = 0.273$ (χ^2 test).

TABLE 66 *Effect of image quality on interobserver variability with HRT*

Parameter	ICC		
	Good (*n* = 21)	**Moderate (*n* = 29)**	**Poor (*n* = 10)**
C:D area ratio	0.784	0.787	0.922
Cup area	0.838	0.816	0.936
Cup volume	0.885	0.787	0.667
Rim area	0.813	0.663	0.841
Rim volume	0.826	0.610	0.719
Cup shape measure	0.770	0.872	0.859
Mean	0.819	0.756	0.824

TABLE 67 *Effect of image quality on interobserver variability with GDx*

	ICC		
Parameter	**Good (*n* = 29)**	**Moderate (*n* = 22)**	**Poor (*n* = 9)**
Number	0.895	0.552	0.979
Symmetry	0.913	0.869	0.967
Superior:nasal ratio	0.961	0.799	0.980
Average thickness	0.984	0.986	0.999
Ellipse modulation	0.863	0.675	0.984
Superior average	0.978	0.977	0.996
Mean	0.932	0.810	0.984

TABLE 68 *Effect of image quality on diagnostic categorisation with HRT and GDx*

	Inconsistent diagnostic category		
Instrument	**Good**	**Moderate**	**Poor**
HRT[a]	7	9	4
GDx[b]	5	9	1

[a] $p = 0.914$ (χ^2 test).
[b] $p = 0.199$ (χ^2 test).

There was no definite relationship between ICC and image quality for either instrument. Both the HRT and GDx showed the lowest ICCs for the 'moderate' category and highest for 'poor' quality images. There was no statistically significant relationship for those who showed inconsistency in their diagnostic categorisation (*Table 68*).

Effect of 'ease' of identifying disc margin

Interobserver variability, subsequent to stratifying the experimental group for the ease with which the disc margin could be identified, is detailed in *Tables 69* and *70*. With both instruments, the highest mean ICC was for the group of patients with discs in the 'intermediate' category and the lowest for discs with margins that were difficult to differentiate. However, the difference was most marked for the HRT, with a mean ICC for the 'difficult' category that is interpreted as having only 'fair agreement'. The range of ICCs for the different parameters was greatest for 'difficult' discs with both instruments.

The likelihood of eyes in a particular category being more prone to receiving a varying diagnostic result does not reach statistical significance (*Table 71*).

Intraobserver variability

For each of the three observers' two sessions, the mean and standard deviation for the parameters and ICCs for intraobserver variability are given in *Tables 72–75*.

From the HRT data, observers 1 and 2 each produced parameters that have 'almost perfect' agreement, compared with observer 3's 'substantial' agreement. The difference was not so marked for the GDx, with all three showing excellent agreement. Observer 3 had the lowest mean ICC for both the HRT and GDx.

Overall, parameter values from both instruments were largely similar for both AJK and DBH, with their distributions appearing alike. Bland and Altman plots (the ordinate scales on graphs of corresponding parameters have been matched between operators to facilitate comparison) are shown in *Figures 51* and *52* (HRT parameters from AJK and JK, respectively) and *Figures 53* and *54* (GDx parameters from DBH and JK, respectively); these combinations were selected on the grounds of having the highest and lowest intraobserver variability for each instrument. These plots show no relationship between values and variability for most parameters. However, the GDx parameter superior:nasal ratio appears to show greater variability at higher values.

To compare consistency of diagnostic category, the kappa statistic for each of the observers was 0.897,

TABLE 69 *Effect of 'ease' of identification of optic disc margin on interobserver variability with HRT*

	ICC		
Parameter	**Easy (*n* = 24)**	**Intermediate (*n* = 15)**	**Difficult (*n* = 21)**
C:D area ratio	0.726	0.929	0.383
Cup area	0.850	0.932	0.310
Cup volume	0.888	0.870	0.180
Rim area	0.757	0.924	0.373
Rim volume	0.622	0.895	0.401
Cup shape measure	0.795	0.866	0.657
Mean	0.773	0.903	0.384

TABLE 70 *Effect of 'ease' of identification of optic disc margin on interobserver variability with GDx*

	ICC		
Parameter	**Easy (*n* = 30)**	**Intermediate (*n* = 20)**	**Difficult (*n* = 10)**
Number	0.904	0.768	0.667
Symmetry	0.883	0.983	0.763
Superior:nasal ratio	0.892	0.913	0.987
Average thickness	0.988	0.994	0.965
Ellipse modulation	0.802	0.917	0.779
Superior average	0.975	0.994	0.984
Mean	0.907	0.928	0.858

TABLE 71 *Effect of 'ease' of identification of optic disc margin on diagnostic categorisation with HRT and GDx*

	Inconsistent diagnostic category		
Instrument	**Easy**	**Intermediate**	**Difficult**
HRT[a]	8	4	8
GDx[b]	5	7	3

[a] $p = 0.418$ (χ^2 test).
[b] $p = 0.420$ (χ^2 test).

TABLE 72 *HRT parameter values: data shown as mean (standard deviation)*

	HRT parameter					
Observer and session number	**C:D area ratio**	**Cup area (mm^2)**	**Cup volume (mm^3)**	**Rim area (mm^2)**	**Rim volume (mm^3)**	**Cup shape measure**
AJK 1	0.357 (0.224)	0.768 (0.595)	0.220 (0.255)	1.222 (0.389)	0.287 (0.155)	–0.166 (0.091)
AJK 2	0.357 (0.227)	0.768 (0.619)	0.226 (0.278)	1.195 (0.374)	0.278 (0.148)	–0.165 (0.089)
DBH 1	0.362 (0.223)	0.762 (0.603)	0.219 (0.267)	1.158 (0.348)	0.269 (0.141)	–0.163 (0.091)
DBH 2	0.358 (0.222)	0.738 (0.613)	0.216 (0.290)	1.107 (0.324)	0.260 (0.129)	–0.157 (0.089)
JK 1	0.779 (0.623)	1.888 (0.536)	0.334 (0.100)	0.376 (0.238)	0.230 (0.301)	0.623 (0.257)
JK 2	0.386 (0.234)	0.790 (0.613)	0.227 (0.268)	1.070 (0.352)	0.246 (0.142)	-0.156 (0.092)

TABLE 73 *GDx parameter values: data shown as mean (standard deviation)*

Observer and session number	GDx parameter					
	Number	Symmetry	Superior: nasal ratio	Average thickness (μm)	Ellipse modulation (average μm)	Superior average (μm)
AJK 1	26.783 (19.168)	1.002 (0.138)	1.884 (0.355)	67.917 (11.913)	2.122 (0.691)	76.029 (13.998)
AJK 2	26.583 (18.832)	1.002 (0.138)	1.888 (0.355)	67.900 (14.040)	2.121 (0.687)	76.148 (14.155)
DBH 1	27.317 (18.72)	1.006 (0.137)	1.872 (0.329)	67.950 (11.942)	2.153 (0.715)	76.185 (14.505)
DBH 2	28.817 (19.719	1.006 (0.138)	1.870 (0.333)	67.850 (11.836)	2.143 (0.788)	76.037 (14.240)
JK 1	31.383 (20.620)	0.998 (0.137)	1.825 (0.319)	67.783 (11.888)	2.126 (0.685)	75.689 (14.241)
JK 2	32.317 (20.692)	1.004 (0.138)	1.837 (0.322)	67.633 (11.987)	2.126 (0.663)	75.070 (14.002)

TABLE 74 *Intraobserver variability for HRT*

Parameter	ICC		
	Observer 1 (AJK)	Observer 2 (DBH)	Observer 3 (JK)
C:D area ratio	0.985	0.982	0.557
Cup area	0.926	0.896	0.666
Cup volume	0.980	0.978	0.697
Rim area	0.987	0.978	0.697
Rim volume	0.920	0.853	0.730
Cup shape measure	0.988	0.981	0.772
Mean	0.964	0.945	0.687

TABLE 75 *Intraobserver variability for GDx*

Parameter	ICC		
	Observer 1 (AJK)	Observer 2 (DBH)	Observer 3 (JK)
Number	0.989	0.976	0.907
Symmetry	0.992	0.987	0.941
Superior:nasal ratio	0.953	0.990	0.982
Average thickness	0.994	0.994	0.986
Ellipse modulation	0.996	0.997	0.994
Superior average	0.991	0.999	0.999
Mean	0.986	0.991	0.968

0.798 and 0.564 with the HRT and 0.827, 1.00 and 0.779 with the GDx.

Effect of clinical diagnosis

Table 76 shows the effect of patient's diagnosis on repeatability for the HRT and GDx. Mean ICC values derived from all six parameters are given.

For HRT parameters, each of the observers showed a different pattern of variation with diagnosis. Observers 1 and 2 agreed with the trend found from the interobserver study, where HRT parameters from normal eyes showed the highest variability. This was also the case for the GDx although, for observers 1 and 2, the differences were small.

Effect of PPA

The effect of PPA on intraobserver variability with the HRT and GDx is shown in *Table 77*. The trend is the same as for interobserver variability, albeit the magnitude of the difference was much lower.

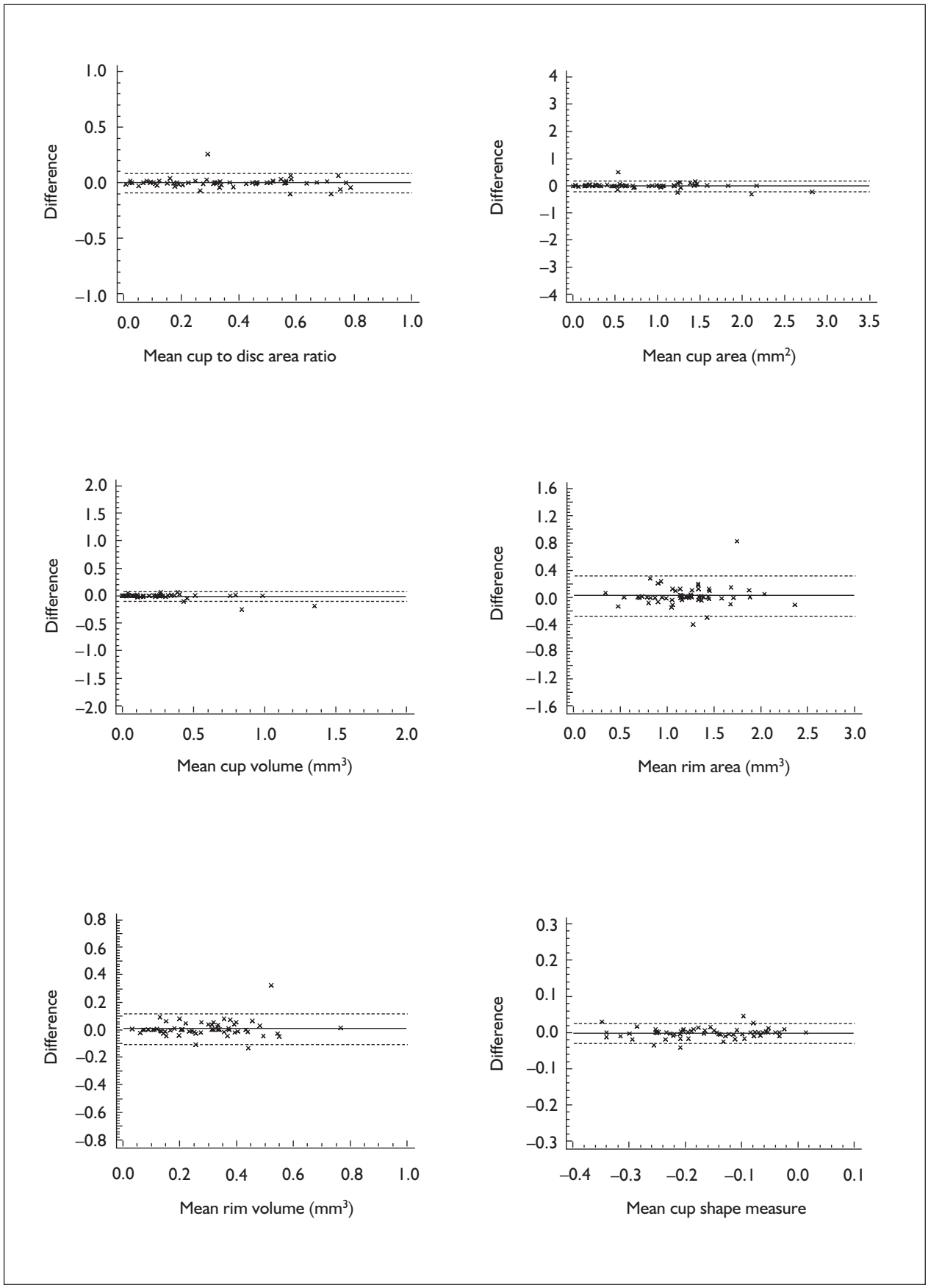


FIGURE 51 *Bland and Altman plots for HRT parameters (observer 1, AJK). Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

