Comparative study of automated imaging technologies for the diagnosis of glaucoma (GATE study). Evaluation of the diagnostic accuracy, performance as triage tests, and cost-effectiveness.

Report commissioned by: NIHR R&D HTA Programme

Augusto Azuara-Blanco[1]
Katie Banister[2]
Charles Boachie[3]
Peter McMeekin[4]
Joanne Gray[5]
Jennifer Burr[6]
Rupert Bourne[7]
David Garway-Heath[8]
Mark Batterbury[9]
Rodolfo Hernández[10]
Gladys McPherson[2]
Craig Ramsay[2]
Jonathan Cook[11]

[1]Centre for Experimental Medicine, Queen's University, Belfast, NI
[2]Health Services Research Unit, University of Aberdeen, UK
[3]Robertson Centre for Biostatistics, University of Glasgow, UK
[4]Health Economics Group, Institute of Health and Society, Newcastle University, UK
[5]Faculty of Health and Life Sciences, Northumbria University, UK
[6]School of Medicine, University of St. Andrews, UK
[7]Vision & Eye Research Unit, Postgraduate Institute, Anglia Ruskin University, UK
[8]NIHR Biomedical Research Centre, Moorfields Eye Hospital and UCL Institute of Ophthalmology, UK
[9]St Paul's Eye Unit, Royal Liverpool and Broadgreen, University Hospitals NHS Trust, UK
[10]Health Economics Research Unit, University of Aberdeen, UK
[11]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK

Correspondence to: Augusto Azuara-Blanco
Centre for Experimental Medicine,
Queen's University, Belfast, NI

Date completed: **10th October 2014**

Version 0.1

Declared competing interests of authors

DGH received grants from NIHR, during the conduct of the study; personal fees and non-financial support from Heidelberg Engineering UK, personal fees and non-financial support from Carl Zeiss Meditec, non-financial support from OptoVue Inc., non-financial support from Topcon, outside the submitted work.

Abstract

Comparative study of automated imaging technologies for the diagnosis of glaucoma (GATE study). Evaluation of the diagnostic accuracy, performance as triage tests, and cost-effectiveness.

Azuara-Blanco A[1]*, Banister K[2], Boachie C[3], McMeekin P[4], Gray J[5], Burr J[6], Bourne R[7], Garway-Heath D[8], Batterbury M[9], Hernández R[10], McPherson G[2], Ramsay C[2], Cook J[11]

[1]Centre for Experimental Medicine, Queen's University, Belfast, NI
[2]Health Services Research Unit, University of Aberdeen, UK
[3]Robertson Centre for Biostatistics, University of Glasgow, UK
[4]Health Economics Group, Institute of Health and Society, Newcastle University, UK
[5]Faculty of Health and Life Sciences, Northumbria University, UK
[6]School of Medicine, University of St. Andrews, UK
[7]Vision & Eye Research Unit, Postgraduate Institute, Anglia Ruskin University, UK
[8]NIHR Biomedical Research Centre, Moorfields Eye Hospital and UCL Institute of Ophthalmology, UK
[9]St Paul's Eye Unit, Royal Liverpool and Broadgreen, University Hospitals NHS Trust, UK
[10]Health Economics Research Unit, University of Aberdeen, UK
[11]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK

*Corresponding author

Background: Many glaucoma referrals from the community to hospital eye services (HES) are unnecessary. Imaging technologies can potentially be useful to triage this population.

Objectives: To assess the diagnostic performance and cost-effectiveness of imaging technologies, as triage tests, for identifying people with glaucoma.

Design: Within patient comparative diagnostic accuracy study. Markov economic model comparing the cost effectiveness of a triage test with usual care.

Setting: Secondary care

Participants: Adults referred from the community to HES for possible glaucoma.

Interventions: Heidelberg Retinal Tomogram (including two diagnostic algorithms, HRT-GPS and HRT-MRA), scanning laser polarimetry (GDx), and optical coherence tomography (OCT). The reference standard was clinical examination by a consultant ophthalmologist with glaucoma expertise including visual field testing and intraocular pressure measurement (IOP).

Main outcome measures: 1) Diagnostic performance of imaging: the eye with most severe disease was used. 2) Composite triage test performance (imaging test, IOP measurement and visual acuity measurement), using data from both eyes, in correctly identifying clinical management decisions, that is, 'discharge' or 'do not discharge'. Outcome measures were sensitivity and specificity, incremental cost per quality adjusted life year (QALY).

Results: Data from 943 of 955 participants were included in the analysis. The average age was 60.5 (SD 13.8) years, and 51.1% were females. Glaucoma was diagnosed by the clinician in at least one eye in 16.8%. 37.9% of participants were discharged after the first visit.

Regarding diagnosing glaucoma, HRT-MRA had the highest sensitivity (87.0%, 95% CI (80.2,92.1)) but lowest specificity (63.9%, 95% CI (60.2,67.4)), GDx had the

lowest sensitivity (35.1%, 95% CI (27.0,43.8)) but the highest specificity (97.2%, 95% CI (95.6,98.3)). HRT-GPS had sensitivity (81.5%, 95% CI (73.9,87.6)) and specificity (67.7%, 95% CI (64.2,71.2)), and OCT had sensitivity (76.9%, 95% CI (69.2,83.4)) and specificity (78.5, 95% CI (75.4,81.4)). Regarding triage accuracy, triage using HRT-GPS had the highest sensitivity (86.0%, 95% CI (82.8,88.7)) but lowest specificity (39.1%, 95% CI (34.0,44.5)), GDx had the lowest sensitivity (64.7%, 95% CI (60.7,68.7)) but the highest specificity (53.6%, 95% CI (48.2,58.9)).

Introducing a composite triage station into the referral pathway to identify appropriate referrals was cost-effective. All triage strategies resulted in a cost reduction compared with standard care (consultant led diagnosis) but with an associated reduction in effectiveness. GDX was the least costly and least effective strategy. OCT and HRT-GPS were not cost effective. Compared to GDX, the cost per QALY gained for HRT-MRA is £22,904. The cost per QALY gained with current practice is £156,985 compared to HRT-MRA. Large savings could be made by implementing HRT-MRA but some benefit to patients will be forgone. Results were sensitive to the triage costs.

Conclusions: Automated imaging can be effective to aid glaucoma diagnosis among individuals referred from the community to HES. A model of care using a triage composite test appears to be cost-effective.

Future work: There are uncertainties about glaucoma progression under routine care, and cost of providing health care. Acceptability of implementing a triage test needs to be explored.

WORD COUNT (487)

Table of contents

List of tables

List of figures

Alphabetical list of abbreviations/glossary

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for convenience.

List of abbreviations

AMD	Age-related macular degeneration
CI	Confidence interval
DOR	Diagnostic odds ratio
GAT	Goldmann applanation tonometry
GPS	Glaucoma Probability Score
HRT	Heidelberg Retinal Tomogram
IOP	Intra-ocular pressure
LE	Left eye
MD	Mean deviation
MRA	Moorfields Regression Analysis
NFI	Nerve fibre indicator
OAG	Open angle glaucoma
OCT	Optical coherence tomography
OHT	Ocular hypertension
PAC	Primary angle closure
PSD	Pattern standard deviation
QALY	Quality adjusted life year
RE	Right eye
ROC	Receiver operating characteristic
RNFL	Retinal nerve fibre layer
SAP	Standard automated perimetry
SD-OCT	Spectral domain optical coherence tomography
STARD	STAndards for the Reporting of Diagnostic accuracy studies
TSC	Trial steering committee
VA	Visual acuity
VF	Visual field
VFI	Visual field index
WTP	Willingness to pay

Scientific summary

Background

Glaucoma describes a group of chronic age-related eye diseases in which there is progressive damage of the optic disc and characteristic visual field loss. Glaucoma is a significant public health problem as it is the second leading cause of blindness in the UK.

Glaucoma care constitutes a major proportion of the workload of hospital eye services (HES). In England there are over one million glaucoma related outpatient visits to the acute sector annually. Considerable NHS resources are required to assess referrals to HES for possible glaucoma that are typically initiated by community optometrists. However, less than one quarter of referrals have glaucoma, and nearly 50% of referred individuals are discharged after their first visit. If referrals could be triaged in an effective and cost-effective manner, resources could be better utilised for other needs.

Glaucoma is diagnosed by clinicians detecting structural changes of the optic nerve head, [the optic disc] and corresponding visual field defects. New imaging techniques for assessment of the structural changes have emerged: scanning laser ophthalmoscopy, commercially available as the Heidelberg Retina Tomograph; (including two diagnostic algorithms, Moorfields Regression Analysis [HRT-MRA] and Glaucoma Probability Score [HRT-GPS]) and scanning laser polarimetry, commercially available as the GDx, and spectral domain optical coherence tomography (OCT), with several commercial devices available.

Imaging technologies are being introduced into glaucoma services but their role in the diagnostic pathway is unclear. Imaging tests are user friendly, safe, provide automated classifications and potentially could reduce the need for an examination by a clinician.

Aim

To assess the relative performance and cost-effectiveness of diagnostic imaging technologies, as triage tests in secondary care, for identifying people with glaucoma.

Objectives

Primary objective: To compare the diagnostic performance (in terms of sensitivity and specificity), in a cohort of patients referred to HES with possible glaucoma, of:

- four imaging tests (HRT-MRA, HRT-GPS, GDx, OCT) for diagnosis of glaucoma
- a composite triage test (combining imaging tests, visual acuity and intraocular pressure (IOP) measurements) in correctly identifying patients to be discharged from secondary care.

Secondary objectives:

a) To explore alternative thresholds for determining abnormal tests;
b) To evaluate the diagnostic performance of combinations of imaging tests;
c) To evaluate the performance of the tests across the spectrum of glaucoma (mild, moderate, severe);
d) To evaluate the cost-effectiveness of incorporating imaging in a triage test in HES compared with current practice of diagnostic examination by a clinician;
e) To evaluate patient preferences of different imaging technologies.

Methods

We designed a pragmatic within-patient comparative diagnostic and triage evaluation of imaging techniques for glaucoma. Participants were adult patients referred from community optometrists or general practitioners with any possible glaucoma-related findings. Five UK NHS centres participated: three academic centres and two district general hospitals.

Participants received all imaging tests: HRT-GPS, HRT-MRA, GDx and OCT. Possible tests results were: within normal limits; borderline; outside normal limits. The Heidelberg Retina Tomograph (HRT) uses confocal laser scanning to allow quantitative structural measurement of the optic disc anatomy. There are two main

classification tools to relate measurements to normative data (a) Moorfields Regression Analysis (HRT-MRA) which requires user definition of the optic disc boundary and (b) Glaucoma Probability Score (HRT-GPS) which is fully automated.

The GDx scanning laser polarimeter measures the retinal nerve fibre layer (RNFL) thickness surrounding the optic disc utilising the birefringent properties of the RNFL. The software provides a discriminating classifier termed the Nerve Fiber Indicator (NFI), which is fully automated.

Spectral Domain Optical Coherence Tomography (OCT) is an optical imaging technique providing high resolution, cross-sectional, imaging of the retina analogous to B-scan ultrasonography but utilising light instead of sound. The Spectralis OCT (Heidelberg Engineering) was used in this study.

The reference standard was a full clinical examination, including visual field testing, by a consultant ophthalmologist with glaucoma expertise to determine (a) a diagnosis of glaucoma (mild, moderate or severe) according to well defined criteria (diagnosis analysis), and (b) whether the patient would be discharged or should be monitored/treated within HES (triage analysis).

Statistical analysis

Sample size calculations were based on standard (McNemar) diagnostic accuracy study methods. A 5% significance level based upon a 2-sided test was used which required a study of 897 individuals to have 90% power to detect an accuracy difference of 9% for the primary outcome of glaucoma diagnosis. Including a 6% indeterminacy rate increased the sample size to 954.

Two diagnostic performance analyses were undertaken: a diagnosis and a triage analysis. For the diagnosis analysis (classification of glaucoma), one eye per patient was used: the eye with most severe disease except for one sensitivity analysis. The test “abnormal” definition was an imaging test result of “outside normal limits”, with borderline cases classified as “normal”. This was compared with a reference standard diagnosis of “glaucoma”.