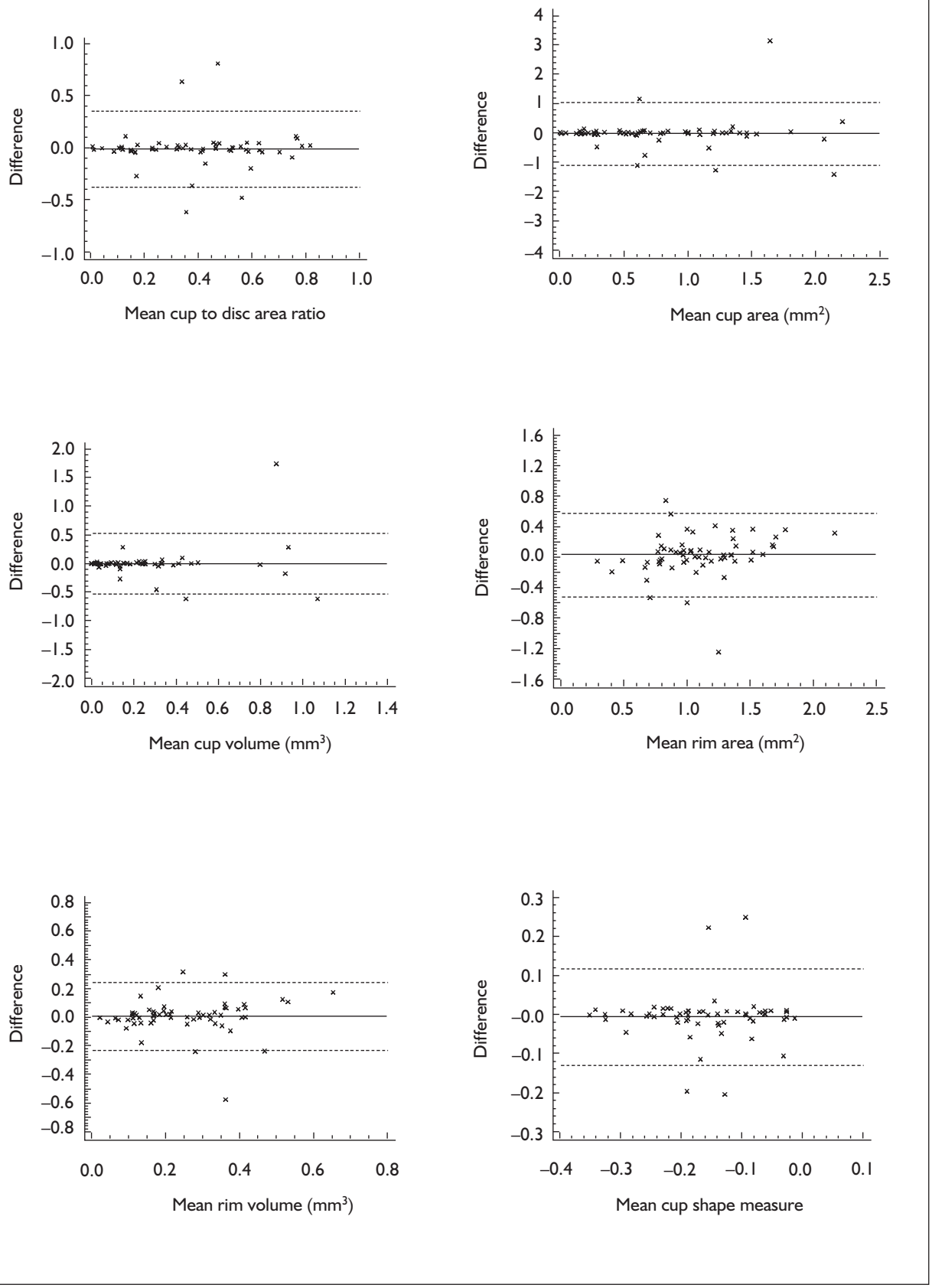


FIGURE 52 *Bland and Altman plots for HRT parameters (observer 3, JK). Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

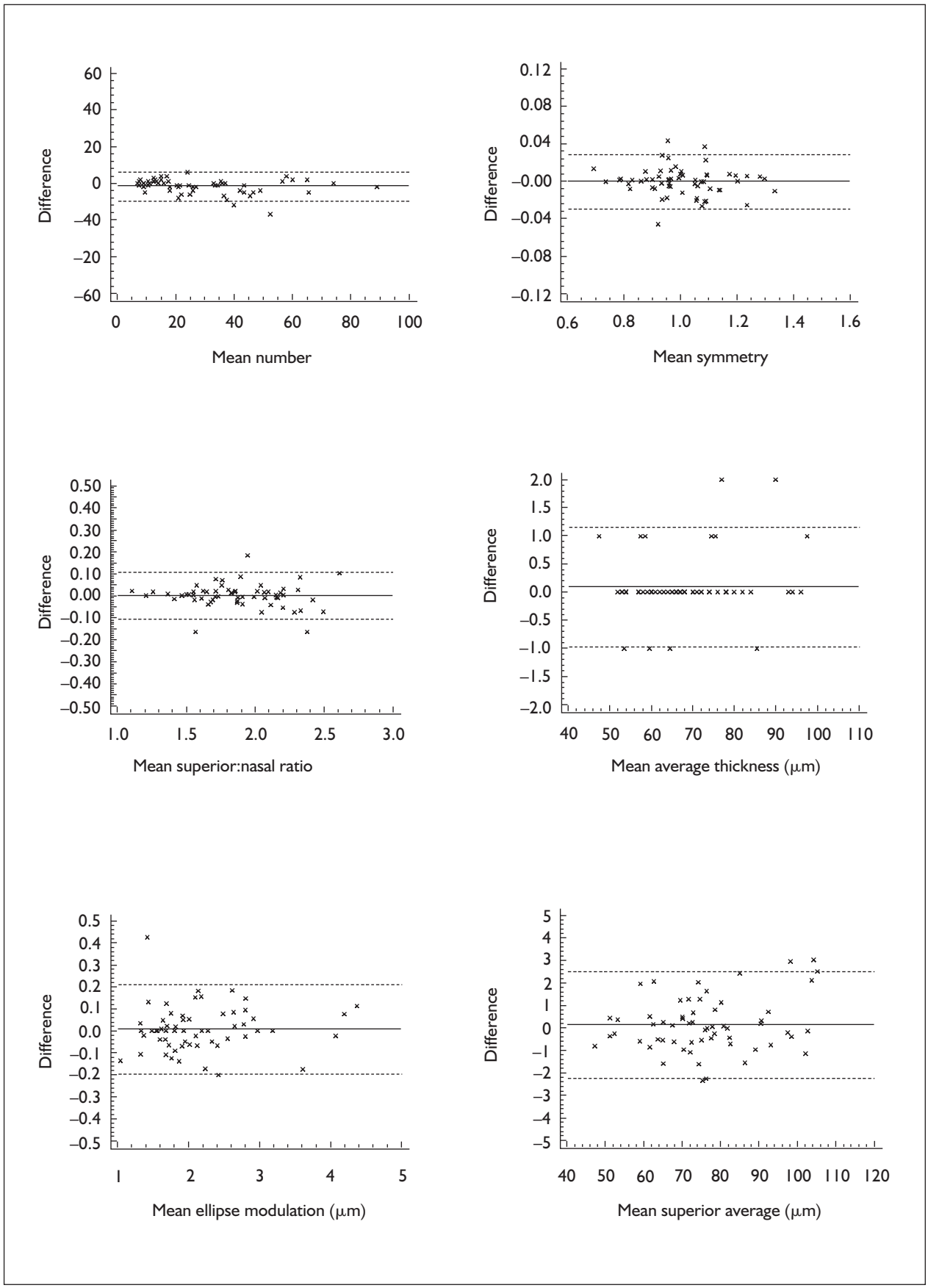


FIGURE 53 *Bland and Altman plots for GDx parameters (observer 2, DBH). Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

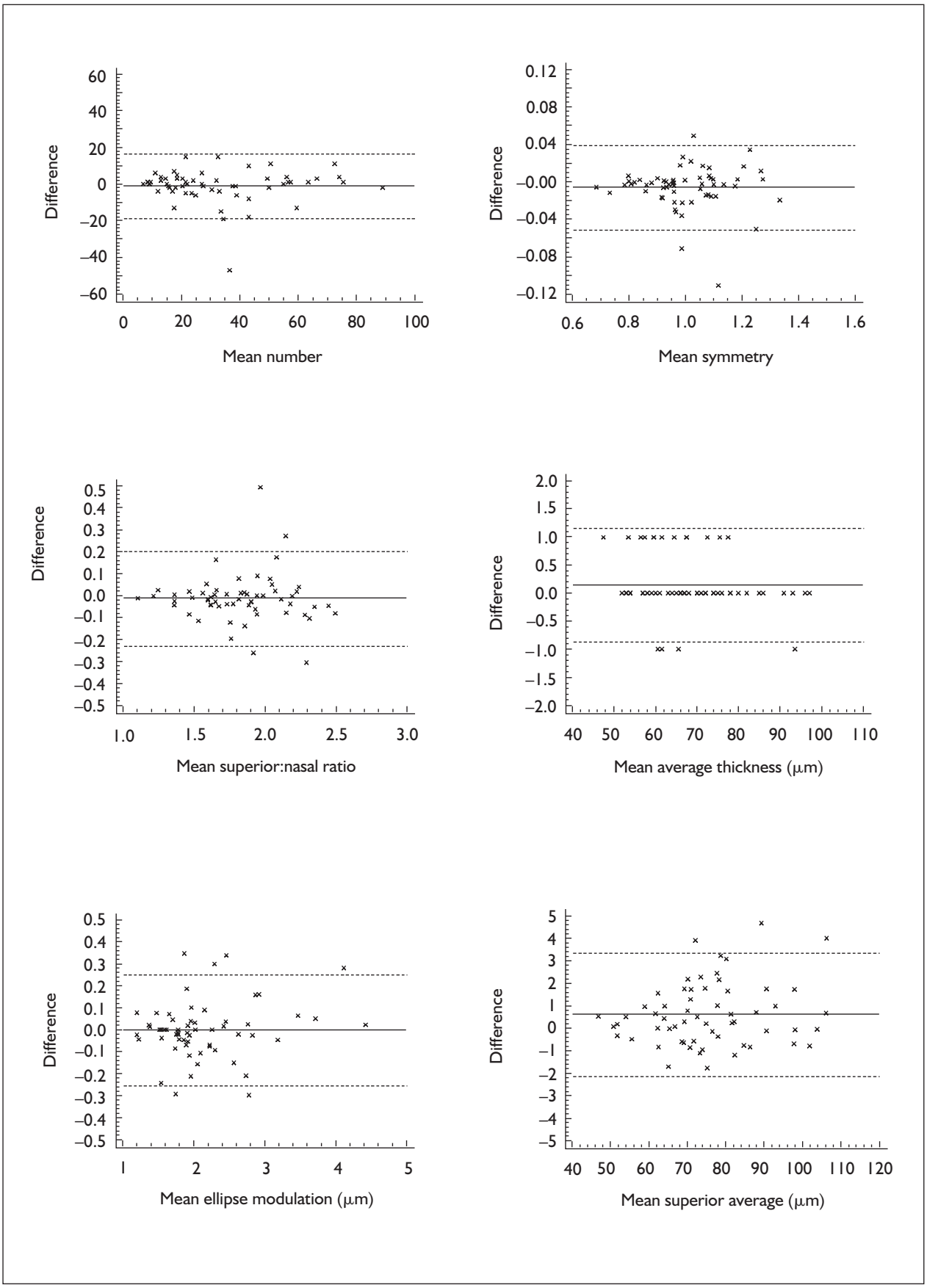


FIGURE 54 *Bland and Altman plots for GDx parameters (observer 3, JK). Solid line = mean; dotted line = mean ± 1.96 standard deviations.*

TABLE 76 *Effect of patient's diagnosis on intraobserver variability with HRT and GDx*

	Mean ICC		
Observer	**Normal**	**High risk**	**Glaucoma**
HRT			
1 (AJK)	0.914	0.966	0.965
2 (DBH)	0.828	0.927	0.980
3 (JK)	0.542	0.889	0.434
GDx			
1 (AJK)	0.976	0.993	0.984
2 (DBH)	0.986	0.989	0.989
3 (JK)	0.907	0.975	0.986

TABLE 77 *Effect of PPA on intraobserver variability with HRT and GDx*

	Mean ICC		
Observer	**With PPA (*n* = 13)**	**Without PPA (*n* = 47)**	**Difference (= with PPA – without PPA)**
HRT			
1 (AJK)	0.973	0.960	0.013
2 (DBH)	0.977	0.934	0.043
3 (JK)	0.726	0.676	0.050
GDx			
1 (AJK)	0.977	0.989	–0.012
2 (DBH)	0.982	0.992	–0.010
3 (JK)	0.965	0.968	–0.003

TABLE 78 *Effect of image quality on intraobserver variability with HRT and GDx*

	Mean ICC		
Observer	**Good**	**Moderate**	**Poor**
HRT			
1 (AJK)	0.969	0.953	0.972
2 (DBH)	0.976	0.897	0.962
3 (JK)	0.582	0.760	0.819
GDx			
1 (AJK)	0.990	0.978	0.991
2 (DBH)	0.990	0.984	0.995
3 (JK)	0.954	0.974	0.982

Effect of image quality

HRT image quality affected each observer differently (with two producing their highest reliability for poor-quality images), whereas GDx quality produced a more consistent effect, with all three observers showing the greatest reliability for images that had been judged to have poor quality (although by only a very small margin in two cases) – see *Table 78*.