For the triage analysis, a composite test (including three components: imaging; IOP measurement; visual acuity) was compared with a reference standard of clinical decision “do not discharge”. The test categorised a patient as needing evaluation by a clinician if any elements of the composite triage test were themselves “abnormal” in either eye: imaging classification “outside normal limits” or IOP>21mmHg or visual acuity of 6/12 or poorer.

Primary diagnostic performance outcomes were sensitivity and specificity of tests. Secondary diagnostic performance outcomes were likelihood ratio and diagnostic odds ratio (DOR). The proportions of indeterminate test results and low quality imaging and need for pupil dilation were measured and patient preference for the tests was ranked. The test performance was assessed with respect to the glaucoma spectrum (mild, moderate, severe), when including glaucoma suspects in the reference standard diagnosis, and when including “borderline” results as abnormal. The diagnostic performance of combinations of tests was also evaluated.

Economic analysis

A current practice pathway model was developed where patients referred to HES were seen by a nurse for visual acuity assessment, a technician for visual field measurement, and by a clinician.

In an alternative triage care pathway model, individuals were seen by a nurse for visual acuity examination and IOP measurement and a technician for imaging assessment. The triage test results classified patients as needing referral for clinician diagnosis, or discharged. Those referred were seen by a technician for visual field measurement and examined by a clinician.

The cost-effectiveness of four triage pathways, using IOP, visual acuity and each of the four imaging technologies (which varied by their diagnostic ability and capital cost) and their subsequent care management pathways were assessed using a multi-state Markov model compared with current practice.

The cohort started in one of six health states: normal; at risk of glaucoma; mild, moderate, severe glaucoma; sight impaired. The sensitivity and specificity of each triage strategy determined if diagnosis was correct, and depending on this, the health state patients would move to and associated progression of any underlying glaucoma. Modelled care pathways were developed in consultation with the study team and the independent steering committee and utilised our previous models in this area, reviewed guidelines, study data and expert opinion.

Consequences were considered in terms of monetary costs (of testing and subsequent management of the patient's condition) to the NHS and in terms of the effects on quality of life (by assigning utility weights). Combining these data with the probabilities of events occurring over time enabled cost, patient outcomes and quality adjusted life years (QALYs) to be estimated for a hypothetical cohort of patients undergoing each modelled strategy.

Model results were analysed as incremental cost per QALYs and incorporated: (i) costs (of testing) and triage diagnostic outcomes, (ii) costs (of testing and subsequent management) and (iii) QALYs. The base-case analysis used a cohort of 40 year-old males using prevalence data from the GATE study and for a 50 year time horizon. Cycle length was 1 year. Results were presented in incremental cost effectiveness ratios (ICERs).

Several deterministic sensitivity analyses were explored which varied: the annual probability of discharged patients having a sight test; the cost of triage tests; the start age of the cohort; the performance of the diagnosing clinician; the diagnostic performance of imaging technologies; the prevalence of glaucoma in referred population; utility weights for those "at risk of glaucoma". The possibility of a hypothetical pathway, in which patients diagnosed as "at risk of glaucoma" were discharged from the service, was explored to investigate the impact in terms of costs and QALYs.

Results

Between April 2011 and July 2013, 2088 participants were identified as potentially eligible: 2013 were invited to take part. Of those invited, 966 (48%) agreed to take part. Following consent, 11 participants were found to be ineligible and did not participate and 12 were excluded as they did not receive all 4 imaging tests. Therefore 943 participants were available for the comparisons of tests.

The average age of participants was 60.5 (SD 13.8) years, and 51.1% were females. Non participants had similar age and gender balance. Most participants (89.2%) were of "White British" ethnicity. The average IOP at referral was 20 mmHg. The most common diagnosis was "no glaucoma-related findings" (31.7% of participants). Co-morbidities were uncommon, except for cataract that was reported in 8.3% of right eyes and 7.4% of left eyes. Glaucoma was diagnosed in at least one eye in 16.8% of the GATE cohort and 6.5% had glaucoma in both eyes at referral. 37.9% of GATE participants were discharged after the first visit.

Performance of the imaging tests in diagnosing glaucoma differed. HRT-MRA had the highest sensitivity (87.0%, 95% CI (80.2,92.1)) but lowest specificity (63.9%, 95% CI (60.2,67.4)), GDx had the lowest sensitivity (35.1%, 95% CI (27.0,43.8)) but the highest specificity (97.2%, 95% CI (95.6,98.3)), and the other two tests providing intermediate results (HRT-GPS sensitivity (81.5%, 95% CI (73.9,87.6)) and specificity (67.7%, 95% CI (64.2,71.2)), and OCT sensitivity (76.9%, 95% CI (69.2,83.4)) and specificity (78.5, 95% CI (75.4,81.4)).

Likelihood ratios showed evidence of both being able to rule in and out the presence of glaucoma for all four imaging tests (95% CIs did not contain 1.0). DORs ranged from 9.24 for HRT-GPS to 18.48 for GDx.

When including borderline imaging results as test positive, the sensitivity increased but with a corresponding decrease in specificity. In this sensitivity analysis, HRT-MRA had the highest sensitivity (94.9%, 95% CI (89.8,97.9)) but the second lowest specificity (43.9%, 95% CI (40.2,47.6)), GDx had the lowest sensitivity (60.4%, 95%

CI (51.6,68.8)) but the highest specificity (82.8%, 95% CI (79.8,85.5)), and the other two tests providing intermediate results.

The impact of combining two imaging tests improved detection of glaucoma but the effect was marginal and smaller than the loss of specificity.

When considering a definition of disease of severe glaucoma only OCT had the highest sensitivity (95.2%, 95% CI (76.2,99.9)) and the second highest specificity (70.9%, 95% CI (67.7, 73.9)), GDx had the lowest sensitivity (78.9%, 95% CI (54.4,93.9)) but the highest specificity (93.7%, 95% CI (91.8,95.2)), and the other two tests providing intermediate results.

The performance of triage tests (a composite assessment comprising imaging test, IOP and visual acuity assessments) in correctly identifying patients to be discharged from secondary care showed that triage including HRT-GPS had the highest sensitivity (86.0%, 95% CI (82.8,88.7)) but lowest specificity (39.1%, 95% CI (34.0,44.3)), GDx had the lowest sensitivity (64.7%, 95% CI (60.7,68.7)) but the highest specificity (53.6%, 95% CI (48.2,58.9)), the other two tests providing intermediate results (HRT-MRA values were very similar to the HRT-GPS results (sensitivity 86.0%, 95% CI (82.8,88.7), and specificity (53.6 95% CI (48.2|,58.9)), and OCT had lower sensitivity 75.4%, 95% CI (71.6,78.9)) but higher specificity (41.0 95% CI (35.8|,46.3)) values than HRT-GPS and HRT-MRA). Likelihood ratios (and 95% CI) showed evidence of all four triage tests being able to rule in and out the presence of abnormalities for all four triage tests (CIs did not contain 1.0). DORs ranged from 2.1 for GDx to 3.9 for HRT-GPS.

Participant preference was collected for 890 participants (94%). Almost half of responders (48.2%) had no preference. Of those participants who gave a preference, OCT was ranked as most preferred (27.6%) followed by GDx (11.9%) and HRT (5.1%). Average time taken to perform the test varied from 5.2 minutes (SD 3.0) for OCT to 7.6 minutes (SD 5.0) for HRT.

Economic analysis results

All triage strategies were more cost-effective than current practice but resulted in reduced health due to missing cases (i.e. fewer expected QALYs). The base-case results suggest that, of the triage pathways modelled, a triage including IOP, visual acuity and HRT-MRA is the most cost-effective strategy. GDX was shown to be the least costly and least effective. OCT and HRT-GPS were not cost effective. Compared to GDX, the cost per QALY gained for HRT-MRA was £22,904.The cost per QALY gained with current practice was £156,985 compared to HRT-MRA. Large savings could be made by implementing HRT-MRA but some benefit to patients would be forgone. .

These results should be interpreted with some caution, particularly in terms of differences among triage strategies since the diagnostic accuracy of all tests (except GDx) and their unit costs are very similar. The incremental cost-effectiveness of the triage strategies compared to current practice is very sensitive to costs included in the model. Indeed, current practice becomes cost-effective when the total cost of a triage test increases to £30 and above. A key assumption used in the model was that clinicians are 100% accurate in their diagnostic ability. Relaxing this assumption increased further the incremental cost effectiveness ratios (favouring triage strategies).

Conclusions

Implications for health care

Imaging technologies can be effective to aid the diagnosis of glaucoma. An alternative pathway for patients referred from community to HES with possible glaucoma using a triage test that includes imaging, IOP and visual acuity appears to be cost-effective compared with current practice. Our findings are based on a relatively inexpensive composite triage test (£<30). The most cost-effective strategy would include HRT-MRA imaging. However, triaging would be associated with a loss of health, and the acceptability of this option among users and clinicians has not been evaluated.

Recommendations for research

- acceptability to patients and health care providers of implementing an efficient triage glaucoma diagnostic system but with reduced health.

- data on glaucoma disease progression, specifically including patients classified as glaucoma suspects and ocular hypertension, associated utility, and cost of providing health care
- investigation of varying the results of the imaging tests beyond the standard options, since the recommended classification may not be the one best suited to the population which GATE recruited from
- the effectiveness of implementation of a composite triage test

Funding

This project was funded by the NIHR Health Technology Assessment programme (09/22/111) and will be published in full in Health Technology Assessment; Vol. X, No. X. See the HTA programme website for further project information.

Word count (2,142, not including funding details)

Plain English summary

Glaucoma is a life-long eye disease. Treatment is usually effective to slow the progression of glaucoma. About 4000 people are registered with sight impairment each year because of glaucoma. Many healthy subjects are unnecessarily referred from the community to hospital eye services to rule out glaucoma.

New imaging tests that investigate the back of the eye can aid the diagnosis of glaucoma, and are safe and easy to perform. These technologies measure with high accuracy the tissues in the back of the eye that are typically thinned in glaucoma. This study was designed to evaluate the performance of four imaging tests at identifying, among patients referred to hospital, those who have glaucoma or are at risk and those who do not have any eye disease. We compared the imaging test results with an experienced eye doctor's diagnosis, and evaluated how well a possible care pathway would perform using imaging results combined with measurements of the eye pressure and vision, to identify whether the individual needed to see an eye doctor.

955 individuals were recruited. The best performing test correctly diagnosed glaucoma in 87 out of 100 patients. If imaging tests with an eye pressure test and a visual acuity test were used to screen out people without eye disease, there would be substantial savings to the health service, but not all patients with disease would be picked up. A relatively small proportion of patients with glaucoma and at risk of glaucoma would be missed (approximately 1 in 7).