Effect of 'ease' of identifying disc margin

For HRT parameters, observers 1 and 2 show the same pattern as for interobserver variability, that is, highest repeatability for 'intermediate' discs and lowest for discs with margins that were difficult to differentiate; by far the larger difference in agreement was between 'moderate' and 'difficult' discs. With the GDx, the same two observers showed highest intraobserver agreement for easier to define disc margins and lowest for more difficult discs. From *Table 79*, it is clear that observer 3 was affected differently by this variable.

Reanalysis of interobserver variability

From *Table 72*, it is evident that JK's first session of outlining the HRT disc margins produced outlying

TABLE 79 *Effect of 'ease' of identification of disc margin on intraobserver variability with HRT and GDx*

	ICC		
Observer	**Easy**	**Intermediate**	**Difficult**
HRT			
1 (AJK)	0.956	0.987	0.876
2 (DBH)	0.950	0.975	0.819
3 (JK)	0.441	0.877	0.428
GDx			
1 (AJK)	0.996	0.986	0.974
2 (DBH)	0.993	0.987	0.981
3 (JK)	0.976	0.938	0.991

TABLE 80 *Reanalysis of interobserver variability for HRT*

Parameter	ICC
C:D area ratio	0.742
Cup area	0.729
Cup volume	0.722
Rim area	0.741
Rim volume	0.704
Cup shape measure	0.763
Mean	0.734

values. These data had been used to calculate interobserver variability as discussed in the section 'Intraobserver variability' (p. 101). In order to investigate whether this had adversely affected the ICCs, the analysis was repeated, replacing JK's first data set with his second; this yielded a mean ICC of 0.734 (compared with 0.802 previously). The full results are detailed in *Table 80*. The number of individuals who showed inconsistent diagnostic classification actually increased from 20 to 21.

Discussion

This study has investigated the effects of inter- and intraobserver variability on HRT and GDx parameters. Its results are highly relevant to the overall conclusions of the Glaucoma Imaging Study, since the utility of the instruments is dependent on them having high repeatability. The analysis has investigated the effects of factors such as diagnosis and the presence of PPA on repeatability – these variables are of particular importance to a population of glaucoma patients.

Data presented in this chapter were collected and analysed fairly early in the study. Therefore, on several counts, the results are not applicable to the most contemporary versions of the equipment (HRT II and GDx VCC).

- The new HRT Windows-based software incorporates a modified system for identifying the optic nerve head margin (as shown in *Figure 3*), which is far easier to use than its predecessor (which was used in this section of the study).
- Data presented in this chapter used HRT images whose 'aligned' image had been saved, in order to facilitate a system where individual planes (of the original 32) could be viewed to enhance identification of the disc margin. This option is not available with the new software.
- The new HRT software does not incorporate the system described above, but uses a pseudo-three-dimensional display on which the outline is overlaid. The effect of this display on inter- and intraobserver variability needs to be established by further research.
- From Chapter 7, optimum diagnostic precision with the HRT was achieved by the MRA.[95] This algorithm was not incorporated in this substudy since, at the time of analysis, we could not access MRA results.
- The GDx VCC uses a completely different form of display (integral screen versus VDU) and a modified form of identification of the disc margin on a reduced-size image. The impact of these differences requires further investigation.

Overall, this study found interobserver agreement of HRT and GDx parameters to be very good. For the former, ICCs of the six selected parameters ranged from 0.731 to 0.840 (mean 0.802); a similar study, which also assessed variability between five observers for six HRT stereometric parameters, found a mean of 0.803,[237] albeit for different parameters. Agreement for GDx parameters was even higher, ranging from 0.847 to 0.988 (mean 0.909); there are no other published data with which the results can be compared.

It is important to recognise that even if an instrument shows very high levels of agreement, it is of no clinical use if its diagnostic accuracy is low. The issue of the scale of a measurement is highlighted by Bailey and colleagues, who note that if concordance between examinations is high, the scale might be too coarse, which would limit the ability of the test to detect change.[242]

The levels of intraobserver agreement found in this study varied between experimenters: AJK and DBH produced parameters that were very repeatable with both the HRT and GDx, achieving mean ICCs of 0.964 and 0.945 (HRT) and 0.986 and 0.991 (GDx). The third researcher (JK) showed higher variability, with mean ICCs of 0.687 (HRT) and 0.968 (GDx). This is possibly attributable to the fact that JK is the least familiar of the three observers with the imaging techniques used in the study.

Comparing the repeatability of the different parameters, our results differ slightly from other findings that show cup shape measure to have low variability with reference ring placement.[237,248] In the interobserver study, cup shape measure ranked third best, but in the intraobserver study, AJK and DBH showed it to be the most repeatable and second most repeatable parameter, respectively. This parameter is considered particularly important, since it has been previously shown to have good sensitivity and specificity for diagnosing glaucoma and, compared with the other parameters, it is considered suitable for monitoring progression because it is independent of change in the reference plane.[87]

Between the two studies of inter- and intraobserver variability, there was no pattern to the variability ranking of HRT parameters; for example, cup area had the highest repeatability of all six parameters in the interobserver study, although for AJK and DBH in the intraobserver study it ranked fifth. Of interest, the distribution of HRT ICC values for these two researchers was most similar within the study; this is likely to be a reflection of their greater experience with the equipment, compared with the other observers.

The GDx, similarly, showed no strong trend in repeatability of parameters. The number tended to have a low ranking in both parts of the study, whereas ellipse modulation and superior average showed higher levels of repeatability amongst the parameters examined. The susceptibility of the number is in accordance with warnings made by the manufacturers that it is sensitive to correct placement of the reference ring.[105]

A particular aim of this research was to investigate the effect of variability in the placement of the contour line on diagnostic categorisation by the instruments. A surprisingly high proportion of patients showed inconsistent diagnostic classification with the HRT Mikelberg DFA (33%) and GDx (25%). This finding is unexpected within the context of the excellent agreement of both HRT and GDx parameters between and within operators. The distributions of data for the Mikelberg DFA and GDx number show the susceptibility of many patients to misclassification owing to proximity of their result to the cut-off values for each category. This occurrence could be due to selection bias, that is, the inclusion of a disproportionate number of 'borderline' patients, although the study aimed to prevent such bias by selecting 20 patients from the normal, borderline and glaucoma groups. The mean visual field MD of the glaucoma group was –5.24 ± 6.09 dB (range –28.04 to 1.04 dB), so possibly the tendency of this group towards patients with mild to moderate visual field defects was associated with their being more prone to misclassification. However, the group of patients considered at 'high risk' of developing glaucoma who, theoretically, might be expected to lie within the borderline range did not, in fact, show a greater frequency of inconsistent diagnosis than the glaucoma patients and normal control individuals.

Even if the experimental population was biased towards borderline cases, it suggests a definite weakness of the systems, since these are the patients for whom examination with these instruments is particularly indicated.

The prevalence of inconsistent diagnostic classification was derived from the proportion of subjects for whom there was not perfect agreement between the five observers. The resulting rates are, clearly, dependent on the number of observers. The analysis could have been enhanced by detailing the mean number of pair-wise disagreements.

An incidental, but very interesting, observation is the large extent of overlap of the normal, high risk and glaucoma groups in their values of the HRT discriminant analysis function and the GDx number. This finding has been widely quoted elsewhere,[90,250] but is not investigated further in this chapter, because such issues have been discussed in much greater detail in Chapters 7, 8 and 9.

For the within- and between-observer studies with both the HRT and GDx, the effect of patient's

diagnosis follows a similar pattern, with high risk and glaucoma eyes having higher ICCs than normal eyes; other research has also found greater agreement for glaucoma patients than normal individuals.[245] The higher levels of repeatability found in glaucoma eyes are likely to be a manifestation of the fact that the disc margins are easier to determine in these cases, where the disc margin is less obscured by the overlying retinal nerve fibres. This finding is consistent with our conclusion that eyes with glaucoma were more likely to have disc margins that were easily differentiated. Most of the observers who performed the study concurred with this observation, and agreed that small, healthy discs are the hardest to outline.

Improved differentiation of the disc margin from the surrounding retina is likely to be the reason why eyes with PPA had more repeatable HRT parameters in both parts of the study. The converse was true for the GDx, where higher variability was found in eyes with PPA. This finding is likely to be due to the modified strategy adopted for placement of the GDx reference ring in eyes with PPA. Patients were only classified as having PPA when the extent of the zone was half the disc diameter or larger. For individuals with a minimal extent of atrophy, GDx repeatability may be unaffected as the damaged area is unlikely to interfere with the measurement zone.

This study's finding that eyes with glaucoma were not significantly more likely to have PPA than their normal counterparts differs from that of other researchers[251,252] and is probably a result of the small sample size used and the definition of PPA.

Image quality had an unpredictable affect on both instruments: images considered to be of poor quality showed the highest repeatability in the inter- and intraobserver studies. This is possibly because some the criteria used to judge image quality (e.g. evenness of illumination, centration) have no actual bearing on placement of the reference ring. In comparing the distribution of numerical GDx quality scores with our categorisation, there was poor separation of the groups, especially for those images that were considered poor. However, experience with the instrument has shown that the GDx will, not infrequently, give a high score to an image that is obviously defocussed or shows eye movement during image capture.

All HRT parameters had lower ICCs for eyes with disc margins that were considered difficult to determine. The magnitude of this difference was much greater than any other variable assessed. The mean ICC of HRT parameters in the group with 'difficult' discs was 0.384, which is interpreted as only 'fair' agreement. The difference was nowhere near as marked for intraobserver repeatability of AJK and DBH. A possible explanation is that their familiarity with the equipment allowed them to cope better with more challenging discs. The same pattern was found for the GDx, although the magnitude was far less. The use of stereo-photographs has been shown to improve repeatability of HRT parameters,[248] which might be an especially useful aid for analysing discs with margins that are considered difficult to define.

After completing their sessions of disc outlining, the observers were asked for their comments and preferences. All reported that they performed the GDx far more quickly than the HRT and found the process far easier. Several expressed frustration with the GDx's lack of flexibility when presented with a disc that did not have vertical or horizontal symmetry, in that a 'perfect' outline could not be achieved. The HRT does not have this limitation, but the consensus of opinion was that this benefit is outweighed by the complexity of the procedure.

A clear omission from the study is a consideration of the repeatability of the HRT parameter disc area. This measurement is particularly important because of the role it plays in the MRA.[95] This parameter has been investigated by other authors, who showed it to have the highest repeatability (both between and within observers) of all measurements considered.[246]

A limitation of this study is its choice of observers, all of whom had optometric, academic or ophthalmological backgrounds. This is unlikely to be the case if the instruments were to be used in a clinical environment, where a skilled technician would perform the examination and analysis. Also, no time constraints were imposed on the task, which allowed the participants to be meticulous in their placement of the reference rings; this might not necessarily be attainable in a busy clinical setting.

Alternative strategies have been aimed at developing methods of analysis that are independent of observer input. Provisional research has shown promising results for such systems,[253,254] but more thorough investigation is required prior to their acceptance as established techniques.

Chapter 12

Summary and conclusions

Glaucoma is a common and potentially blinding condition that poses a significant burden to the NHS: patients who are suspected of having the condition may require repeated examinations to confirm their status, and individuals who are diagnosed with glaucoma require life-long follow-up.