Chapter 1 Introduction

Glaucoma describes a group of eye diseases in which there is progressive damage of the optic nerve. It is characterised by a specific pattern of optic nerve head and visual field loss leading to impaired vision and sometimes blindness if inadequately treated. Primary glaucoma can be classified as open angle glaucoma (OAG) or angle closure glaucoma, the former being the most common.[2] Glaucoma is a significant public health problem, second to macular degeneration as the most common cause of blindness in the UK,[3-5] and is the leading cause of irreversible blindness worldwide.[6] The impact on patients is considerable with the risks of moderate visual field loss (which affects the ability to drive) and long term blindness reported as the most important consequences.[7] Late detection is a major risk factor for glaucoma blindness.[8] However, if glaucoma is identified in the early stages, treatment is effective at reducing progressive disease.[9]

There are a number of factors which increase the risk of developing glaucoma, including elevated intraocular pressure (IOP), older age, ethnic background and family history of glaucoma. Of these, the level of IOP is the most important risk factor and is the only one which is treatable. Ocular hypertension, generally defined as an IOP of ≥21mmHg (2 standard deviations above the mean) used to be considered as a part of the definition of glaucoma, but population studies have consistently found that many people with glaucoma have an IOP below this level.[10-14] However, the risk of developing glaucoma, and for worsening of existing disease, increases with increasing IOP.[15-17] This is supported by the fact that those presenting with advanced glaucoma at diagnosis are more likely to have higher IOP.[13,18]

The estimated prevalence of glaucoma in the UK is over 1% of the population over 40 years of age.[19-22] Approximately 4000 new cases of severe sight impairment due to glaucoma are registered every year in the UK. Many more glaucoma patients have sight impairment not severe enough to be registered but with significant impact on their quality of life (e.g. loss of driving licence). In England and Wales, in 2007 there were over 5 million outpatient attendances at hospital eye services (around 10% of all annual outpatient attendances) in the NHS. Of these, approximately 1,400,000 were

new patients (costing over £140 million). As the population ages these numbers are likely to increase.[23]

Estimates based on official population projections and epidemiological prevalence surveys have predicted that the number of glaucoma cases in England and Wales will increase by one third by 2021 and continue to increase at a similar pace until 2031.[24]

Management of patients with glaucoma and those at risk of suffering from glaucoma constitutes a major part of the workload of any secondary care eye services. In two independent surveys, between 8%[25] and 13%[26] of all new referrals to secondary eye care were due to glaucoma, and 25% of all follow-up attendances were glaucoma related. In England alone there are over one million glaucoma related outpatient visits in the NHS hospital eye service annually (approximately 1% of all outpatient activity).[27] Currently, referrals for suspected glaucoma are usually initiated by a community optometrist and are assessed in hospital eye services by clinicians. However, the reported referral accuracy of glaucoma by optometrists is suboptimal. Less than a quarter of referrals actually have glaucoma, and nearly 50% of referred individuals are discharged after the first visit.[28] Thus, many referrals are unnecessary and overburden the already busy hospital eye services. It also causes distress and worry to the patient which could be avoided. Interventions such as glaucoma training[29] or agreed guidelines[30] may not always have an effect in the rates of false positives referrals by community optometrists.

Diagnosing glaucoma

Glaucoma is diagnosed primarily by detecting glaucomatous optic neuropathy (i.e. characteristic changes of the optic nerve head-the optic disc) and a compatible visual field defect. According to current NICE guidelines,[27] a definitive glaucoma diagnosis is based on the expertise of a clinician who subjectively interprets the appearance of the optic disc and the results of visual field testing. In addition to diagnosing glaucoma, the clinical examination will include a visual acuity test (to measure central vision), anterior chamber angle exam (to determine the mechanism of glaucoma, e.g. open-angle or angle-closure), and intraocular pressure measurement (which is a risk factor for glaucoma and also for disease progression).

Accurate clinical diagnosis of glaucoma is limited by subjectivity, reliance on the examiner's experience, and a wide variation of optic disc structure in the population. Imaging techniques for assessment of the structural changes at the optic nerve head and retinal nerve fibre layer (RNFL) have emerged and are in routine use in the NHS: the Heidelberg Retinal Tomogram (HRT-III), scanning laser polarimetry (GDx), and spectral domain optical coherence tomography (SD-OCT). These techniques can be easily performed by trained technicians and provide an automatic glaucoma classification index. Some clinicians now routinely incorporate the information from such imaging technologies to help make a diagnosis of glaucoma, although there is no strong evidence of their effectiveness.

Using an automated imaging quantitative test for glaucoma diagnosis may have advantages over visual field testing in that the majority of people can be imaged. [19]

Comparison of glaucoma diagnostic technologies

In 1997 the HTA programme funded a study entitled "The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma".[31] At the time, this study was the largest and most rigorous head to head comparison of tests for diagnosing glaucoma. However, this study used the first prototypes of the HRT and GDx, now outdated. Another serious limitation was the small study sample (250 participants), in addition to a potentially biased selection of patients as these were not consecutively selected.

A systematic review of the performance of technologies for detecting glaucoma both as screening or diagnostic tests for glaucoma identified that the evidence is of poor quality and that no one test was clearly superior.[19] In this systematic review it was also found that populations studied were varied and biased. Furthermore, there were only six studies that performed a direct comparison of the available diagnostic instruments (with the average size of the studies less than 300 patients), the threshold for definitions of glaucoma cases were not consistent, and there were no studies reporting on the performance of GDx and OCT that met the inclusion criteria for this systematic review. However, the review did suggest that some diagnostic technologies have better performance than others (e.g. HRT performed relatively well), but the

credible intervals around the estimates were wide, reflecting considerable uncertainty, and therefore recommended that the available diagnostic tests need to be evaluated in an appropriately powered directly comparative study.

In the published NICE guideline[27] the authors searched for evidence comparing the diagnostic performance of HRT, GDx and OCT with expert clinical examination. No studies met the inclusion criteria for the guideline review (Section 4.6.1, Page 85, NICE).

Triage tests in secondary care eye services

There are considerable NHS resources required to assess all patients referred to hospital eye services with suspected glaucoma. In June 2009, the chairman published on behalf of the Professional Standards Committee of the Royal College of Ophthalmology a statement that the interpretation of NICE glaucoma guidelines was putting considerable strain onto secondary care eye services through the increase in false–positive referrals from community optometrists. The statement proposed that Eye Departments should consider innovative and efficient clinics for the initial assessment of patients.[32]

If referrals could be triaged, to identify suitable referrals and discharge unsuitable referrals, in an effective and cost-effective manner, the resources could be better utilised for patient eye care services. Imaging technologies are being introduced into glaucoma services in both hospital and community settings but their role in the diagnostic pathway either as triage, replacement or add on tests has not been evaluated. The tests to be evaluated in this study are the currently available imaging technologies with characteristics that suggest that they could be valuable triage tests, and are in current use in the NHS. They do not require patient input, are user friendly,[33] provide automated quantitative classifications and potentially could reduce the need for an extensive examination by an expert glaucoma clinician. The diagnostic performance of these imaging technologies has not been evaluated in a triage setting and in a robust manner.

Research question, aims and research objectives

Research question

What is the relative performance and the cost-effectiveness of new diagnostic imaging technologies, as triage tests in secondary care, for identifying people with glaucoma?

Aim

To assess the relative performance and the cost-effectiveness of new diagnostic imaging technologies, as triage tests in secondary care, for identifying people with glaucoma.

Research objectives

Primary objective

To compare the performance of imaging technologies (HRT-MRA, HRT-GPS, GDx, and OCT) as diagnostic and triage tests for patients referred to hospital eye services with possible glaucoma. Triage tests include an imaging technology, visual acuity and intraocular pressure.

Secondary objectives

a) to explore alternative thresholds for determining test positivity
b) to evaluate the diagnostic performance of combinations of the imaging tests
c) to evaluate the performance of the tests across the spectrum of glaucoma (mild, moderate, severe)
d) to evaluate the cost-effectiveness of adopting individual tests or combination of tests as triage tests compared with the current practice of diagnostic examination by a clinician in a secondary care setting
e) to evaluate patient preferences of different imaging technologies.

Chapter 2 Methods

This chapter describes the GATE (**G**laucoma **A**utomated **T**ests **E**valuation) study design and methods for the diagnostic performance evaluation, and follows the STAndards for the Reporting of Diagnostic accuracy studies (STARD).[34] The methods for the health economic evaluation are described separately (*see Chapter 6*).

Overview of the study design

An overview of the GATE study design is shown in Figure 1. The GATE study is a pragmatic within-patient comparative diagnostic evaluation of four imaging techniques for glaucoma in patients referred to hospital eye services. Specifically this study was designed to evaluate (1) diagnostic accuracy of imaging tests for detecting glaucoma in an eye; and (2) diagnostic accuracy of triage tests that consisted of a combination of an imaging test, visual acuity and intraocular pressure (IOP) measurement, for identifying patients requiring referral to hospital eye services.

All patients recruited to the study received four different imaging tests (using three different devices) which were compared with a reference standard (i.e. a comprehensive clinical examination). The study was co-ordinated from a central study office in the Health Services Research Unit, University of Aberdeen.

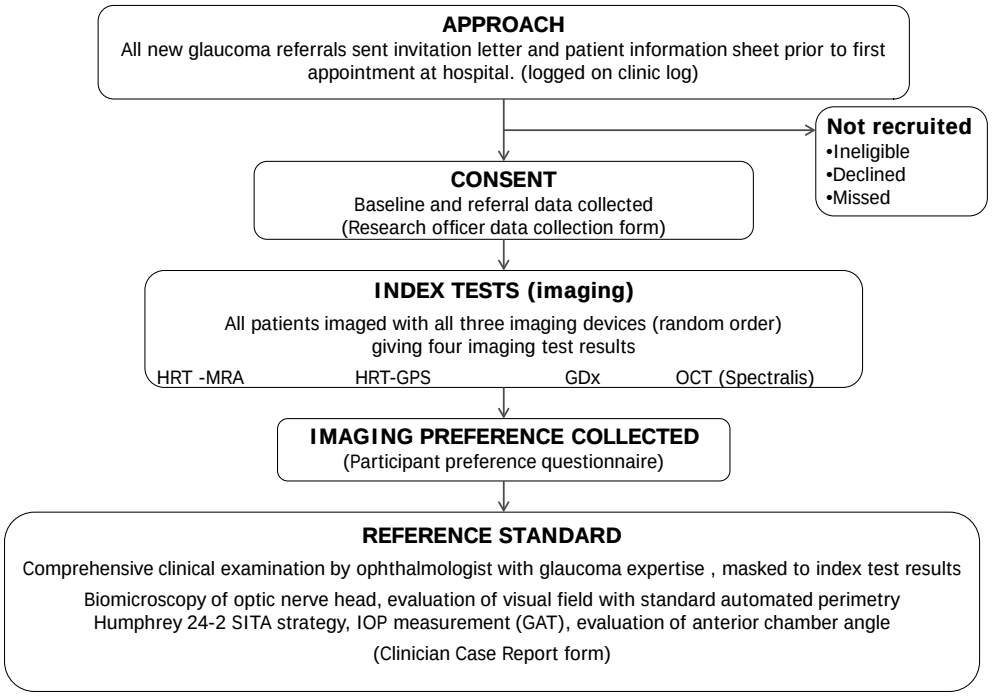


Figure 1 Overview of study design

Participants

Inclusion criteria

Adult patients referred from community optometrists or general practitioners to hospital eye services with any glaucoma-related findings, including those with ocular hypertension.

Exclusion criteria

Patients referred to hospital eye services because of other ocular disease; patients < 18 years old; patients who could not give informed consent; patients who had already been diagnosed with glaucoma; and patients referred from within secondary care.

Setting

Five NHS hospital eye services in the UK participated in this study: Aberdeen Royal Infirmary (Aberdeen), Bedford Hospital (Bedfordshire), Hinchingbrooke Hospital, Huntingdon (Cambridgeshire), Moorfields Eye Hospital (London) and St. Paul's Eye Unit (Liverpool). The participating units consisted of three academic units of different size and two district general hospitals (Hinchingbrooke and Bedford).

Identification of participants and recruitment process

Consecutive eligible patients referred from community optometrists to hospital eye services with a glaucoma-related finding were identified by the research officer in each centre at the time of referral. Patients were identified from their referral letter as being referred with a possible glaucoma diagnosis or glaucoma-related finding, including high IOP, possible abnormalities in the optic disc or visual field tests, and possible narrow anterior chamber angle. To ensure a full cross-section of referrals were identified, existing referral refinement schemes in two of the participating centres were suspended for the duration of the study in order not to introduce selection bias. In the largest centre (Moorfields Eye Hospital) only those patients booked to see a clinician trained in the study protocol to provide the reference standard were identified as eligible. Information about this study was sent to potentially eligible patients together with the date of the appointment (*see Appendix 2*). Patients were approached by the local research officer on their first visit to hospital eye services to discuss the study and those patients who agreed to participate and signed the consent form (*see Appendix 2*) were enrolled (i.e. before their consultation with the ophthalmologist). Each research centre kept a clinic log of eligible patients invited (*see Appendix 3*) which included patient demographics (age, gender) and, for those who declined to take part or were found to be ineligible, reason for not taking part if given.

Diagnostic technologies being assessed (index tests)

Four diagnostic tests from three imaging devices were evaluated:

(i) HRT-3: Confocal laser scanning imaging technology, employed by the Heidelberg Retina Tomograph (Heidelberg Engineering) exploits the principle of confocal laser scanning to allow quantitative structural information of the optic disc anatomy. The topographic image is derived from multiple optical sections at consecutive focal depth planes. Each image consists of numerous pixels, with each pixel corresponding to the retinal height at its location. Images are given a measure of quality – the mean topography standard deviation (SD) which the manufacturer recommends should be ≤40 μm. There are two main classification tools to define normality/outside normal limits: (a) the Moorfields Regression Analysis (MRA)[35] which requires the user to draw a contour line to define the optic disc boundary and (b) the

Glaucoma Probability Score (GPS)[36] which is fully automated and independent of operator input.

a. The HRT-MRA produces an overall ("global") classification as well as by 6 segments ("temporal", "temporal superior", "temporal inferior", "nasal", "nasal superior" and "nasal inferior") of the eye. For each a classification of "within normal limits", "borderline" and "outside normal limits" is given based upon whether the observed value is within the 95% prediction interval, between the 95.0% and the 99.9% prediction interval, or below the 99.9% predictions interval of the preset data, respectively. The final classification is based upon the most abnormal of any of the 7 classifications. If any one of these are "outside normal limits" then overall classification is "outside normal limit". Where there are none "outside normal limits" but at least one "borderline" then the final classification is "borderline". Only where the global and all 6 segment probabilities are "within normal limits" is the final classification "within normal limits".
b. HRT-GPS produces an overall probability of the presence of glaucoma ("global") and by segment ("temporal", "temporal superior", "temporal inferior", "nasal", "nasal superior" and "nasal inferior") for each eye. The default "final" classification is based upon applying cut-off to the overall and 6 segment probabilities; 0<0.28 is "within normal limits", >=0.28 and < 0.65 is "borderline" and ≥0.65 is "outside normal limits".[36] If any one of these are "outside normal limits" then overall classification is "outside normal limit". Where there are none "outside normal limits" but at least one "borderline" then the final classification is "borderline". Only where the global and all 6 segment probabilities are "within normal limits" is the final classification "within normal limits".

(ii) GDx-ECC: Scanning Laser Polarimetry measures the retinal nerve fibre layer (RNFL) thickness (Carl Zeiss Meditec). Measurements are based on the birefringent properties of the RNFL, which has its neurotubules disposed in an organised, parallel fashion. The software provides a discriminating classifier of glaucoma/normality, the Nerve Fiber Indicator (NFI) value, which is fully automated and is calculated for each eye. The manufacturers' reported cut-

offs for the GDx-ECC NFI value are based upon 95 and 99% coverage of the normative database population and are 1-35 ("normal"), 36-55 ("abnormal 95") and 56+ ("abnormal 99")(GDxPro Manual). The difference between "abnormal 95" and "abnormal 99" maybe viewed in a similar manner to the "borderline" category for HRT-GPS, HRT-MRA and OCT classifications. The "TSNIT" parameters used in the calculation of the NFI are also produced overall and by eye segment (superior and inferior) and an inter-eye symmetry is also produced. Images are given a quality figure (Q) which the manufacturer recommends should be ≥7. In this study GDx-ECC measurements were made using either the GDx-Pro (3 centres) or GDx-VCC with updated ECC module (2 centres).

(iii) OCT: Spectral Domain Optical Coherence Tomography (Spectralis, Heidelberg Engineering) is an optical imaging technique capable of providing high resolution, cross-sectional, imaging of the human retina in a fashion analogous to B-scan ultrasonography but utilising light instead of sound. OCT utilises the principles of low coherence interferometry using light echoes from the scanned structure to determine the thickness of the tissue. The glaucoma detection software of the Spectralis machine used in this study produces an average RNFL thickness value for the global and 6 segments of the eye and automatically compares sectors of RNFL thickness with a normative database An overall assessment of "within normal limits", "borderline" or "outside normal limits" is produced[35] based upon the global and the 6 individual segments. Inter-eye symmetry is also produced for each segment. Images are given a quality figure (Q) which the manufacturer recommends should be >15.

Sample reports generated by each of the imaging tests are shown in Appendix 4.

Reference standards

Eye level (i.e. for the diagnostic performance analysis)

The glaucoma diagnosis reference standard chosen for this study represents current clinical practice in the UK, which consists of clinical examination (biomicroscopy) of

the appearance of the optic nerve head, and evaluation of the visual field with standard automated perimetry (SAP) Humphrey 24-2 SITA strategy by an ophthalmologist with glaucoma expertise. In addition the clinician measured the intraocular pressure and examined the anterior chamber angle. The imaging tests were not available to the ophthalmologist when measuring the reference standard. The clinician recorded the status of each eye as described in Table 1 (i.e. glaucoma, OHT, glaucoma suspect, other eye morbidities or normal). If a clinical diagnosis could not be established at the first visit (e.g. unreliable visual field measurement requiring repeated measurement at a further appointment), an inconclusive diagnosis was recorded. In order to ensure valid and consistent application of the agreed reference standard, a limited number of consultant ophthalmologists provided the reference standard (1 or 2 clinicians in four centres, and 5 different clinicians at one centre). Principal investigators collaborating in each of the participating units gathered at the start of the project to review and agree on the reference standard (definitions of glaucoma, OHT, glaucoma suspect and normal) and how to define the spectrum of the disease (mild, moderate and severe). For this purpose training material was used including a series of cases with glaucoma related findings and also with normal subjects. Clinicians who were incorporated into the study at a later date to recruit and provide the reference standard were trained individually by the chief investigator with the same material.

For the eye level analysis, reference standard positive was classified as a diagnosis of glaucoma based upon the “worse” eye. Sensitivity analyses explored the diagnostic performance of the tests when also including glaucoma suspects in the definition of reference standard positive along with using the “best” eye (*see Statistical analysis methods below for full details).*

Table 1 Clinical diagnosis definitions

Diagnosis		**Definition**
Glaucoma	Severe	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss** *Severe*: MD worse than or equal to -12.01 dB
	Moderate	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss** *Moderate*: MD between -6.01dB and -12 dB
	Mild	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss** *Mild*: MD better than or equal to -6 dB
Glaucoma suspect	Disc suspect	Appearance suggestive of glaucomatous optic neuropathy but may also represent a variation of normality, with normal visual fields (with or without high IOP)
	Visual field suspect	Visual field loss suggestive of glaucoma, but may also represent a variation of normality, with normal appearance of the optic disc (with or without high IOP)
	Visual field and disc suspect	Both the optic disc and visual field have some features that resemble glaucoma but may also represent a variation of normality (with or without high IOP)
Ocular Hypertension (OHT)		When both the visual field and optic nerve appear normal in the presence of elevated pressure > 21mmHg
Primary Angle Closure (PAC)		Closed anterior chamber angle (appositionally or synechial) in at least 270°, and at least one of the following two: IOP > 21 mmHg and/or presence of peripheral anterior synechiae. Both visual field and optic nerve appear normal

PAC suspect	Closed anterior chamber angle (appositionally without any synechiae) in at least 270°, with IOP $\leq$ 21 mmHg. Both visual field and optic nerve appear normal
* *Evidence of optic nerve damage from any of the following: Optic disc or retinal nerve fibre layer structural abnormalities. Diffuse thinning, focal narrowing, or notching of the optic disc rim, especially at the inferior or superior poles. Documented, progressive thinning of the neuroretinal rim with an associated increase in cupping of the optic disc. Diffuse or localised abnormalities of the peripapillary retinal nerve fibre layer, especially at the inferior or superior poles. Disc rim or peripapillary retinal nerve fibre layer haemorrhages. Optic disc neural rim asymmetry of the two eyes consistent with loss of neural tissue.*	
** *Reliable visual field abnormality considered a valid representation of the subject's functional status. Visual field damage consistent with retinal nerve fibre layer damage (e.g. nasal step, arcuate field defect, or paracentral depression in clusters of test sites). Visual field loss in one hemifield that is different from the other hemifield, i.e. across the horizontal midline (in early/moderate cases). Absence of other known explanations.*	
REFERENCE STANDARD: For the eye level analysis, reference standard positive was classified as a diagnosis of glaucoma. Sensitivity analyses explored the diagnostic performance of the tests when also including glaucoma suspects in the definition of reference standard positive (see Statistical analysis methods below for full details)	

Patient level (i.e. for the triage performance analysis)

For each patient the clinical management decision made was recorded, that is, "discharge" or "don't discharge". Additionally, the reason for non-discharge (and which eye(s) it refers to) of "treatment" or "monitoring" was also collected. Clinicians were advised to follow NICE guidelines in deciding whether to discharge or not.[27]

Outcomes

For each of the four tests (HRT-MRA, HRT-GPS, GDx and OCT) the following outcomes were measured:

Diagnostic performance of imaging technologies

The primary diagnostic performance outcomes were sensitivity and specificity. Secondary diagnostic performance outcomes were likelihood ratio, and diagnostic odds ratio. The overall diagnostic performance of combinations of these four tests was also evaluated (HRT-MRA with each of the other three tests), and their relative performance. The diagnostic performance of the tests (and corresponding combinations) was also assessed according to the spectrum of glaucoma (mild, moderate, severe), as defined by the glaucoma expert.

Other outcomes

The proportions of indeterminacy results, and low quality imaging according to the manufacturer's recommendation, along with the participant's preference regarding the four tests were recorded for each test. Additionally, the number of participants who required pupil dilation to perform the imaging was also recorded. Dilation was attributed to the first imaging technology. Where a high quality test result was not available for a participant ("no result"), one of the following categorises applied:

A. test performed and imaging report produced but quality is lower than manufacturer quality cut-off
B. test performed and imaging report produced but no overall classification generated by machine
C. test performed but there was a clear imaging artefact on the report;

D. test attempted but no imaging could be acquired from the patient's eyes – no report generated
E. missing imaging output (due to study related or data collection issues).

Indeterminacy of the result was calculated as categories B-D over the total number of non-missing cases. The proportion of low quality imaging was A over the total number of non-missing cases minus categories A-D.

Diagnostic performance of a triage test (imaging test, visual acuity and intraocular pressure measurement)

As for the diagnosis analyses, the primary diagnostic performance outcomes of the triage test were sensitivity and specificity in correctly identifying patients who would be discharged from secondary care. Clinicians were advised to follow NICE guidelines in deciding whether to discharge or not.[27] Secondary diagnostic performance outcomes included likelihood ratios and the diagnostic odds ratios.

Delivery of interventions and data collection

Enrolled participants attended a diagnostic station for imaging (index test) and visual field measurement immediately prior to their meeting with the ophthalmologist. In three centres (Hinchingbrooke, Bedford, Liverpool), the visual field and imaging measurements took place on a separate day prior to the ophthalmologist appointment (within 2 weeks). Pupils were not routinely dilated. However, in those patients in whom adequate quality imaging could not be obtained, pupil dilatation could be used to try to improve image quality. In exceptional circumstances, where dilation was required in centres offering split visits, some or all of the imaging tests could be delayed until the clinic appointment, but always ahead of the clinical reference standard. Imaging technicians and the patient were therefore masked to the patient's underlying condition at the time of testing. In the remaining two centres (Aberdeen and Moorfields) all measurements were undertaken on the same day. All participants in each of the centres underwent testing with the three imaging devices, in a random order (to avoid bias when collecting participant preference) in one sitting. The random test order was automatically generated for each patient from the study website.

Imaging technicians employed at each centre performed the imaging tests. Between one and three technicians were identified at each centre and trained in study procedures prior to recruitment (*see Appendix 5*). There was no restriction for the same technician to perform all imaging tests on an individual. Across all centres, most technicians were experienced in performing the test prior to the study; where technicians were not already experienced they received training from the manufacturer or local imaging lead prior to collecting study data.