Within the glaucoma clinic, the assessment of patients typically involves a triad of examinations: measurement of IOP, ophthalmoscopic examination of the ONH and RNFL and perimetric examination of the visual field. There are several shortcomings to each of these techniques:

- Measurement of IOP is a poor diagnostic predictor of glaucoma since it will not identify patients who have normal tension glaucoma and will mis-classify individuals who have ocular hypertension.
- Perimetry involves a lengthy and tiring examination, which patients universally dislike. The test is subjective and requires an attentive patient with accurate observational skills. Examinations frequently require repeating in cases of unreliable results or when the result was spoiled by a testing artefact.
- Examination of the ONH and RNFL is compromised by high levels of inter- and intraobserver variability and the extent of overlap in appearance of normal and glaucomatous eyes.

There is, therefore, a clear requirement for objective techniques that provide rapid and repeatable examination to assist in the diagnosis and monitoring of glaucoma.

Two approaches that may fulfil this role are the HRT and GDx, which provide topographic examination of the ONH and measurement of RNFL thickness, respectively. Each of the instruments has been widely described in the literature, although there is a noticeable lack of large-scale longitudinal studies that review the outcome of HRT and GDx examination in groups of glaucoma suspects and patients.

The research described in this report was collected over a 6-year period by the University of Manchester at MREH. The study aimed to address the hiatus in the literature and provide much needed longitudinal data. The study examined two groups of patients cross-sectionally (150 glaucoma patients and 100 normal individuals) and two groups longitudinally (240 high risk eyes and 75 glaucoma patients). Each examination session involved image capture with the HRT and GDx and assessment with the 24-2 full threshold program of the HFA. Amongst the wider aims of the study was to investigate fully the equipment with regard to the feasibility of incorporating the HRT and GDx as part of the normal 'armoury' of the glaucoma clinic.

The study was compromised by technical problems (such as unstable back-up media) and data loss – a significant amount of data was lost at the outset of the study and further episodes followed.

A significant problem was identified that is highly relevant to the viability of conducting longitudinal research: whilst the study was under way, significant advances in the technology underlying the instruments occurred. This is particularly relevant to the GDx – the original GDx that was used for this research was superseded by the GDx VCC, which allows customised adjustments for individual variability in corneal polarisation. The instrument used for the Glaucoma Imaging Study was not amenable to upgrading, since both hardware and software developments are incorporated in its newer counterpart. The lack of backward compatibility is likely to have had little manifestation on the study's longitudinal results, since any error will have been constant throughout the study; it is, however, not possible to estimate the consequence to the results from the cross-sectional study.

Advances in the HRT actually assisted the study, since our system (an original 'Mark I' HRT) was recently upgraded to use the software designed for the HRT II (the instrument's newer equivalent). The more contemporary software is Windows (rather than DOS) based and is, subsequently,

far easier and more efficient to use. Using the improved technology greatly assisted analysis of the HRT data.

The upgrading of the HRT has caused slight inconsistencies within this report. For instance, results from one of the newer diagnostic algorithms, the MRA, are not reported in Chapter 11. In a study of this size, analysis was performed in several phases, some of which were completed with the earlier software.

Issues of developments in instrument hardware and software and of data loss relate not only to the research environment, but also to the clinical situation. If the tests were adopted within the NHS for clinical use, similar problems could occur. This matter is of particular relevance to glaucoma, where patients require long-term monitoring and, consequently, instruments must be capable of providing long-term follow-up.

In addition to the imaging devices, developments have also taken place in perimetry. When the study was established, the gold standard examination in glaucoma research was the HFA 24-2 full threshold program. However, subsequently, the new generation SITA algorithms have become widely accepted in both clinical and academic roles. Differences exist in the test duration and repeatability.

Another aspect of long-term research is that definitions and/or scientific opinion can change while the study is under way. An example of this problem is in defining glaucoma. The original study protocols stated that glaucoma should be defined in terms of visual field status and/or the appearance of the ONH and RNFL. However, subsequent to the beginning of the study, opinion changed, because it is considered that if a study is determining diagnostic accuracy of a technique, then the condition should be defined by an independent classifier. Consequently, patients within the study were recategorised on the basis of their visual fields alone, since if they had been categorised on the basis of their ONH and RNFL characteristics, it may have led to elevation of the apparent diagnostic precision of the HRT and GDx.

Technological developments in the HRT and GDx may decrease the relevance of findings from data presented in Chapter 4, which reported the duration of examinations with the HRT, GDx and HFA. It concluded that, in an average 3.5-hour clinical session, the three instruments could each be used to examine 17, 18 and seven patients, respectively. A greater number of patients per session could probably be examined with the newer imaging devices; similarly, the number of perimetric subjects could possibly be increased by utilising newer testing algorithms. This study defined the length of examination from the moment the patient was seated at the instrument until completion of the printout. This strategy was used to prevent the preferential appearance of a technique with rapid examination, but lengthy analysis phase.

Chapter 5 discussed economic aspects of utilising the equipment. Since there are many similarities in consumables between the instruments, the main differential financial factors are the number of patients that can be examined per session and the purchase and maintenance costs. Although the imaging devices can be used to examine over twice as many patients per session, the outlay for equipment is much greater, with the HRT and GDx each costing approximately twice that of a basic HFA; maintenance packages for the imaging devices are also more expensive than those for the HFA.

The issue of instrument development also applies to the study described in Chapter 6: applicability of the techniques. The number of patients who can successfully be examined with an instrument is clearly dependent on characteristics of that device. If the latter changes, then patients who were unable to be examined with the original appliance may be able to undergo examination with the more contemporary version and vice versa. The Glaucoma Imaging Study found that, from an unselected population of glaucoma patients, 98.3, 80.4 and 88.3% of patients could successfully undergo examination with the HFA, HRT and GDx, respectively. A common cause of failed examination with the HRT and GDx was that patients were unable to maintain fixation on the contra-lateral fixation target. The HRT II and GDx VCC each uses an ipsi-lateral fixation target. However, whether the newer devices will be able to successfully examine eyes with coexisting pathology that involves fixation remains to be determined. Another common cause of failed examination was that patients could not keep their eyes adequately still for the duration of image capture (1.6 and 0.7 seconds for the HRT and GDx, respectively). Reduced capture times with the novel devices may lead to increased applicability.

The cross-sectional study described in Chapter 7 determined the sensitivity and specificity of HRT

and GDx parameters. A limiting factor in comparing data from previous research is that different studies use diagnostic criteria with varying levels of specificity. For a condition such as glaucoma, it is thought to be ideal to set the diagnostic criteria to yield a specificity of 95% and estimate the corresponding sensitivity.

Diagnostic precision is highly vulnerable to the thresholds used to define abnormality. The glaucoma imaging study took the approach of defining glaucoma by three different perimetric criteria. At 95% specificity, optimum sensitivity from all the global HRT parameters was for the Bathija DFA, with sensitivity of 57%. The MRA achieved 59% sensitivity at a similar level of specificity. The diagnostic ability of the GDx was inferior to that of the HRT, but this may be a manifestation of the fact that our instrument was not capable of providing individualised corneal compensation: the optimum sensitivity (45%) was for the GDx number at 95% specificity.

The study used another measure for determining diagnostic precision: area underneath the ROC curve: this technique has the advantage of being a criterion-free method of establishing overall diagnostic precision. The majority of global HRT parameters had areas underneath the ROC curve that were interpreted as good or fair. This was not the case for the GDx, where a few parameters fell into the good or fair categories, but more were categorised as very poor; indeed, many measurements yielded a result just over 0.5, which represented little better than pure chance.

Chapter 8 described the outcome of longitudinal follow-up of the high risk cohort. Patients in this group had ocular hypertension and/or a fellow eye with POAG. Different methods were employed to determine if a visual field defect had developed, with a low level of agreement between the techniques. In all, 72 patients (30%) showed some evidence of the development of a visual field defect (termed perimetric conversion); however, the level of certainty differed between patients. Seven patients (3%) showed very definite conversion. The extent of agreement of detection of conversion appeared to be related to the definition employed. When stringent criteria were used, agreement improved. However, with less rigorous criteria, there was little agreement between the three measures. This finding suggests that the three techniques may detect different aspects of the development of glaucomatous optic neuropathy, and that no single method of examination can be used to detect early glaucoma.

The finding that ONH and RNFL analysis detects different groups of patients from visual field analysis challenges previous scientific thinking that dictated that structural alteration precedes functional change. However, findings from the Glaucoma Imaging Study are in line with contemporary findings from other research groups.

Chapter 9 reviewed longitudinal examination of glaucoma patients. The conclusions are similar to those from the high risk category, where there is little agreement between the three techniques in identifying patients who show deterioration. (Also, similarities between the two groups exist regarding the pattern of agreement depending on the strictness of criteria used.) Detection of perimetric progression was identified by three different methods, but few patients demonstrated a significant amount of change. The large degree of noise associated with glaucomatous visual fields may have incurred a number of FP progressors, which may have confounded the relationship between the HFA and the imaging devices. With regard to noise, the GDx parameter appeared to be less prone to intraindividual variation. Consequently, it may have a promising role in detection of glaucomatous RNFL changes.

Another finding from the longitudinal cohorts was that the overall amount of change within the groups was small. There may be an element of selection bias in that patients who volunteered to participate were meticulous and conscientious individuals who were more likely to be compliant with their medication and attendance at clinic appointments were selected. Differences between the rate of deterioration encountered in the Glaucoma Imaging Study population and normal clinic populations will form the basis of further research within the University of Manchester over the next 2 years.

Chapters 10 and 11 determined the effects of inter- and intraobserver variability in image capture and the placement of the peri-papillary contour line. Overall, variability was low, with most operators showing good agreement. However, the level of agreement and outcome of diagnostic categorisation varied between individuals. Such a study is necessary if the instruments are to be used in the normal clinical environment, where multiple operators may use the equipment. In the NHS, it is likely that a variety of operators might perform examinations such as HRT and GDx. The Glaucoma Imaging Study emulated this situation since, over the duration of the study, a total of 10 members of staff were involved in patient examination.

The most fundamental question of the Glaucoma Imaging Study is whether perimetry could be replaced by the HRT or GDx. Data presented in this report suggest that such a substitution could not be made – there is little agreement between the three techniques, which appear to identify changes at different times in different patients. Our findings are in line with the philosophy that glaucoma is not one homogeneous disease, with uniform appearance, but a variety of conditions, each with individual variations. A hypothesis is that the different methods of examination may each be able to detect change in separate types of glaucoma. However, further research would be necessary to investigate this issue further.

It has been suggested that the HRT and GDx could be used for 'remote' or 'telemedicine' clinics, where review of data and patient examination are separated in time and/or location. However, neither instrument provides adequate information for such an examination. In order to provide a full investigation of the glaucoma patient or suspect, the clinician needs a stereoscopic examination of the ONH and RNFL with true colour representation. This examination is necessary in order to detect changes such as RNFL haemorrhages or pallor of the neuro-retinal rim, which are significant findings in glaucoma. Such alterations cannot easily be detected with the HRT or GDx. The findings of the glaucoma imaging study suggest that, although the HRT and GDx provide good-quality digital images, their data may contribute little to a patient's clinical diagnosis but would add significantly to the cost of their assessment.

Acknowledgements

Special acknowledgment must go to the lead applicant, Professor David Henson, for his immense contribution to the research described within this report. In addition, thanks must go to the other co-authors, Dr Robert Harper, Miss Fiona Spencer and Professor David McLeod.

The administrative support of Miss Katherine Dickinson and Miss Joanne Bradley is gratefully acknowledged, as is the contribution of Chapter 10 by Miss Annemiek Coops.

Thanks must go to the volunteer patients, Marilyn Simkin, Eric Simkin and Professor David Henson, and also to study patients EW and MB who agreed to participate in the supplementary study covered in Chapter 10. The operators of Chapter 10 were Annemiek Coops, Amanda Jones, Dr Inma Pérez-Gomez, Dr Robert Harper, and Shauna Quinn. Thanks are also due to Miss Amanda Jones, Dr Robert Harper, Mr Jeffrey Kwartz, Professor Steve Parrish and Professor David Henson for their participation in the study covered in Chapter 11.

From April 1998 to March 2004, the study benefited from the assistance of a team of data collection technicians. Acknowledgment must be made to the hard work and commitment of Harpreet Josan, Anne Bjerre, Amanda Jones, Kathryn Dandy, Dr Inma Pérez-Gomez, Emma Gowan and Derek McPhee.

Contribution of authors

Anna Kwartz (Research Associate) was the project management lead author, responsible for running the study and writing the final report. David Henson (Professor of Ophthalmology) was the lead applicant. He designed the study and was closely involved in its day-to-day running. Robert Harper (Principal Optometrist) was a contributing author, and assisted in the design of the project and contributed to the final report. Fiona Spencer (Consultant Ophthalmologist) contributed to the clinical aspects of the study, assisted with recruitment and gave clinical input. David McLeod (Professor of Ophthalmology and Head of Department) oversaw the study and participated in biannual review meetings.