With the exception of HRT-MRA which required an experienced user to identify a contour line at the optic disc margin, all imaging tests generated the glaucoma classification automatically once an image had been acquired. The research officer kept printed copies of the images and uploaded the imaging results to the study website. Imaging reports were identified using a unique study number and date of birth.

The participant was asked to grade the tests in order of preference, or record no preference, using a standard form (*see Appendix 3*). Visual field measurements were undertaken with standard automated perimetry Humphrey SITA 24-2 strategy (Carl Zeiss Meditec) for each participant after all imaging tests had been completed. In exceptional circumstances, visual field measurements were undertaken ahead of the imaging tests due to clinic demand for equipment. Participants were then examined by an experienced glaucoma clinician who performed a comprehensive ocular examination including intraocular pressure (IOP) measurement with Goldmann applanation tonometry (GAT), gonioscopy, biomicroscopic examination of the optic disc (with pupil dilated in patients without narrow anterior chamber angle), and evaluated the visual field test results. The clinician provided the reference standard masked to the results of the imaging technologies and completed a clinical data collection form (*see Appendix 3*).

The research officer collated the results for each participant (*see Appendix 3*) including a copy of the visual field test, completed forms for each participant, uploaded the information onto the webpage, and posted original consent forms to the central office. Information uploaded onto the webpage included demographics,

referral IOP, refractive error, patient preference, need for pupil dilation, and Humphrey visual field reliability and global indices MD, PSD and VFI.

Data management

A web-based secure study database was developed for the GATE study which research staff could access remotely. Password protected access was provided such that centres could only view data from their own centre. All data collected during the course of the research was kept strictly confidential and accessed only by members of the study team. Minimal patient details were recorded and were stored under the guidelines of the 1988 Data Protection Act. Patients were allocated an individual study number and this number was used to identify study paperwork. Study data were entered and imaging reports uploaded onto the database by the research officer working in each centre. Whenever possible, drop-down boxes were employed to select appropriate responses and minimise typographical errors. Automated range checks and validation were inbuilt to ensure that inappropriate values could not be recorded.

Staff in the Study Office monitored data centrally and worked closely with local research officers to ensure that the data were as complete and accurate as possible. Missing forms and primary outcome data were automatically identified on the study website and distributed to local research officers on a regular basis. Uploaded imaging reports for each participant were checked by the central office, following an agreed checklist, and errors flagged for correction to the appropriate research team on a regular basis. This resulted in a low percentage of missing primary outcome data (1% reference standard: 1-3% imaging data). The content of approximately 50 case report forms and imaging reports selected at random were checked against entered data to ensure data entry accuracy. If consistent errors or discrepancies were found this triggered a further training session with the research officer to discuss and resolve data collection and entry issues.

The chief investigator checked a random sample of HRT-MRA imaging reports from each centre (5 reports for each operator at each centre) for accurate location of the

optic disc margin. A high error rate (more than 2 of 5 checked) at one centre triggered a complete check of the data at that centre: images with incorrectly placed contour lines were excluded from the default analysis and classified as artefact as described in Chapter 4.

Statistical analyses

Sample size

The sample size calculation and analysis were based on standard diagnostic accuracy study methods.[37] The sensitivity and specificity of each of the automated imaging tests were compared. A 5% significance level based upon a 2-sided test was used in the sample size calculations. A study of 897 individuals would have 90% power to detect a difference in accuracy of 9% for the primary outcome of diagnosis of glaucoma. This is based upon conservative assumptions of a probability of disagreement of 0.18 (maximum level possible), a glaucoma rate of 25% (as seen in similar populations) and a sensitivity of 86% as found in a systematic review for HRT.[19] Given this sample size, there would also be 80% power for detecting a 6% difference in accuracy should the sensitivity be 93% (the current best estimate from meta analyses of high quality diagnostic studies). For specificity, we would have over 90% power to detect a 5% difference. Based upon current available evidence, a rate of 6% indeterminacy of tests results was assumed which increased the sample size to 954 in total. A sample of this size would be of sufficient size for other measures of diagnostic performance (e.g. the sensitivity and specificity of individual technologies would be estimated to a 95% confidence interval (CI) of width 10% and 5% respectively).

Overview of planned analyses

To address the primary objective, two sets of pre-planned statistical analyses and sensitivity analyses of the diagnostic performance were carried out. They were:

1. "glaucoma diagnosis" analyses focused upon the clinical diagnosis of glaucoma (*see Chapter 4*)
2. "triage" analyses those focused upon the clinical discharge decision (*see Chapter 5*)

Glaucoma diagnosis analyses of diagnostic performance

The diagnostic performance of the four imaging tests (HRT-GPS and HRT-MRA outputs, GDx-ECC, and OCT) from three imaging devices for detecting glaucoma was calculated and compared. The "worst" eye of each participant as defined by the clinical reference standard was used in these analyses except for one sensitivity analysis which used the "best" eye of each participant. The reference standard was a clinical diagnosis of glaucoma (mild, moderate or severe) by an ophthalmologist (*see section 'Reference standards' above*). Diagnosis was ranked in order of decreasing severity as severe glaucoma, moderate glaucoma, mild glaucoma, glaucoma suspect (of any kind), PAC, OHT, normal (including all other diagnoses). The "worst" eye on the basis of comparing eyes using this ranking was used. If the two eyes had a similar spectrum of disease then a random eye was chosen. The primary analysis definition did not include a glaucoma suspect (whether disc and/or visual field based suspicion). The initial "positive" test definition under the respective imaging assessment was a test result of "outside normal limits" for HRT-MRA, HRT-GPS, OCT and NFI≥56 for GDx, with borderline cases classified as "negative".

Triage analyses of diagnostic performance

This set of analyses focussed upon the clinical decision for the management of a participant (discharged or not discharged). The reference standard for these analyses was person level clinical decision ("not discharged" or "discharged"). "Not discharged" was defined a "positive" test result for the reference standard. The decision to "not discharge" a patient may have been due to the diagnosis of an eye condition which needs treatment (glaucoma or otherwise) or the need for monitoring in one or both eyes. As visual acuity and IOP influence the clinical decision to discharge or not a patient for conditions other than glaucoma, and are routinely collected, these data were incorporated and a composite triage test was defined. In these analyses the discharge status of the patient was compared with a composite "test" which will be a combination of results from: an imaging test, the measurement of IOP, and visual acuity.

Following the statistical analysis plan, the diagnosis results (according to diagnosis performance and proportion of indeterminate tests) were considered prior to conduct

of the triage analysis. Corresponding triage analyses of all four imaging tests were then conducted according to the following definitions. An “abnormal” result for the imaging component was defined as including borderline as “abnormal”. An “abnormal” result for the IOP measurement component was a pressure of greater than 21mmHg as measured by the ophthalmologist. Similarly for visual acuity, an “abnormal” test result was defined as 6/12 or poorer as measured prior to referral by an optometrist. The visual acuity cut-off point (6/12) was chosen because below this level patients would not be able to drive and would merit further investigation to justify the reduced vision. Visual acuity was assumed not to be abnormal if it was not mentioned in the referral letter. The composite test was classified as “abnormal” if any of three components tests were judged to be abnormal for either eye.

Statistical analysis methods

Diagnostic performance analysis methods

Diagnostic measures (sensitivity, specificity, likelihood ratios and diagnostic odds ratio) were calculated for each test with appropriate CI.[38,39] All analyses were conducted at a 5% (2 sided) significance level with 95% CIs produced where appropriate. Under the diagnoses analyses, the diagnostic performance (sensitivity and specificity) of the alternative imaging tests was compared using McNemar’s test (default analyses only).[37] Corresponding CIs for the paired difference were generated.[40] No missing imaging, IOP or reference standard data was imputed. Visual acuity was assumed not to meet the abnormal criteria if not reported.

Sensitivity analyses of diagnostic performance

A range of sensitivity analyses were conducted for the diagnosis and/or triage analyses. These were:

- varying the imaging test cut-off to explore possible threshold effects. This was done by classifying borderline as diseased for the overall classification and also by utilising the parameters reported by each imaging test. A Receiver operation characteristic (ROC) curve and the area under the curve with the corresponding 95% confidence interval was calculated for each parameter using a non-parametric approach (SAS Logistic command). The results of the threshold assessment are given in Appendix 6 [diagnosis analysis only]

- varying the reference standard definition of abnormal (e.g. inclusion of glaucoma suspects for diagnoses analyses) [both diagnosis and triage analyses]
- removing the imaging quality requirement and/or assuming indeterminate results were abnormal [both diagnosis and triage analyses]
- using a combination of (two) tests for diagnostic performance. The choice of combinations was informed by the individual imaging test glaucoma diagnosis analyses [diagnosis analysis only]
- assess the impact of using "best" eye instead of the "worst" eye for each participant as defined by the clinical reference standard [diagnosis analysis]
- varying the IOP cut-off value for the pressure component of the test to be classified as "abnormal". A further analysis using a cut-off of IOP> 25mmHg was carried out [triage analysis only]
- using the referral IOP measurement instead of the ophthalmologist's measurement to define the positive IOP component of the triage test. For this analysis IOP >21 mmHg will be used as the cut-off for OHT [triage analysis only]
- varying the threshold for the visual acuity component of the composite test to be classified as "abnormal" [triage analyses only]
- utilising a composite test without a visual acuity component (i.e. only imaging and IOP components) [triage analyses only].

Diagnostic analyses to populate the Health economic model

A third set of analyses were produced in order to provide the most appropriate diagnostic performance data to populate the economic model (*see Appendix 7 for the results*). Under these analyses, the reference standard was detection of glaucoma and those "at risk" of glaucoma (i.e. a patient who was a glaucoma suspect of any kind, PAC or OHT). This is because people with these potential diagnoses need to remain monitored in secondary care according to the NICE guidelines. Any modelled triage system would need to reflect standard practice.[27]

Other outcomes

Two other outcomes were used to evaluate each of the four tests: Indeterminacy of tests and participant preferences. Indeterminacy of tests was quantified as the

proportion of tests that are indeterminate for each of the four imaging tests. This outcome was calculated in two ways, those which meet the manufacturer's suggested quality requirements and those for which a test result was produced. Participants' preference ranking of the three imaging technologies was summarised.

Patient and public involvement

Representatives from a UK based charity for glaucoma patients, the International Glaucoma Society, were involved in the study oversight throughout the project through the steering committee. This included: review and development of the study protocol and patient paperwork; monitoring the study progress; review and discussion of the final results of the study, including the care pathways and sensitivity analyses for the economic analyses, with particular reference to the patient perspective; proposing further research priorities particularly the acceptability of this new model of care. Additionally a patient with glaucoma reviewed and commented on the lay summary of the report.

Study oversight and management arrangements

The University of Aberdeen sponsored the study. An independent Trial Steering Committee (TSC) was established. The TSC comprised an independent chairperson (Ophthalmologist and senior academic), three further independent members (two Ophthalmologists and the chief executive of a UK based charity for glaucoma patients, the International Glaucoma Association), and the study grant-holders. The TSC met approximately annually over the course of the study. A patient (I.R.) agreed to provide advice on certain aspects of the study, but was not a member of the TSC. No Data Monitoring Committee was used as there were no safety concerns as the diagnostic technologies under evaluation were non-invasive, were routinely performed in clinical settings, and patient management did not change.

The day-to day running of the study was the responsibility of the chief investigator (Azuara-Blanco) supported by the research manager, research fellow and data support staff. A project management group consisting of the co-applicants provided strategic, management and content expertise to the study.

Ethical arrangements and regulatory approvals

The study and subsequent amendments were reviewed and given a favourable opinion by the North of Scotland Research Ethics Committee (reference 10/S0801/58) and local Research and Development Departments. The study was conducted according to the principles of Good Clinical Practice.

Protocol amendments after study initiation

A number of minor protocol revisions were made after study initiation.

Versions of the study protocol

Version 1, 28th July 2010

Version 1.1, 31 January 2011 (minor typographical changes)

Version 1.2, 17 April 2012 (extension of recruitment timescale)

Version 1.3, 11April 2013 (extension of recruitment timescale)

Version 1.4 4 July 2013 (updated list of grant holders and TSC members)

These are included in Appendix 1.

Chapter 3 Participant characteristics

This chapter provides an overview of the baseline characteristics of participants in the GATE study.

Recruitment of participants

Between April 2011 and July 2013, 2088 participants were identified as potentially eligible to take part in the study: 2013 were sent letters of invitation and patient information sheets. Of those invited, 966 (48%) agreed to take part, and 265 (13%) expressed a preference for not participating. Characteristics of non-participants are detailed in Table 2.

Following consent, 11 participants were subsequently excluded from the study: 10 were ineligible (4 had pre-existing glaucoma, 4 were referred from secondary care, 2 were not referred for glaucoma) and 1 person withdrew from the study. Therefore 955 participants were available for the index test comparison. Additionally, due to administrative and research processes, imaging was not implemented for all imaging tests in 12 participants and these were excluded from all analyses. The baseline measurements presented in this chapter relate to the remaining 943 participants.

Figure 2 shows a consort style diagram of the recruitment to the study. Full details of patient flow through the diagnostic performance analysis are described within the results Chapters 4 and 5.

Aberdeen and Hinchingbrooke were the highest recruiting centres (Table 3). Over 2/3 of GATE participants were recruited from these two sites.

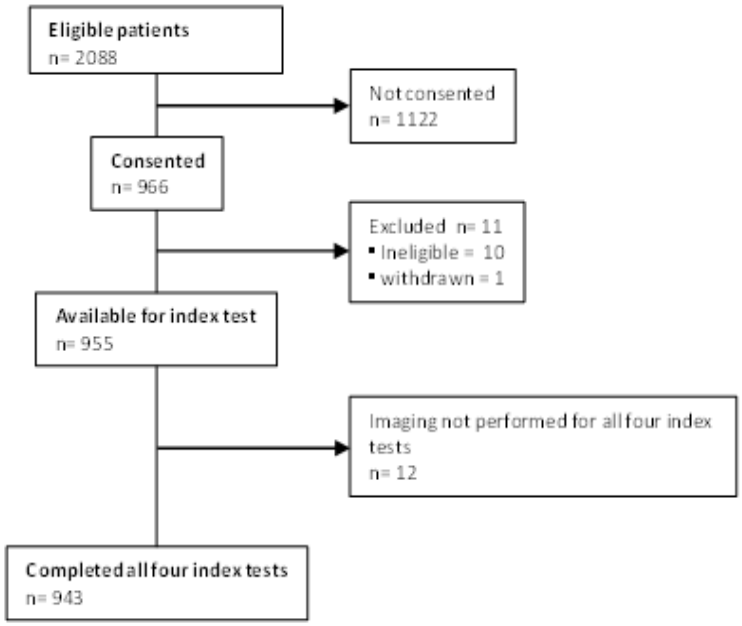


Figure 2 Consort style diagram showing recruitment to the study

Table 2 Characteristics of non-participants

Characteristic	Value
N	**1122**
Age[a] Mean – (SD)	61.7 (15.1)
Female – n (%)	592 (52.8)
Reasons for not taking part n (%)	
Screened but not sent information sheet	75 (6.7)
Refusal	265 (23.6)
Equipment malfunction	33 (2.9)
Missed	93 (8.3)
Non-attendance	134 (11.9)
Other reason	247 (22.0)
Reason not given	275 (24.5)

[a]Age calculated as (year of test-year of birth)

Table 3 Centre recruitment

Centre	Participants recruited n (%)
Aberdeen Royal Infirmary	353 (37.0)
Bedford Hospital	74 (7.7)
Hinchingbrooke Hospital NHS Trust, Huntingdon	343 (35.9)
Moorfields Eye Hospital, London	157 (16.4)
Royal Liverpool Hospital	28 (2.9)
TOTAL	**955**

Baseline characteristics of participants

Demographic characteristics of participants and non-participants were similar, with an average age slightly above 60 years of age (*Tables 2 and 4*) and similar gender distribution. Among participants, nearly 90% were of White British ethnicity (self reported ethnicity, *Table 4*).

Ocular characteristics recorded in the referral letter from the optometrist are detailed in Table 5. In the majority of referrals (77%), the optometrist had highlighted abnormalities in both eyes (referral eye). The average IOP at referral was 20 mmHg. Where the method of IOP measurement was reported on the referral letter (52%), the most commonly reported method of measurement was non-contact tonometry (NCT). Data on visual acuity and refractive error at referral are summarised in Table 5

Table 4 Baseline demographics of included participants

Characteristic	Value		
	All participants	**Glaucoma**	**Non -glaucoma**
N	**943**	**158**	**770**
Age Mean - (SD)	60.5(13.8)	67.4 (12.7)	59.2(13.6)
Female - n (%)	482 (51.1)	74 (46.8)	401 (52.1)
Ethnicity[a] – n (%)			
Black or Black-Caribbean	25 (2.7)	4 (2.5)	21 (2.7)
Black or Black British-African	20 (2.1)	6 (3.8)	14 (1.8)
Asian or Asian British-Indian	18 (1.9)	5 (3.2)	13 (1.7)
Asian or Asian British-Pakistani	4 (0.4)	0 (0)	4 (0.5)
Chinese	1 (0.1)	1 (0.6)	0 (0)
Other Asian Background	4 (0.4)	1 (0.6)	3 (0.4)
Mixed-White and Black African	1 (0.1)	1 (0.6)	0 (0)
White- British	826(89.2)	140 (88.6)	686(89.1)
Other	29 (3.1)	0 (0)	29 (3.8)

[a] There was no ethnicity recorded in 15 participants

Table 5 Ocular characteristics of participants at referral

Characteristic	Right eye	Left eye
Referral eye n/N (%)	97/939 (10.3)	116/939 (12.3)
Both eyes	725/939 (76.9)	
Not answered	1/939 (0.1)	
IOP on referral (mmHg) Mean (SD), n	19.6 (5.7), 918	19.9 (5.6), 918
Method Of IOP Assessment n/N (%)		
Non-contact tonometry	260/943 (27.6)	
Goldmann applanation tonometry	231/943 (24.5)	
Other[a]	452/943 (47.9)	
Refraction		
Mean sphere (dp) Mean (SD), n	0.4 (3.3), 571	1.0 (3.6), 561
Myopes greater than -5dp n/N (%)	37/943 (3.9)	36/943 (3.8)
Hyperopes greater than +5dp n/N (%)	38/943 (4.0)	51/943 (5.4)
Astigmatics greater than 3dp n/N (%)	16/943 (1.7)	16/943 (1.7)
Visual acuity Mean (SD)		
BCVA Snellen[b]	1.0 (0.3), 925	1.0 (0.3), 926
Log MAR[c]	0.0 (0.3), 925	0.0 (0.3), 926

[a] Includes those where the method of assessment was not recorded on referral

[b] Best corrected visual acuity, Snellen chart

[c] Logarithm of the Minimum Angle of Resolution chart

Reference standard diagnosis characteristics

Tables 6 to 14 describe the tests used to determine the reference standard, and the diagnoses in the GATE population. The average clinician IOP measured with GAT was similar to the referral IOP (*see Table 6*), and highest among patients with OHT and glaucoma (*see Table 7*). Visual field testing was outside the manufacturer recommended reliability in one quarter of participants. The average MD among those diagnosed with glaucoma and with reliable VF tests was -6.0 (SD 6.4) dB in the right eye and -7.5 (SD 6.8) dB in the left eye (*see Table 7*).

Table 8 displays the diagnosis of the GATE population per eye according to the agreed reference standard (*see Chapter 2*). The most common diagnosis

(approximately 40%) was “no glaucoma-related findings”. Glaucoma was diagnosed in about 11% of eyes. Co morbidities were uncommon, except for cataract that was reported in approximately 8% of eyes (*see Table 9*).

Among those eyes with glaucoma, mild disease was most prevalent (above half), while severe glaucoma was diagnosed in a relatively small proportion of eyes with the disease (28 of 219 eyes, 12.7%, *see Table 10*).

Over 1/3 of the GATE participants were discharged after the first visit (*see Table 11*). Table 13 describes the diagnosis by worst eye (ranked as shown) and by best eye. Glaucoma was diagnosed in at least one eye in 16.8% of the GATE cohort and 6.5% had glaucoma in both eyes at referral (*see Table 12*).

Table 6 Data from hospital eye service examination. Visual field and intra-ocular pressure

Characteristic	Right eye	Left eye
Visual field reliability[a] - n/N (%)		
Reliable	706/941 (75.0)	707/940 (75.2)
Unreliable	212/941 (22.5)	210/940 (22.3)
Not done	23/941 (2.4)	23/940 (2.4)
Reliable Visual field measures - Mean (SD), n		
MD (dB)	-1.9 (4.0), 703	-2.2 (4.1), 702
PSD (dB)	2.8 (2.6), 703	2.8 (2.6), 702
VFI (%)	95.0 (10.1), 688	94.9 (10.3), 682
Visual Field Measures Including Unreliable- Mean (SD), n		
MD (dB)	-1.8 (4.0), 893	-2.0 (4.1), 887
PSD (dB)	2.8 (2.5), 893	2.8 (2.5), 887
VFI (%)	95.0 (10.2), 866	95.0 (10.1), 859
IOP(mmHg) – Ophthalmologist GAT- Mean (SD), n		
	19.2 (5.1), 932	19.3 (5.1), 932

[a] Visual field reliability as defined by Humphrey VF output

Table 7 Data from hospital eye service examination. IOP and MD by diagnosis

	Right eye- Mean (SD), n	Left eye- Mean (SD), n

IOP (mmHg)GAT		
Glaucoma	23.0 (6.4), 116	22.6 (6.9), 103
Glaucoma Suspects	17.9 (4.4), 201	18.8 (5.2), 194
OHT	25.2 (3.5), 122	25.2 (3.1), 123
PAC/PAC Suspect	17.8 (4.1), 120	17.8 (3.8), 126
Normal	17.1 (3.2), 367	17.2 (3.1), 379
Reliable Visual Field MD (dB)		
Glaucoma	-6.0 (6.4), 85	-7.5 (6.8), 77
Glaucoma Suspects	-2.2 (3.4), 150	-2.2 (3.4), 153
OHT	-0.6 (2.2), 85	-0.8 (2.0), 92
PAC/PAC Suspect	-1.1 (3.0), 91	-1.4 (2.9), 89
Normal	-1.1 (3.0), 280	-1.3 (3.0), 279
All Visual Field MD (dB) including unreliable		
Glaucoma	-5.6 (6.1), 103	-7.2 (6.6), 89
Glaucoma Suspects	-2.2 (3.5), 195	-2.0 (3.3), 187
OHT	-0.3 (2.3), 113	-0.7 (2.1), 111
PAC/PAC Suspect	-0.9 (2.9), 115	-1.3 (2.9), 121
Normal	-1.1 (3.4), 352	-1.4 (3.4), 364

Table 8 Data from hospital eye service examination. Diagnosis

Diagnosis	Right Eye - n (%)	Left Eye - n (%)
N	**932**	**931**
Glaucoma	116 (12.4)	103 (11.1)
Disc Suspect	146 (15.6)	126 (13.5)
VF Suspect	29 (3.1)	35 (3.8)
VF+Disc Suspect	26 (2.8)	33 (3.5)
OHT	122 (13.0)	123 (13.2)
PAC	30 (3.2)	29 (3.1)
PAC Suspect	90 (9.6)	97 (10.4)
No glaucoma - related findings	367 (39.2)	379 (40.7)
Undetermined	6 (0.6)	6 (0.6)

Table 9 Data from hospital eye service examination. Co-morbidity

Co-morbidity	Right Eye n (%)	Left Eye n (%)
N	**936**	**936**
Age-related Macular Degeneration (AMD)	7 (0.7)	11 (1.2)
Cataract	78 (8.3)	70 (7.4)
Neurological	6 (0.6)	8 (0.8)
Other	65 (6.9)	63 (6.7)