References

1. Bathija R, Gupta N, Zangwill L, Weinreb RN. Changing definition of glaucoma. *J Glaucoma* 1998;**7**:165–9.
2. Harrington DO. *The visual fields: a textbook and atlas of clinical perimetry*. 4th ed. St Louis, MO: Mosby; 1976.
3. Lee BL, Bathija R, Weinreb RN. The definition of normal-tension glaucoma. *J Glaucoma* 1998; **7**:366–71.
4. Shields MB. Classification of the glaucomas. In Ritch R, Shields MB, editors. *The glaucomas: clinical science*. St Louis, MO: Mosby; 1996. pp. 717–25.
5. Phelps CD. Ocular hypertension: to treat or not to treat? *Arch Ophthalmol* 1977;**95**:588–9.
6. Pohjanpelto PE, Palva J. Ocular hypertension and glaucomatous optic nerve damage. *Acta Ophthalmol* 1974;**52**:194–200.
7. Bengtsson B. The prevalence of glaucoma. *Br J Ophthalmol* 1981;**65**:46–9.
8. David R, Zangwill L, Stone D, Yassur Y. Epidemiology of intraocular pressure in a population screened for glaucoma. *Br J Ophthalmol* 1987;**71**:766–71.
9. Georgopoulos G, Andreanos D, Liokis N, Papakonstantinou D, Vergados J, Theodossiadis G. Risk factors in ocular hypertension. *Eur J Ophthalmol* 1997;**7**:357–63.
10. Aclimandos W, Galloway N. Blindness in the city of Nottingham (1980–1985). *Eye* 1988;**2**:431–4.
11. Grey RH, Burns-Cox CJ, Hughes A. Blind and partial sight registration in Avon. *Br J Ophthalmol* 1989;**73**:88–94.
12. King AJW, Reddy A, Thompson JR, Rosenthal AR. The rates of blindness and of partial sight registration in glaucoma patients. *Eye* 2000; **14**:613–19.
13. Stromberg U. Ocular hypertension. *Acta Ophthalmol* 1962;**69**:S7–S75.
14. Luntz MH, Sevel D, Lloyd JPF. Incidence of unsuspected chronic glaucoma in a population sample of Oxford. *BMJ* 1963;**ii**:1237–40.
15. Leibowitz HM, Krueger DE, Maunder LR, Milton RC, Kini MM, Kahn HA, *et al*. The Framingham Eye Study monograph: an ophthalmological and epidemiological study of cataract, glaucoma, diabetic retinopathy, macular degeneration, and visual acuity in a general population of 2631 adults, 1973–1975. *Surv Ophthalmol* 1980;**24**:335–610.
16. Vernon SA, Henry DJCL, Jones SJ. Screening for glaucoma in the community by non-ophthalmically trained staff using semi-automated equipment. *Eye* 1990;**4**:88–97.
17. Klein BE, Klein R, Sponsel WE, Franke T, Cantor LB, Martone J, *et al*. Prevalence of glaucoma. The Beaver Dam Eye Study. *Ophthalmology* 1992;**99**:1499–504.
18. Coffey M, Reidy A, Wormald R, Xian WX, Wright L, Courtney P. Prevalence of glaucoma in the west of Ireland. *Br J Ophthalmol* 1993; **77**:17–21.
19. Dielemans I, Vingerling JR, Wolfs RC, Hofman A, Grobbee DE, de Jong PT. The prevalence of primary open-angle glaucoma in a population-based study in The Netherlands. The Rotterdam Study. *Ophthalmology* 1994;**101**:1851–5.
20. Mitchell P, Smith W, Attebo K, Healey PR. Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology* 1996;**103**:1661–9.
21. Bonomi L, Marchini G, Marraffa M, Bernardi P, De Franco, I, Perfetti S, *et al*. Prevalence of glaucoma and intraocular pressure distribution in a defined population. The Egna-Neumarkt Study. *Ophthalmol* 1998;**105**:209–15.
22. Kurtz S, Goldenfeld M, Melamed S. Early detection of glaucoma by a mobile unit – results from 10,000 examinees. *Harefuah* 2000;**138**: 273–6, 342.
23. Bonomi L, Marchini G, Marraffa M, Morbio R. The relationship between intraocular pressure and glaucoma in a defined population. Data from the Egna-Neumarkt glaucoma study. *Ophthalmologica* 2001;**215**:34–8.
24. McNaught AI. Current diagnostic tests. Highlights report: the glaucoma challenge: tackling issues and improving outcomes. Presented at the European Glaucoma Society Congress, London, 2000.
25. Whitacre MM, Stein R. Sources of error with use of Goldmann-type tonometers. *Surv Ophthalmol* 1993; **38**:1–30.
26. Hollows FC, Graham PA. Intra-ocular pressure, glaucoma, and glaucoma suspects in a defined population. *Br J Ophthalmol* 1966;**50**:570–86.

27. Sommer A, Tielsch JM, Katz J, Quigley HA, Gottsch JD, Javitt J, *et al*. Relationship between intraocular pressure and primary open angle glaucoma among white and black Americans. The Baltimore Eye Survey. *Arch Ophthalmol* 1991; **109**:1090–5.

28. Shah S, Chatterjee A, Mathai M, Kelly SP, Kwartz J, Henson DB, *et al*. Relationship between corneal thickness and measured intraocular pressure in a general ophthalmology clinic. *Ophthalmology* 1999;**106**:2154–60.

29. Kitazawa Y, Horie T. Diurnal variation of intraocular pressure in primary open-angle glaucoma. *Am J Ophthalmol* 1975;**79**:557–66.

30. Newell FW, Krill AE. Diurnal tonography in normal and glaucomatous eyes. *Trans Am Ophthalmol Soc* 1964;**62**:349–56.

31. Abrams L, Scott I, Spaeth G, Quigley H, Varma R. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology* 1994;**101**:1662–7.

32. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Ophthalmol Soc UK* 1976;**74**:532–72.

33. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol* 1989;**107**:453–64.

34. Artes PH, Chauhan BC. Visual field and optic disc progression in open-angle glaucoma. Presented at the Annual Meeting of the Glaucoma Society of the International Congress of Ophthalmology, Paris, 2003.

35. Artes PH, Nicolela MT, LeBlanc RP, Chauhan BC. Visual field and optic disk progression in a longitudinal study of open-angle glaucoma. Presented at the Annual Meeting of the Association of Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2004. Abstract available at www.arvo.org.

36. Zangwill LM, Bowd C, Berry C, Williams J, Blumenthal E, Sánchez-Galeana C, *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.

37. Drance SM, Begg IS. Sector haemorrhage – a probable acute ischaemic disc change in chronic simple glaucoma. *Can J Ophthalmol* 1970;**5**:137–41.

38. Airaksinen PJ, Mustonen E, Alanko HI. Optic disc hemorrhages. Analysis of stereophotographs and clinical data of 112 patients. *Arch Ophthalmol* 1981;**99**:1795–801.

39. Airaksinen PJ, Mustonen E, Alanko HI. Optic disc hemorrhages precede retinal nerve fibre layer defects in ocular hypertension. *Acta Ophthalmol* 1981;**59**:627–41.

40. Drance SM. Disc hemorrhages in the glaucomas. *Surv Ophthalmol* 1989;**33**:331–7.

41. Airaksinen PJ, Drance SM, Douglas GR, Mawson DK, Nieminen H. Diffuse and localized nerve fiber loss in glaucoma. *Am J Ophthalmol* 1984;**98**:566–71.

42. Jonas JB, Nguyen NX, Naumann GOH. The retinal nerve fiber layer in normal eyes. *Ophthalmology* 1989;**96**:627–32.

43. Hitchings RA, Poinoosawmy D, Poplar N, Sheth GP. Retinal nerve fibre layer photography in glaucomatous patients. *Eye* 1987;**1**:621–5.

44. Henson DB, Artes PH, Chauhan BC. Diffuse loss of sensitivity in early glaucoma. *Invest Ophthalmol Vis Sci* 1999;**40**:3147–51.

45. Aulhorn E, Harms H. Early visual field defects in glaucoma. In Leydhecker W, editor. *Glaucoma Symposium Tutzing Castle*. New York: Karger; 1967. pp. 151–85.

46. Drance SM. The glaucomatous visual field. *Br J Ophthalmol* 1972;**56**:186–200.

47. Drance SM. The glaucomatous visual field. *Invest Ophthalmol Vis Sci* 1972;**11**:85–96.

48. Greve EL, Langerhorst CT, van den Berg TTJP. Perimetry and other visual function tests in glaucoma. In Cairns JE, editor. *Glaucoma*. Vol. I. London: Grune and Stratton; 1986. pp. 37–78.

49. Hitchings RA. *Fundamentals of clinical ophthalmology: glaucoma*. London: BMJ Publishing Group; 2000.

50. Heijl A, Lindgren G, Olsson J, Åsman P. Test retest variability in glaucomatous visual fields. *Am J Ophthalmol* 1989;**108**:204–8.

51. Wood JM, Wild JM, Hussey MK, Crews SJ. Serial examination of the normal visual field using Octopus automated projection perimetry – evidence for a learning effect. *Acta Ophthalmol* 1987;**65**:326–33.

52. Heijl A, Lindgren G, Olsson J. The effect of perimetric experience in normal subjects. *Arch Ophthalmol* 1989;**107**:81–6.

53. Werner EB, Adelson A, Krupin T. Effect of patient experience on the results of automated perimetry in clinically stable glaucoma patients. *Ophthalmology* 1988;**95**:764–7.

54. Marchini G, Pisano F, Bertagnin F, Marraffa M, Bonomi L. Perimetric learning effect in glaucoma patients. *J Glaucoma* 1991;**1**:102–6.

55. Johnson CA, Adams CW, Lewis R A. Fatigue effects in automated perimetry. *Appl Opt* 1988;**27**:1030–7.

56. Wildberger H, Robert Y. Visual fatigue during prolonged visual field testing in optic neuropathies. *Neuro-Ophthalmology* 1988; 8:167–74.

57. Hudson C, Wild JM, O'Neill EC. Fatigue effects during a single session of automated static threshold perimetry. *Invest Ophthalmol Vis Sci* 1994;**35**:268–80.

58. Heijl A, Drance SM. Changes in differential threshold in patients with glaucoma during prolonged perimetry. *Br J Ophthalmol* 1983; **67**:512–16.

59. Heijl A, Drance SM. Deterioration of thresholds in glaucoma patients during perimetry. *Doc Ophthalmol Proc Ser* 1985;**42**:129–36.

60. Katz J, Quigley HA, Sommer A. Repeatability of the glaucoma hemifield test in automated perimetry. *Invest Ophthalmol Vis Sci* 1995; **36**:1658–64.

61. Tuck MW, Crick RP. The projected increase in glaucoma due to an ageing population. *Ophthalmic Physiol Opt* 2003;**23**:175–9.

62. Bell R, O'Brien C. Accuracy of referral to a glaucoma clinic. *Ophthalmic Physiol Opt* 1997; **17**:7–11.

63. Henson DB, Spencer AF, Harper RA, Cadman E. Community refinement of glaucoma referrals. *Eye* 2003;**17**:21–6.

64. Henson DB. Results from a postal survey of UK consultant ophthalmologists. Presented at the Managing Glaucoma in the NHS conference, Manchester, 2001.

65. Brenton RS, Phelps CD, Rojas P, Woolson RF. Fluctuation on the Humphrey and Octopus perimeters. *Invest Ophthalmol Vis Sci* 1986; **28**:767–71.

66. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1987; **105**:1544–9.

67. Tuck MW, Crick RP. The age distribution of primary open angle glaucoma. *Ophthalmic Epidemiol* 1998;**5**:173–83.

68. National population projections. URL: http://www.statistics.gov.uk/downloads/ theme_population/PP2No23.pdf. Accessed 2003.

69. Bartsch DU, Intaglietta M, Bille JF, Dreher AW, Gharib M, Freeman WR. Confocal laser tomographic analysis of the retina in eyes with macular hole formation and other focal macular diseases. *Am J Ophthalmol* 1989;**108**:277–87.

70. Zinser G, Wijnaendts-van-Resandt RW, Dreher AW, Weinreb RW, Harbarth U, Burk ROW. Confocal laser tomographic scanning of the eye. *Proc SPIE* 1989;**1161**:334–7.

71. Burk ROW, Rendon R. Clinical detection of optic nerve damage: measuring changes in cup steepness with use of a new image alignment algorithm. *Surv Ophthalmol* 2001;**45**:S297–S303.

72. Burk ROW. Reference plane for three dimensional topographic optic disc analysis with the Heidelberg Retina Tomograph. *Invest Ophthalmol Vis Sci* 1995;**36**(4):S627.