Table 10 Data from hospital eye service examination. Glaucoma severity

Glaucoma Severity[a]	Right Eye - n (%)	Left Eye - n (%)
N	**116**	**103**
Mild	69 (59.5)	53 (51.5)
Moderate	31 (26.7)	29 (28.2)
Severe	11 (9.5)	17 (16.4)
Severity not recorded	5 (4.3)	4 (3.9)

[a] *see Chapter 2* for severity definitions

Table 11 Data from hospital eye service examination. Action after first consultation

Action	n (%)			
N	**933**			
Discharged – person level	357 (38.3)			
For those not discharged	**Right eye**		**Left eye**	
Treat	291	(31.2)	287	(30.8)
Monitor Only	214	(22.9)	216	(23.2)
Repeat Assessment Required	33	(3.5)	39	(4.1)
Not recorded	37	(4.0)	33	(3.5)

Table 12 Data from hospital eye service examination. Diagnosis by worst eye and best eye

	Worst eye		Best eye	
	n	**(%)**	**n**	**(%)**
N	**932**		**931**	
Diagnosis by clinician				
Glaucoma	158	(17.0)	61	(6.6)
Disc Suspect	170	(18.2)	102	(11.0)
VF Suspect	36	(3.9)	28	(3.0)
VF+ Disc Suspect	36	(3.9)	23	(2.5)
OHT	115	(12.3)	130	(14.0)
PAC	31	(3.3)	28	(3.0)
PAC Suspect	83	(8.9)	104	(11.2)
No glaucoma - related findings	299	(32.1)	447	(48.0)
Undetermined	4	(0.4)	8	(0.8)
Co-morbidity				
AMD	9	(1.0)	9	(1.0)
Cataract	75	(8.0)	73	(7.7)
Neurological	7	(0.7)	7	(0.7)
Other	68	(7.2)	60	(6.4)
Action				
Treat	320	(33.9)	258	(27.4)
Monitor Only	210	(22.3)	220	(23.3)
Repeat Assessment Required	39	(4.1)	33	(3.5)

Table 13 Data from hospital eye service examination. Severity of disease by worst and best eye for those diagnosed with glaucoma

	Worst eye n (%)		Best eye n(%)	
N	**158**		**61**	
Glaucoma severity				
Mild	78	(49.4)	19	(31.1)
Moderate	45	(28.5)	27	(44.3)
Severe	26	(16.5)	15	(24.6)
Severity not recorded	9	(5.7)	0	(0)

Table 14 Data from hospital eye service examination. Glaucoma mechanism for those diagnosed with glaucoma or glaucoma suspect, by worst and best eye

	Worst Eye	Best Eye
Clinical diagnosis n/N (%)		
Glaucoma	158/936 (16.8)	61/936 (6.5)
Open angle	123	46
Angle closure	26	12
Other	1	0
Missing	8	3
Disc Suspect	170/936 (18.0)	102/936 (10.8)
Open angle	150	94
Angle closure	11	6
Other	2	0
Missing	7	2
VF Suspect	36/936 (3.8)	28/936 (3.0)
Open angle	27	21
Angle closure	6	5
Other	1	2
Missing	2	0
VF+ Disc Suspect	36/936 (3.8)	23/936 (2.4)
Open angle	33	21
Angle closure	3	2
Other	0	0
Missing	0	0

Chapter 4 Diagnostic analysis results

Overview

This chapter reports the results of the diagnosis analyses which aimed to assess the diagnostic performance of the four imaging tests (HRT-MRA, HRT-GPS, GDx and OCT) and the other outcomes associated with the imaging tests (indeterminacy and participant preference). Results of the triage analyses are provided in the following chapter. The specific diagnostic performance analyses covered in this chapter are the *default diagnosis analysis*, five sensitivity analyses (*diagnosis sensitivity analyses 1-5*) and the use of a combination of the imaging tests (*Combination of tests analysis – .see Table 15)* for a list with definitions. The default analysis was defined as one where the reference standard definition of disease was a clinical diagnosis of glaucoma only, the imaging test definition of an abnormal result was "outside normal limits" for the overall classification of the respective imaging test *(see Chapter 2).*

Additionally, only cases where there was a good quality image with an overall classification available were included *(see Chapter 2).* The five sensitivity analyses assessed the impact of varying assumptions made in the default analysis relating to the reference standard definition of disease (including all types of glaucoma suspects as diseased), the definition of an abnormal test result (including borderline results as abnormal), and how cases where the test did not produce an overall classification were handled in the analysis. There were four test related reasons why an overall classification may not have been available in addition to missing data *(see no results categories, Table 15).* Sensitivity analyses assessed the impact of removing the requirement of a "good" quality image and using the provided assessment, along with setting other cases which did not produce an overall classification result as abnormal.

The combination of test analysis investigated using pairs of imaging tests to produce a composite imaging test result, under the same assumptions as the default analysis. Given the findings of the default and sensitivity analyses, only three pairs of test combinations were evaluated; HRT-MRA with each of the other tests respectively. For all analyses, a STARD flow diagram[34,37] was produced which shows the flow of participants. The subset of participants who received all four tests and were

considered in the statistical analyses are separated out into three groups according to whether each imaging test result was “abnormal”, “normal”, or “no result” (the imaging test result was not available whether because the test was inconclusive or the result was missing). For each of these three groups the group status according to the reference standard (“glaucoma present” or “glaucoma absent”) for each participant is given or alternative the reference standard was stated to be missing or inconclusive. The final categorisations of the imaging test result by reference standard status provides the four possible combinations (true and false positive, false and true negative) from which the diagnostic performance can be assessed. Sensitivity, specificity, likelihood ratios and DOR are provided with associated 95% CIs summarised for each analysis.

Of the 966 (46%) who agreed to take part in GATE, 11 were excluded from the study: 10 were ineligible and 1 person withdrew prior to participating in the study. Additionally, due to administrative and research processing errors, imaging was not implemented for all 4 imaging tests in 12 participants and these were excluded from all analyses. The analyses in this chapter pertain to the remaining 943 participants. Of these, no reference standard finding was available for 11 participants with an inconclusive finding in a further 4 cases.

Table 15 Diagnosis analyses

Analysis	**Reference standard definition of disease**	**Abnormal test result**	**Handling of 'no result' categories**	**Figure no.**	**Table no.**
Default diagnostic analysis	Glaucoma in the "worse" eye	Outside normal limits	A-E excluded	3	16, 17, 18, 19
Diagnosis sensitivity analysis 1	Glaucoma in the "worse" eye	Outside normal limits or borderline	A-E excluded	4	22
Diagnosis sensitivity analysis 2	Glaucoma or glaucoma suspect in the "worse" eye	Outside normal limits	A-E excluded	5	23
Diagnosis sensitivity analysis 3	Glaucoma or glaucoma suspect in the "worse" eye	Outside normal limits or borderline	A-E excluded	6	24
Diagnosis sensitivity analysis 4	Glaucoma or glaucoma suspect in the "worse" eye	Outside normal limits or borderline	A imaging classification B-D abnormal E excluded	7	25
Diagnosis sensitivity analysis 5	Glaucoma in the "worse" eye	Outside normal limits	A imaging classification B-D abnormal E excluded	8	26

Diagnosis sensitivity analysis 6	Glaucoma in the "best" eye	Outside normal limits	A-E excluded	9	27
Combinations of diagnosis imaging tests	Glaucoma in the "worse" eye	Outside normal limits	A-E excluded	10	28
No result categories *A: test performed and imaging report produced but quality is lower than manufacturer quality cut-off* *B: test performed and imaging report produced but no overall classification generated by machine* *C: test performed but there was a clear imaging artefact on the report* *D: test attempted but no imaging could be acquired from the patient's eyes – no report generated* *E: missing imaging output (due to study related or data collection issues)*					

Default diagnosis analysis

The results for the default diagnosis analysis are presented in three sections:

- diagnostic performance of the imaging tests
- paired comparisons of imaging tests
- diagnostic performance with restricted reference standard definition of disease.

Diagnostic performance of the imaging tests

For the default analysis, abnormal imaging test results were those classified as 'outside normal limits' and the corresponding reference standard definition of disease was a diagnosis of glaucoma in the 'worse' eye. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The flow of study participants according to the default diagnosis analysis is shown in Figure 3 with respective number of abnormal, normal and 'no result' cases by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 158 were classified as disease positive and 770 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 16. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity (87.0%, 95% CI (80.2,92.1)) but lowest specificity (63.9%, 95% CI (60.2,67.4)), GDx had the lowest sensitivity (35.1%, 95% CI (27.0,43.8)) but the highest specificity (97.2%, 95% CI (95.6,98.3)), and the other two tests provided intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had very similar sensitivity and specificity values). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 9.24 for HRT-GPS to 18.48 for GDx.

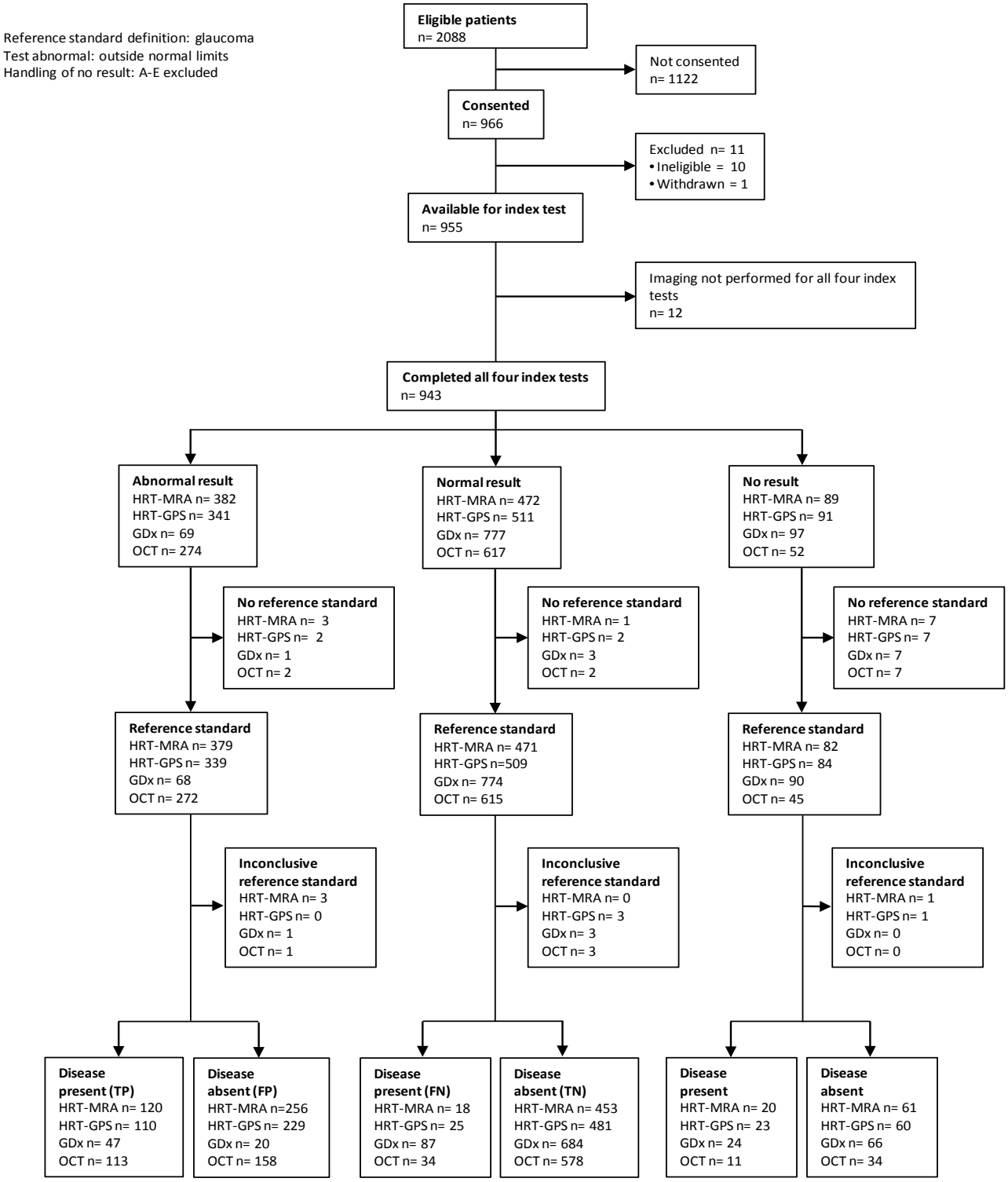


Figure 3 Flow diagram: default diagnostic analysis

Table 16 Diagnostic performance: default diagnosis analysis

Test	Diagnostic parameter	Point Estimate	95% CI Lower	95% CI Upper
HRT-MRA	Sensitivity %	87.0	80.2	92.1
	Specificity %	63.9	60.2	67.4
	Positive Likelihood Ratio	2.41	2.14	2.71
	Negative Likelihood Ratio	0.20	0.13	0.32
	DOR	11.80	7.02	19.81
HRT-GPS	Sensitivity %	81.5	73.9	87.6
	Specificity %	67.7	64.2	71.2
	Positive Likelihood Ratio	2.53	2.21	2.89
	Negative Likelihood Ratio	0.27	0.19	0.39
	DOR	9.24	5.82	14.67
GDx	Sensitivity %	35.1	27.0	43.8
	Specificity %	97.2	95.6	98.3
	Positive Likelihood Ratio	12.35	7.57	20.14
	Negative Likelihood Ratio	0.67	0.59	0.76
	DOR	18.48	10.46	32.63
OCT	Sensitivity %	76.9	69.2	83.4
	Specificity %	78.5	75.4	81.4
	Positive Likelihood Ratio	3.58	3.04	4.22
	Negative Likelihood Ratio	0.29	0.22	0.40
	DOR	12.16	7.97	18.54