73. Heidelberg Engineering. *Heidelberg Retina Tomograph operation manual software version 2.01*. Heidelberg: Heidelberg Engineering; 1997.

74. Burk ROW, Rohrschneider K, Völcker HE, Zinser G. Analysis of three-dimensional optic disk topography by laser scanning tomography. Parameter definition and evaluation of parameter inter-dependence. In Nasemann J, Burk ROW, editors. *Laser scanning ophthalmoscopy and tomography*. Munich: Quintessenz; 1990. pp. 161–75.

75. Lusky M, Bossem ME, Weinreb RN. Reproducibility of optic nerve head topography measurements in eyes with undilated pupils. *J Glaucoma* 1993;**2**:104–9.

76. Tomita G, Honbe K, Kitazawa Y. Reproducibility of measurements by laser scanning tomography in eyes before and after pilocarpine treatment. *Graefes Arch Clin Exp Ophthalmol* 1994;**232**:406–8.

77. Chauhan BC, LeBlanc RP, McCormick TA, Rogers JB. Test–retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. *Am J Ophthalmol* 1994;**118**:9–15.

78. Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg Retina Tomograph. *J Glaucoma* 1993;**2**:101–3.

79. Rohrschneider K, Burk ROW, Völcker HE. Reproducibility of topometric data acquisition in normal and glaucomatous optic nerve heads with the laser tomographic scanner. *Graefes Arch Clin Exp Ophthalmol* 1993;**231**:457–64.

80. Menezes AV, Giunta M, Chisholm L, Harvey PT, Tuli R, Devenyi RG. Reproducibility of topographic measurements of the macula with a scanning laser ophthalmoscope. *Ophthalmology* 1995;**102**:230–5.

81. Sharma K, Tan J, Hitchings RA. Sources of variability in three-dimensional optic nerve head quantitative analysis. *Invest Ophthalmol Vis Sci* 2001; **42**(4):S134.

82. Orgül S, Cioffi GA, Bacon D, Van Buskirk EM. Sources of variability of topometric data with a scanning laser ophthalmoscope. *Arch Ophthalmol* 1995;**113**:161–4.

83. Tan JC, Garway-Heath DF, Fitzke FW, Hitchings RA. Reasons for rim area variability in scanning laser

tomography. *Invest Ophthalmol Vis Sci* 2003; **44**:1126–31.

84. Dreher AW, Tso PC, Weinreb RN. Reproducibility of topographic measurements of the normal and glaucomatous optic nerve head with the laser tomographic scanner. *Am J Ophthalmol* 1991; **111**:221–9.

85. Charman WN. Imaging in the 21st century. *Ophthalmic Physiol Opt* 1998;**18**:210–23.

86. Zangwill LM, Irak I, Berry CC, Garden V, de Souza Lima M, Weinreb, RN. Effect of cataract and pupil size on image quality with confocal scanning laser ophthalmoscopy. *Arch Ophthalmol* 1997;**115**:983–90.

87. Mikelberg FS, Parfitt CM, Swindale NV, Graham SL, Drance SM, Gosine R. Ability of the Heidelberg Retina Tomograph to detect early glaucomatous visual field loss. *J Glaucoma* 1995; **4**:242–7.

88. Bathija R, Zangwill L, Berry CC, Sample PA, Weinreb RN. Detection of early glaucomatous structural damage with confocal scanning laser tomography. *J Glaucoma* 1998;**7**:121–7.

89. Greaney MJ, Hoffman DC, Garway-Heath DF, Nakla M, Coleman AL, Caprioli J. Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci* 2002;**43**:140–5.

90. Zangwill LM, van Horn S, de Souza LM, Sample PA, Weinreb RN. Optic nerve head topography in ocular hypertensive eyes using confocal scanning laser ophthalmoscopy. *Am J Ophthalmol* 1996;**122**:520–5.

91. Hatch WV, Flanagan JG, Etchells EE, Williams-Lyn DE, Trope GE. Laser scanning tomography of the optic nerve head in ocular hypertension and glaucoma. *Br J Ophthalmol* 1997;**81**:871–6.

92. Uchida H, Brigatti L, Caprioli J. Detection of structural damage from glaucoma with confocal laser image analysis. *Invest Ophthalmol Vis Sci* 1996;**37**:2393–401.

93. Kamal DS, Garway-Heath DF, Hitchings RA, Fitzke FW. Use of sequential Heidelberg retina tomograph images to identify changes at the optic disc in ocular hypertensive patients at risk of developing glaucoma. *Br J Ophthalmol* 2000; **84**:993–8.

94. Garway-Heath DF, Hitchings RA. Sources of bias in studies of optic disc and retinal nerve fibre layer morphology. *Br J Ophthalmol* 1998;**82**:986.

95. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology* 1998;**105**:1557–63.

96. Ford BA, Artes PH, McCormick TA, Nicolela MT, LeBlanc RP, Chauhan BC. Comparison of data analysis tools for detection of glaucoma with the Heidelberg Retina Tomograph. *Ophthalmology* 2003;**110**:1145–50.

97. Miglior S, Guareschi M, Albè E, Gomarasca S, Vavassori M, Orzalesi N. Detection of glaucomatous visual field changes using the Moorfields regression analysis of the Heidelberg retina tomograph. *Am J Ophthalmol* 2003; **136**:26–33.

98. Mardin CY, Horn F, Budde WM, Jonas JB. Monitoring of morphometric changes of optic discs with morphologic progression of glaucomatous optic atrophy by means of laser scanner tomography. *Klin Monatsbl Augenheilkd* 2000;**217**:82–7.

99. Tan JC, Hitchings RA. An approach for detecting glaucoma induced structural change in the optic nerve head. *Invest Ophthalmol Vis Sci* 2001; **42**(4):S118.

100. Chauhan BC, Blanchard JW, Hamilton DC, LeBlanc RP. Technique for detecting serial topographic changes in the optic disc and peripapillary retina using scanning laser tomography. *Invest Ophthalmol Vis Sci* 2000; **41**:775–82.

101. Weinreb RN, Dreher AW, Coleman A, Quigley H, Shaw B, Reiter K. Histopathologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness. *Arch Ophthalmol* 1990; **108**:557–60.

102. Choplin NT. Essentials of clinical scanning laser polarimetry. In Lemij HG, Schuman JS, editors. *The shape of glaucoma*. The Hague: Kugler Publications; 2000. pp. 39–62.

103. Lemij HG, Reus NJ, Colen TP. Sensitivity and specificity of new GDx parameters. *Invest Ophthalmol Vis Sci* 2001;**42**(4):S17.

104. Colen TP, Tang NE, Mulder PG, Lemij HG. Sensitivity and specificity of new GDx parameters. *J Glaucoma* 2004;**13**:28–33.

105. Laser Diagnostic Technologies *GDx nerve fiber analyzer primer*. San Diego, CA: Laser Diagnostic Technologies; 2000.

106. Paczka JA, Quigley HA, Friedman DS, Barrón Y, Vitale S. Capabilities of frequency-doubling technology, scanning laser polarimetry and nerve fiber layer photographs to distinguish glaucomatous damage. *Invest Ophthalmol Vis Sci* 1999;**40**(4):S841.

107. Sinai MJ, Essock EA, Fechtner RD, Srinivasan N. Diffuse and localized nerve fiber layer loss measured with a scanning laser polarimeter: sensitivity and specificity of detecting glaucoma. *J Glaucoma* 2000;**9**:154–62.

108. 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.

109. Lemij HG. The value of polarimetry in the evaluation of the optic nerve in glaucoma. *Curr Opin Ophthalmol* 2001;**12**:138–42.

110. Colen TP, Tjon-Fo-Sang MJ, Mulder PG, Lemij HG. Reproducibility of measurements with the nerve fiber analyzer (NFA/GDx). *J Glaucoma* 2000;**9**:363–70.

111. Hoh ST, Greenfield DS, Liebmann JM, Hillenkamp J, Ishikawa H, Mistlberger A, *et al*. Effect of pupillary dilation on retinal nerve fiber layer thickness as measured by scanning laser polarimetry in eyes with and without cataract. *J Glaucoma* 1999;**8**:159–63.

112. Holló G, Nagymihály A, Vargha P. Scanning laser polarimetry in corneal haze after excimer laser refractive surgery. *J Glaucoma* 1997;**6**:359–62.

113. Notting I, Colen TP, Reus NJ, Lemij HG. The effect of motion artifacts in scanning laser polarimetry. *Invest Ophthalmol Vis Sci* 2001; **42**(4):S132.

114. Kremmer S, Pflug A, Heiligenhaus A, Fanihagh F, Steuhl KP. Laser scanning topography and polarimetry with implantation of intraocular lenses before and after cataract surgery. *Klin Monatsbl Augenheilkd* 1999;**214**:378–85.

115. Hoh ST, Greenfield DS, Liebmann JM, Maw R, Ishikawa H, Chew SJ, *et al*. Factors affecting image acquisition during scanning laser polarimetry. *Ophthalmic Surg Lasers* 1998;**29**:545–51.

116. Gazzard G, Foster PJ, Devereux JG, Oen F, Chew PT, Khaw PT, *et al*. Effect of cataract extraction and intraocular lens implantation on nerve fibre layer thickness measurements by scanning laser polarimeter (GDx) in glaucoma patients. *Eye* 2004;**18**:163–8.

117. 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.

118. Tjon-Fo-Sang MJ, de Vries J, Lemij HG. Measurement by nerve fiber analyzer of retinal nerve fiber layer thickness in normal subjects and patients with ocular hypertension. *Am J Ophthalmol* 1996;**122**:220–7.

119. Paczka JA, Friedman DS, Quigley HA, Barron Y, Vitale S. Diagnostic capabilities of frequency-doubling technology, scanning laser polarimetry, and nerve fiber layer photographs to distinguish glaucomatous damage. *Am J Ophthalmol* 2000; **131**:188–97.

120. Horn FK, Jonas JB, Martus P, Mardin CY, Budde WM. Polarimetric measurement of retinal nerve fiber layer thickness in glaucoma diagnosis. *J Glaucoma* 1999;**8**:353–62.

121. Mills R, Chen PP. Studies with the GDx nerve fiber analyzer. In Lemij HG, Schuman JS, editors. *The shape of glaucoma: quantitative neural imaging techniques*. The Hague: Kugler Publications; 2000. pp. 129–48.

122. Corneal Compensation Update. URL: www.laserdiagnostic.com. Accessed 2001.

123. 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.

124. Greenfield DS, Knighton R, Feuer W, Schiffman J, Zangwill L, Weinreb RN. Correction for corneal polarization axis improves discrimination power of scanning laser polarimetry. *Invest Ophthalmol Vis Sci* 2001;**42**(4):S314.

125. Garway-Heath DF, Greaney MJ, Caprioli J. Correction for the erroneous compensation of anterior segment birefringence with the scanning laser polarimeter for glaucoma diagnosis. *Invest Ophthalmol Vis Sci* 2002;**43**:1465–74.

126. Zhou Q, Knighton R, Trost P, Boyett D, Dreher A. Variable corneal compensation on the assessment of the retinal nerve fiber layer thickness in eyes with corneal birefringence anomaly. *Invest Ophthalmol Vis Sci* 2001;**42**(4):S16.

127. Bowd C, Zangwill LM, Weinreb RN. Association between scanning laser polarimetry measurements using variable corneal polarization compensation and visual field sensitivity in glaucomatous eyes. *Arch Ophthalmol* 2003;**121**:961–6.

128. Weinreb RN, Bowd C, Zangwill LM. Glaucoma detection using scanning laser polarimetry with variable corneal polarization compensation. *Arch Ophthalmol* 2003;**121**:218–24.

129. Emerson MV, DeBroff B, Shafranov G. Effect of cataract surgery on retinal nerve fiber layer parameters measured by scanning laser polarimetry. *Invest Ophthalmol Vis Sci* 2001; **42**(4):S17.

130. 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.

131. Poinoosawmy D, Tan JC, Bunce C, Membrey LW, Hitchings RA. Longitudinal nerve fibre layer thickness change in normal-pressure glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2000;**238**:965–69.

132. Advanced Glaucoma Intervention Study. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology* 1994; **101**:1445–55.

133. Vingrys AJ, Demirel S. False-response monitoring during automated perimetry. *Optom Vis Sci* 1998; **75**:513–17.

134. Bengtsson B. Reliability of computerized perimetric threshold tests as assessed by reliability indices and threshold reproducibility in patients with suspect and manifest glaucoma. *Acta Ophthalmol* 2000;**78**:519–22.

135. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol* 1997; **75**:368–75.

136. Bengtsson B, Heijl A, Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. Swedish Interactive Thresholding Algorithm. *Acta Ophthalmol* 1998;**76**:165–9.

137. Bengtsson B, Heijl A. Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol* 1998;**76**:268–72.

138. Wild JM, Pacey IE, O'Neill EC, Cunliffe IA. The SITA perimetric threshold algorithms in glaucoma. *Invest Ophthalmol Vis Sci* 1999; **40**:1998–2009.

139. Wild JM, Pacey IE, Hancock SA, Cunliffe IA. Between-algorithm, between-individual differences in normal perimetric sensitivity: full threshold, FASTPAC, and SITA. Swedish Interactive Threshold algorithm. *Invest Ophthalmol Vis Sci* 1999;**40**:1152–61.

140. Wild JM, Pacey IE, O'Neill EC, Cunliffe IA. The SITA perimetric threshold algorithms in glaucoma. *Invest Ophthalmol Vis Sci* 1999; **40**:1998–2009.

141. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci* 2002;**43**:2654–9.

142. Chi T, Ritch R, Stickler D, Pitman B, Tsai C, Hsieh FY. Racial differences in optic nerve head parameters. *Arch Ophthalmol* 1989;**107**:836–9.

143. Dandona L, Quigley HA, Brown AE, Enger C. Quantitative regional structure of the normal human lamina cribrosa. A racial comparison. *Arch Ophthalmol* 1990;**108**:393–8.

144. Mansour AM. Racial variation of optic disk size. *Invest Ophthalmol Vis Sci* 1989;**30**(4):S367.

145. Poinoosawmy D, Nagasubramanian S, Wormald R, Hitchings R. Glaucoma and race. *Lancet* 1989; **i**:1134.

146. Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt J. Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey. *JAMA* 1991; **266**:369–74.

147. Varma R, Tielsch JM, Quigley HA, Hilton SC, Katz J, Spaeth GL, *et al*. Race-, age-, gender-, and refractive error-related differences in the normal optic disc. *Arch Ophthalmol* 1994;**112**:1068–76.

148. The Association for Research in Vision and Ophthalmology. URL: http://www.faseb.org/arvo/helsinki.htm. Accessed 2002.

149. Heidelberg Engineering. URL: http://www.heidelbergengineering.de/docs/hrt2-tutorial1.pdf. Accessed 2002.

150. Colen TP, Tjon-Fo-Sang MJH, Mulder PGH, Lemij HG. Understanding your clinical NFA/GDx data. In Lemij HG, Schuman JS, editors. *The shape of glaucoma, quantitative neural imaging techniques*. The Hague: Kugler Publications; 2000. pp. 307–19.

151. Henson DB, Earlam RA. Correcting lens system for perimetry. *Ophthalmic Physiol Opt* 1995; **15**:59–62.

152. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; 354:1896–900.

153. Henson DB, Coops A, Kwartz AJ, Harper RA, Spencer AF,McLeod D. Rate of functional and structural change in glaucoma suspects. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2004. Abstract available at www.arvo.org.

154. Kwartz AJ, Jones AK, Henson DB, Harper RA, Spencer AF. The effect of inter- and intra-observer variability in the definition of the optic disc margin on HRT and GDx parameters. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2002. Abstract available at www.arvo.org.

155. Kwartz AJ. Aspects of the clinical application of confocal scanning laser ophthalmoscopy and scanning laser polarimetry. PhD thesis, University of Manchester, 2002.

156. Kwartz AJ, Henson DB, Spencer AF, Harper RA, McLeod D. The ability of the Heidelberg Retina Tomograph and GDx to detect patients with early glaucoma. In: Henson DB, Wall M, editors. *Perimetry update*. The Hague: Kugler Publications; 2002: pp. 299–307.

157. Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D. Ability of HRT summary measures to identify perimetric glaucomatous progression. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2004. Abstract available at www.arvo.org.

158. McLeod D, Joseph AJ, Henson DB, Harper RA, O'Donoghue EP, Spencer AF. Applicability of scanning laser polarimetry, retinal tomography and threshold perimetry to an unselected population of patients attending a glaucoma clinic. *Invest Ophthalmol Vis Sci* 1999;**40**(4):S73.

159. Lan YW, Henson DB, Kwartz AJ. The correlation between optic nerve head topographic measurements, peripapillary nerve fibre layer thickness, and visual field indices in glaucoma. *Br J Ophthalmol* 2003;**87**:1135–41.

160. Garway-Heath DF. Discs in practice: employing new technologies in the clinical setting. In Vernon SA, Ménage M, Migdal C, editors. *Key advances in the effective management of glaucoma*. London: Royal Society of Medicine Press; 2000. pp. 27–35.

161. Shirato S, Inoue R, Fukushima K, Suzuki Y. Clinical evaluation of SITA: a new family of perimetric testing strategies. *Graefes Arch Clin Exp Ophthalmol* 1999;**237**:29–34.

162. Roggen X, Herman K, Van Malderen L, Devos M, Spileers W. Different strategies for Humphrey automated perimetry: FASTPAC, SITA standard and SITA fast in normal subjects and glaucoma patients. *Bull Soc Belge Ophtalmol* 2001;23–33.

163. Nordmann JP, Brion F, Hamard P, Mouton-Chopin D. Evaluation of the Humphrey perimetry programs SITA Standard and SITA Fast in normal probands and patients with glaucoma. *J Fr Ophtalmol* 1998;**21**:549–54.

164. Price of computer components. URL: www.dabs.com/uk. Accessed 2004.

165. Carrington Reid M, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA* 1995;**274**:645–51.

166. Harper RA, Reeves BC. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. *Invest Ophthalmol Vis Sci* 1999;**40**:1650–7.

167. Kremmer S, Ayertey HD, Selbach JM, Steuhl KP. Scanning laser polarimetry, retinal nerve fiber layer photography, and perimetry in the diagnosis of glaucomatous nerve fiber defects. *Graefes Arch Clin Exp Ophthalmol* 2000;**238**:922–6.

168. Vitale S, Smith SD, Quigley HA, Kerrigan-Baumrind LA, Pease ME, Varma R, *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.

169. Humphrey Systems. *Humphrey Field Analyzer II user's guide*. 2nd ed. Dublin, CA: Carl Zeiss; 1998.

170. Viswanathan AC, Fitzke FW,Hitchings RA. Humphrey visual field reliability indices in glaucoma: a large scale study. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2001. Abstract available online at www.arvo.org.

171. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci* 2000;**41**:2201–4.

172. Henson DB, Evans J, Chauhan BC, Lane C. Influence of fixation accuracy on threshold variability in patients with open angle glaucoma. *Invest Ophthalmol Vis Sci* 1996;**37**:444–50.

173. Joseph AJ. An investigation into threshold establishment with the threshold related suprathreshold programme of the Dicon LD 400 Autoperimeter. MSc dissertation, University of Manchester Institute of Science and Technology, 1997.

174. Kirkpatrick JNP, Manivannan A, Gupta AK, Hipwell J, Forrester JV, Sharp PF. Fundus imaging in patients with cataract: role for a variable wavelength scanning laser ophthalmoscope. *Br J Ophthalmol* 1995;**79**:892–9.

175. Pflug A, Kremmer S, Selbach JM. Laser scanning topography and polarimetry before and after cataract surgery. *Invest Ophthalmol Vis Sci* 1997; **38**(4):S835.

176. Collur S, Carroll AM, Cameron BD. Human lens effect on in vivo scanning laser polarimetric measurements of retinal nerve fiber layer thickness. *Ophthalmic Surg Lasers* 2000;**31**:126–30.

177. Laser Diagnostic Technologies. URL: http://www.laserdiagnostic.com/products/gdxaccess/benefits.asp. Accessed 2002.

178. Stamper RL. Glaucoma screening. *J Glaucoma* 1998;**7**:149–50.

179. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;**8**:283–98.

180. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.

181. Swets JA. Sensitivities and specificities of diagnostic tests. *JAMA* 1982;**248**:548–50.

182. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;**39**:561–77.

183. The area under an ROC curve. URL: http://gim.unmc.edu/dxtests/roc3.htm. Accessed 2004.

184. A comment on the ROC curve and the area under it as performance measures. URL: http://www.nhn.ou.edu/~marzban/roc.pdf. Accessed 2004.

185. Laser Diagnostic Technologies. GDx instruction manual – for versions 2.0.06 and higher. San Diego, CA: Laser Diagnostic Technologies, 2000.

186. King AJW, Bolton N, Aspinall P, O'Brien CJ. Measurement of peripapillary retinal nerve fiber layer volume in glaucoma. *Am J Ophthalmol* 2000;**129**:599–607.

187. 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;367–70.

188. 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.

189. Schuman JS, Hee MR, Puliafito CA, Wong C, Pedut-Kloizman T, Lin CP, *et al*. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol* 1995;**113**:586–96.

190. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999;**106**:2144–53.

191. Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology* 1999;**106**:653–62.

192. Viswanathan AC, Crabb DP, McNaught AI, Westcott MC, Kamal D, Garway-Heath DF, *et al*. Interobserver agreement on visual field progression in glaucoma: a comparison of methods. *Br J Ophthalmol* 2003;**87**:726–30.

193. Henson DB, Artes PH, Chauhan BC. Diffuse loss of sensitivity in early glaucoma. *Invest Ophthalmol Vis Sci* 1999;**40**:3147–51.

194. Burk ROW, Völcker HE. Current imaging of the optic disk and retinal nerve fiber layer. *Curr Opin Ophthalmol* 1996;**7**:99–108.

195. Mardin CY, Horn FK, Jonas JB, Budde WM. Preperimetric glaucoma diagnosis by confocal scanning laser tomography of the optic disc. *Br J Ophthalmol* 1999;**83**:299–304.

196. Iester M, Mikelberg FS, Drance SM. The effect of optic disc size on diagnostic precision with the Heidelberg Retina Tomograph. *Ophthalmology* 1996;**104**:545–8.

197. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994;**309**:102.

198. Britton RJ, Drance SM, Schulzer M, Douglas GR, Mawson DK. The area of the neuroretinal rim of the optic nerve in normal eyes. *Am J Ophthalmol* 1987;**103**:497–504.

199. Jonas JB, Gusek GC, Naumann GOH. Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Invest Ophthalmol Vis Sci* 1988;**29**:1151–8.

200. Iester M, Mikelberg FS, Drance SM. The effect of optic disc size on diagnostic precision with the Heidelberg retina tomograph. *Ophthalmology* 1997; **104**:545–8.

201. Kerrigan-Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, Mitchell RS. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci* 2000;**41**:741–8.

202. Kamal DS, Viswanathan AC, Garway-Heath DF, Hitchings RA, Poinoosawmy D, Bunce C. Detection of optic disc change with the Heidelberg retina tomograph before confirmed visual field change in ocular hypertensives converting to early glaucoma. *Br J Ophthalmol* 1999;**83**:290–4.

203. Greenfield DS, Knighton RW, Feuer WJ, Schiffman JC, Zangwill L, Weinreb RN. Correction for corneal polarization axis improves the discriminating power of scanning laser polarimetry. *Am J Ophthalmol* 2002;**134**:27–33.

204. Weinreb RN, Zangwill L, Berry CC, Bathija R, Sample PA. Detection of glaucoma with scanning laser polarimetry. *Arch Ophthalmol* 1998;**116**:1583–9.

205. Quigley HA, Addicks EM, Green WR. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fibre loss and visual field defect in glaucoma, ischaemic neuropathy, papilledema and toxic optic neuropathy. *Arch Ophthalmol* 1982;**100**:135–46.

206. Bengtsson B. The variation and covariation of cup and disc diameters. *Acta Ophthalmol* 1976; **54**:804–18.

207. Jonas JB, Gusek GC, Naumann GOH. Pseudoglaucomatous physiologic large cups. *Am J Ophthalmol* 1989;**107**:137–44.

208. Ramrattan RS, Rolfs RCW, Horman A, deJong PTVM. Are gender differences in disk and rim area due to differences in refractive error or height? The Rotterdam study. *Invest Ophthalmol Vis Sci* 1997;**38**(4):S824.

209. Johnson CA, Sample PA, Zangwill LM, Vasile CG, Cioffi GA, Liebmann JR, *et al*. Structure and function evaluation (SAFE): II. Comparison of optic disk and visual field characteristics. *Am J Ophthalmol* 2003;**135**:148–54.

210. Pratt AB, Medeiros FA, Bowd C, Sample PA, Boden RN, Weinreb RN, *et al*. Predicting early glaucomatous visual field loss using the Heidelberg Retina Tomograph. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2004. Abstract available at www.arvo.org.

211. Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, *et al*. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular

hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:701–13.

212. Armaly MF, Krueger DE, Maunder L. Biostatistical analysis of the collaborative glaucoma study. I. Summary report of the risk factors for glaucomatous visual field defects. *Arch Ophthalmol* 1980;**98**:2163–71.

213. Davanger M, Ringvold A, Blika S. The probability of having glaucoma at different IOP levels. *Acta Ophthalmol* 1991;**69**:565–8.

214. Goldberg I. Relationship between intraocular pressure and preservation of visual field in glaucoma. *Surv Ophthalmol* 2003;**48** Suppl 1: S3–S7.

215. 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. A*rch Ophthalmol* 2002;**120**:1268–79.

216. Bowd C, Weinreb RN, Lee B, Emdadi A, Zangwill L. Optic disc topography after medical treatment to reduce intraocular pressure. *Am J Ophthalmol* 2000;**130**:280–6.

217. Irak I, Zangwill L, Garden V, Shakiba S, Weinreb RN. Change in optic disk topography after trabeculectomy. *Am J Ophthalmol* 1996; **122**:690–5.

218. Chu DH, Greenstein VC, Thienprasiddhi P, Liebmann JM, Hood DC, Rich R. Patterns of visual field abnormalities in normal versus high tension glaucoma. Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, 2004. Abstract available at www.arvo.org.

219. Greenfield DS, Knighton RW. Stability of corneal polarization axis measurements for scanning laser polarimetry. *Ophthalmology* 2001;**108**:1065–9.

220. Girkin CA. Relationship between structure of optic nerve/nerve fiber layer and functional measurements in glaucoma. *Curr Opin Ophthalmol* 2004;**15**:96–101.

221. Leskea MC, Heijl A, Hyman L, Bengtsson B, Komaroff E. Factors for progression and glaucoma treatment: the Early Manifest Glaucoma Trial. *Curr Opin Ophthalmol* 2004;**15**:102–6.

222. Leske MC, Heijl A, Hussein M, Bengtsson B, Hyman L, Komaroff E. Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial. *Arch Ophthalmol* 2003; **121**:48–56.

223. Collaborative Normal Tension Glaucoma Study Group. Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. *Am J Ophthalmol* 1998;**126**:487–97.

224. Collaborative Normal Tension Glaucoma Study Group. The effectiveness of intraocular pressure reduction in the treatment of normal-tension glaucoma. *Am J Ophthalmol* 1998;**126**:498–505.

225. Leske MC, Heijl A, Hyman L, Bengtsson B, Komaroff E. Factors for progression and glaucoma treatment: the Early Manifest Glaucoma Trial. *Curr Opin Ophthalmol* 2004;**15**:102–6.

226. Pereira ML, Kim CS, Zimmerman MB, Alward WL, Hayreh SS, Kwon YH. Rate and pattern of visual field decline in primary open-angle glaucoma. *Ophthalmology* 2002;**109**:2232–40.

227. Crabb DP, Fitzke FW, McNaught AI, Edgar DF, Hitchings RA. Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology* 1997;**104**:517–24.

228. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. I*nvest Ophthalmol Vis Sci* 2002;**43**:1400–7.

229. Henson DB, Chaudry S, Artes PH. The relationship between sensitivity and variability in normal and glaucomatous visual fields. In Wall M, Wild SM, editors. *Perimetry update 1998/1999*. Amsterdam: Kugler and Ghedini; 2000. pp. 95–101.

230. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci* 2000;**41**:417–21.

231. Holmin C, Krakau C. Variability of glaucomatous visual field defects in computerised perimetry. *Graefes Arch Clin Exp Ophthalmol* 1979;**210**:235–50.

232. Leskea MC, Heijl A, Hyman L, Bengtsson B, Komaroff E. Factors for progression and glaucoma treatment: the Early Manifest Glaucoma Trial. *Curr Opin Ophthalmol* 2004;**15**:102–6.

233. Schwartz B, Takamoto T, Martin J. Increased rate of visual field loss associated with larger initial visual field threshold values on follow-up of open-angle glaucoma. *J Glaucoma* 2004;**13**:120–9.

234. Cioffi GA, Robin AL, Eastman RD, Perell HF, Sarfarazi FA. Confocal laser scanning ophthalmoscope. Reproducibility of optic nerve head topographic measurements with the confocal laser scanning ophthalmoscope. *Ophthalmology* 1993;**100**:57–62.

235. Brigatti L, Weitzman M, Caprioli J. Regional test–retest variability of confocal scanning laser tomography. *Am J Ophthalmol* 1995;**120**:433–40.

236. Spencer AF, Sadiq SA, Pawson P, Vernon SA. Vertical optic disk diameter: discrepancy between

planimetric and SLO measurements. *Invest Ophthalmol Vis Sci* 1995;**36**:796–803.

237. Hatch WV, Flanagan JG, Williams-Lyn DE, Buys YM, Farra T, Trope GE. Interobserver agreement of Heidelberg Retina Tomograph parameters. *J Glaucoma* 1999;**8**:232–7.

238. Kogure S, Chiba T, Kinoshita T, Kowa H, Tsukahara S. Effects of artefacts on scanning laser polarimetry of retinal nerve fibre layer thickness measurement. *Br J Ophthalmol* 2000;**84**:1013–17.

239. Kook MS, Sung K, Park RH, Kim KR, Kim ST, Kang W. Reproducibility of scanning laser polarimetry (GDx) of peripapillary retinal nerve fiber layer thickness in normal subjects. *Graefes Arch Clin Exp Ophthalmol* 2001;**239**:118–21.

240. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:15–174.

241. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.

242. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci* 1991;**32**:422–32.

243. Zambarakji HJ, Evans JE, Amoaku WM, Vernon SA. Reproducibility of volumetric measurements of normal maculae with the Heidelberg Retina Tomograph. *Br J Ophthalmol* 1998;**82**:884–91.

244. Zambarakji HJ, Vernon SA, Spencer AF, Amoaku WM. Reproducibility of volumetric macular measurements in diabetic patients with the Heidelberg Retina Tomograph. *Doc Ophthalmol* 1999;**97**:349–60.

245. Roff EJ, Hosking SL, Barnes DA. The influence of contour line size and location on the reproducibility of topographic measurement with the Heidelberg Retina Tomograph. *Ophthalmic Physiol Opt* 2001;**21**:173–81.

246. Garway-Heath DF, Poinoosawmy D, Wollstein G, Viswanathan A, Kamal DS, Fontana L, *et al*. Inter- and intra-observer variation in the analysis of optic disc images: comparison of the Heidelberg Retina Tomograph and computer assisted planimetry. *Br J Ophthalmol* 1999;**83**:664–9.

247. Hatch WV, Trope GE, Buys YM, Macken P, Etchells EE, Flanagan JG. Agreement in assessing glaucomatous discs in a clinical teaching setting with stereoscopic disc photographs, planimetry and laser scanning tomography. *J Glaucoma* 1999;**8**:99–104.

248. Iester M, Mikelberg FS, Courtright P, Burk RO, Caprioli J, Jonas JB, *et al*. Interobserver variability of optic disk variables measured by confocal scanning laser tomography. *Am J Ophthalmol* 2001;**132**:57–62.

249. Bland JM, Altman DG. Statistics notes: measurement error and correlation coefficients. *BMJ* 1996;**313**:41–2.

250. Miglior S, Casula M, Guareschi M, Marchetti I, Iester M, Orzalesi N. Clinical ability of Heidelberg retinal tomograph examination to detect glaucomatous visual field changes. *Ophthalmology* 2001;**108**:1621–7.

251. Jonas JB, Budde WM, Panda-Jonas S. Ophthalmoscopic evaluation of the optic nerve head. *Surv Ophthalmol* 1999;**43**:293–320.

252. Park KH, Park SJ, Lee YJ, Kim JY, Caprioli J. Ability of peripapillary atrophy parameters to differentiate normal-tension glaucoma from glaucomalike disk. *J Glaucoma* 2001;**10**:95–101.

253. Iester M, De Ferrari R, Zanini M. Topographic analysis to discriminate glaucomatous from normal optic nerve heads with a confocal scanning laser: new optic disk analysis without any observer input. *Surv Ophthalmol* 1999;**44**:S33–S40.

254. Swindale NV, Stjepanovic G, Chin A, Mikelberg FS. Automated analysis of normal and glaucomatous optic nerve head topography images. *Invest Ophthalmol Vis Sci* 2000;**41**:1730–42.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme, Department of Pharmacology & Therapeutics,
University of Liverpool

Professor Bruce Campbell, Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Dr Edmund Jessop, Medical Advisor, National Specialist, Commissioning Advisory Group (NSCAG), Department of Health, London

Professor Jon Nicholl, Director, Medical Care Research Unit, University of Sheffield, School of Health and Related Research

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme, Department of Pharmacology & Therapeutics,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield, School of Health and Related Research

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine

Dr Jeffrey Aronson
Reader in Clinical Pharmacology, Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford

Professor Deborah Ashby, Professor of Medical Statistics, Department of Environmental and Preventative Medicine, Queen Mary University of London

Professor Ann Bowling, Professor of Health Services Research, Primary Care and Population Studies, University College London

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford

Professor John Cairns, Professor of Health Economics, Public Health Policy, London School of Hygiene and Tropical Medicine, London

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York

Mr Jonathan Deeks, Senior Medical Statistician, Centre for Statistics in Medicine, University of Oxford

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge

Professor Sallie Lamb, Professor of Rehabilitation, Centre for Primary Health Care, University of Warwick

Professor Stuart Logan, Director of Health & Social Care Research, The Peninsula Medical School, Universities of Exeter & Plymouth

Dr Linda Patterson, Consultant Physician, Department of Medicine, Burnley General Hospital

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine

Professor Mark Sculpher, Professor of Health Economics, Centre for Health Economics, Institute for Research in the Social Services, University of York

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Birmingham

Ms Kate Thomas, Deputy Director, Medical Care Research Unit, University of Sheffield

Ms Sue Ziebland, Research Director, DIPEx, Department of Primary Health Care, University of Oxford, Institute of Health Sciences

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Lay Member, Bolton

Professor Max Bachmann Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant Paediatrician/ Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London

Professor William Rosenberg, Professor of Hepatology, Liver Research Group, University of Southampton

Dr Susan Schonfield, Consultant in Public Health, Specialised Services Commissioning North West London, Hillingdon Primary Care Trust

Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne

Dr Margaret Somerville, PMS Public Health Lead, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Associate Dean for Education, Head of Department of Obstetrics and Gynaecology, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust

Professor Tony Avery, Head of Division of Primary Care, School of Community Health Services, Division of General Practice, University of Nottingham

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Consultant in Pharmaceutical Public Health, National Public Health Service for Wales, Cardiff

Mrs Sharon Hart, Head of DTB Publications, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, South Gloucestershire Primary Care Trust

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, The Royal Marsden Hospital, Sutton

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Aileen Clarke,
Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, London

Dr Matthew Cooke, Reader in A&E/Department of Health Advisor in A&E, Warwick Emergency Care and Rehabilitation, University of Warwick

Dr Carl E Counsell, Clinical Senior Lecturer in Neurology, Department of Medicine and Therapeutics, University of Aberdeen

Ms Amelia Curwen, Executive Director of Policy, Services and Research, Asthma UK, London

Professor Gene Feder, Professor of Primary Care R&D, Department of General Practice and Primary Care, Barts & the London, Queen Mary's School of Medicine and Dentistry, London

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Bec Hanley, Co-Director, TwoCan Associates, Hurstpierpoint

Ms Maryann L Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, Academic Department of Radiology, The Institute of Cancer Research, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Professor Neil McIntosh, Edward Clark Professor of Child Life & Health, Department of Child Life & Health, University of Edinburgh

Professor James Neilson, Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust

Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

Dr Vimal Sharma, Consultant Psychiatrist/Hon. Senior Lecturer, Mental Health Resource Centre, Cheshire and Wirral Partnership NHS Trust, Wallasey

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, Department of Public Health, University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman,
Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond,
Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark,
Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson,
Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Mr John Dunning,
Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Services, West Middlesex University Hospital, Isleworth

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Alastair Gray,
Professor of Health Economics, Department of Public Health, University of Oxford

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Dr Duncan Keeley,
General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping,
Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay,
Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Julian Little,
Professor of Human Genome Epidemiology, Department of Epidemiology & Community Medicine, University of Ottawa

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Sue Moss, Associate Director,
Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Mrs Julietta Patnick,
Director, NHS Cancer Screening Programmes, Sheffield

Professor Tim Peters,
Professor of Primary Care Health Services Research, Academic Unit of Primary Health Care, University of Bristol

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Professor Peter Sandercock,
Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Dr Ken Stein,
Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter

Professor Sarah Stewart-Brown,
Professor of Public Health, University of Warwick, Division of Health in the Community Warwick Medical School, LWMS, Coventry

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278