Paired comparisons of imaging tests

Table 17 shows the paired difference (with 95% CI) and corresponding McNemar's test p-value for comparisons between pairs of tests. There was evidence that the sensitivity of all tests differed from each other except for HRT-GPS versus OCT.

HRT-MRA and GDx had the highest and lowest sensitivity respectively. Differences varied from -6.7% (HRT-GPS versus HRT-MRA) to 55.6% (HRT-MRA versus GDx). Similarly there was evidence that all specificities of all tests varied from each other (according to McNemar's test);[37] the 95% paired difference CI for HRT-GPS versus HRT-MRA just overlapped with zero.

Table 17 Paired comparisons of sensitivity and specificity between the imaging tests

Tests Compared	Parameter	Test	Value (95% CI)	P-value (McNemar's)
HRT-GPS vs GDx	Sensitivity	HRT-GPS	81.1 (74.2 to 88.1)	
		GDx	34.4 (26.0 to 42.9)	
		Difference	46.7 (37.0 to 54.9)	<.001
	Specificity	HRT-GPS	67.5 (64.0 to 71.1)	
		GDx	97.5 (96.3 to 98.7)	
		Difference	-30.0 (-33.6 to -26.3)	<.001
GDx vs OCT	Sensitivity	GDx	36.4 (28.1 to 44.7)	
		OCT	77.5 (70.3 to 84.7)	
		Difference	-41.1 (-49.2 to -31.6)	<.001
	Specificity	GDx	97.5 (96.3 to 98.7)	
		OCT	79.8 (76.8 to 82.8)	
		Difference	17.7 (14.9 to 20.8)	<.001
GDx vs HRT-MRA	Sensitivity	GDx	33.1 (24.8 to 41.3)	
		HRT-MRA	88.7 (83.1 to 94.3)	
		Difference	-55.6 (-63.8 to -45.6)	<.001
	Specificity	GDx	97.3 (96.1 to 98.5)	
		HRT-MRA	63.7 (60.1 to 67.4)	
		Difference	33.6 (29.8 to 37.3)	<.001
HRT-GPS vs HRT-MRA	Sensitivity	HRT-GPS	81.3 (74.7 to 87.9)	
		HRT-MRA	88.1 (82.6 to 93.5)	
		Difference	-6.7 (-13.2 to -0.6)	<.001
	Specificity	HRT-GPS	67.8 (64.3 to 71.3)	
		HRT-MRA	64.1 (60.5 to 67.6)	
		Difference	3.7 (-0.1 to 7.5)	<.001
HRT-MRA vs OCT	Sensitivity	HRT-MRA	86.5 (80.7 to 92.3)	
		OCT	75.2 (67.8 to 82.5)	
		Difference	11.3(3.4 to 19.2)	<.001
	Specificity	HRT-MRA	63.9 (60.3 to 67.5)	
		OCT	79.4 (76.4 to 82.4)	

Tests Compared	Parameter	Test	Value (95% CI)	P-value (McNemar's)
		Difference	-15.5 (-19.8 to -11.2)	<.001
HRT-GPS vs OCT	Sensitivity	HRT-GPS	82.3 (75.7 to 88.9)	
		OCT	75.4 (68.0 to 82.8)	
		Difference	6.9(-1.6 to 15.4)	0.106
	Specificity	HRT-GPS	67.7 (64.2 to 71.2)	
		OCT	79.7 (76.7 to 82.7)	
		Difference	-12.0 (-16.3 to -7.6)	<.001

Impact of severity of disease

Two further analyses looked at the impact of changing the reference standard definition of disease to moderate and severe glaucoma only and severe glaucoma (*see Chapter 2 for disease definitions*). The only change from the default analysis was in terms of the reference standard. The diagnostic performance for the four imaging tests where the reference standard definition of disease was moderate and severe glaucoma only is given in Table 18.

Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-GPS had the highest sensitivity (92.7%, 95% CI (82.4, 98.0)) but with the second lowest specificity (63.5%, 95% CI (60.1, 66.9)), GDx had the lowest sensitivity (60.0%, 95% CI 45.9,73.0)) but the highest specificity (95.7%, 95% CI (94.0,97.0)), and the other two tests provided intermediate results (HRT-MRA values were very similar to the HRT-GPS results, and OCT had a similar sensitivity but higher specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 12.44 for HRT-MRA to 33.04 for GDx. Compared to the default analysis, the diagnostic performance of GDx and OCT both performed relatively better and HRT-GPS and HRT-MRA relatively poorer.

Table 18 Diagnostic performance: default diagnosis analysis (reference standard definition of disease of moderate and severe glaucoma)

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	89.7	78.8	96.1
	Specificity %	58.9	55.4	62.4
	Positive Likelihood Ratio	2.18	1.93	2.46
	Negative Likelihood Ratio	0.18	0.08	0.38
	DOR	12.44	5.28	29.30
HRT-GPS	Sensitivity %	92.7	82.4	98.0
	Specificity %	63.5	60.1	66.9
	Positive Likelihood Ratio	2.54	2.26	2.86
	Negative Likelihood Ratio	0.11	0.04	0.29
	DOR	22.22	7.95	62.12
GDx	Sensitivity %	60.0	45.9	73.0
	Specificity %	95.7	94.0	97.0
	Positive Likelihood Ratio	13.82	9.32	20.47
	Negative Likelihood Ratio	0.42	0.30	0.58
	DOR	33.04	17.43	62.65
OCT	Sensitivity %	89.1	78.8	95.5
	Specificity %	73.9	70.7	76.9
	Positive Likelihood Ratio	3.41	2.95	3.94
	Negative Likelihood Ratio	0.15	0.07	0.30
	DOR	23.02	10.34	51.25

The diagnostic performance for the four imaging tests where the reference standard definition of disease was severe glaucoma only is given in Table 19. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: OCT had the highest sensitivity (95.2%, 95% CI (76.2,99.9)) and the second highest specificity (70.9%, 95% CI (67.7, 73.9)), GDx had the lowest sensitivity (78.9%, 95% CI (54.4,93.9)) but the highest specificity (93.7%, 95% CI (91.8,95.2)), and the other two tests providing intermediate results (HRT-GPS and HRT-MRA results were very similar and had a similar sensitivity to OCT though lower a specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0) though not always ruling out the disease. DORs ranged from 23.63 for HRT-MRA to 55.31 for OCT. Compared to the default analysis, the diagnostic performance of GDx and OCT both performed relatively better and HRT-GPS and HRT-MRA relatively poorer.

Table 19 Diagnostic performance: default diagnosis analysis (reference standard definition of disease of severe glaucoma)

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	94.7	74.0	99.9
	Specificity %	56.8	53.3	60.2
	Positive Likelihood Ratio	2.19	1.92	2.50
	Negative Likelihood Ratio	0.09	0.01	0.63
	DOR	23.63	3.14	177.85
HRT-GPS	Sensitivity %	94.7	74.0	99.9
	Specificity %	61.1	57.7	64.5
	Positive Likelihood Ratio	2.44	2.13	2.79
	Negative Likelihood Ratio	0.09	0.01	0.58
	DOR	28.32	3.76	213.16
GDx	Sensitivity %	78.9	54.4	93.9
	Specificity %	93.7	91.8	95.2
	Positive Likelihood Ratio	12.43	8.75	17.66
	Negative Likelihood Ratio	0.22	0.09	0.54
	DOR	55.31	3.76	172.63
OCT	Sensitivity %	95.2	76.2	99.9
	Specificity %	70.9	67.7	73.9
	Positive Likelihood Ratio	3.27	2.84	3.77
	Negative Likelihood Ratio	0.07	0.01	0.2
	DOR	48.69	6.50	364.73

Other outcomes

Indeterminacy results are shown in Table 20. GDx had the highest percentage of low quality imaging results followed by HRT-GPS and HRT-MRA, with OCT giving the lowest percentage number of low quality results.

Table 20 Classification and quality of imaging results (default analysis)

Class	**HRT-MRA n (%) N=943**	**HRT-GPS n (%) N=943**	**GDx n (%) N=943**	**OCT n (%) N=943**
Normal	319 (33.8)	310 (32.9)	640 (67.9)	447 (47.4)
Borderline	153 (16.2)	201 (21.3)	137 (14.5)	170 (18.0)
Abnormal	382 (40.5)	341 (36.2)	69 (7.3)	274 (29.1)
Indeterminacy (No result categories A-D)	58 (6.3)	75 (8.0)	79 (8.4)	40 (4.2)
Missing data (No result category E)	31 (3.2)	16 (1.7)	18 (1.9)	12 (1.3)
Quality[a]	**N=887**	**N=887**	**N=907**	**N=906**
Good quality	854 (96.3)	852 (96.1)	846 (93.3)	891 (98.3)
Low quality	33 (3.7)	35 (3.9)	61 (6.7)	15 (1.7)

[a] excluding 'no result' categories B-E

Table 21 shows the participants' preference ranking of imaging tests (HRT-GPS and HRT-MRA have the same results), time taken to conduct the test and the proportion which received dilation also shown. Participant preference was collected for 890 participants (94%). Almost half of responders (48.2%) had no preference. Of those participants who gave a preference, OCT was ranked as most preferred (27.6%) following by GDx (11.9%) and HRT-GPS/HRT-MRA with the lowest preference (5.1%). Average time taken to perform the test varied from 5.2 (OCT) to 7.6 (HRT-GPS/HRT-MRA) minutes. More participants received dilation under HRT-GPS/HRT-MRA (2.2%) than the other two tests. No adverse events were reported during the study.

Table 21 Participant preference, test conduct time and dilation results

Test		Test Preference N=890	Test conduct time (minutes) Mean (SD)	Dilation N=918
HRT(MRA/GPS)	1	49 (5.1)	N=900	20 (2.2)
	2	150 (15.6)	7.6 (5.0)	-
	3	229 (23.9)	-	-
GDx	1	114 (11.9)	N=886	16 (1.7)
	2	162 (16.9)	7.5 (5.1)	-
	3	152 (15.8)	-	-
OCT	1	265 (27.6)	N=904	6 (0.7)
	2	116 (12.1)	5.2 (3.0)	-
	3	44 (4.6)	-	-
All	None	462 (48.2)	-	-

Diagnosis sensitivity analysis 1

Diagnosis sensitivity analysis 1 differed from the default analysis in that a borderline finding on the imaging test was also classified as an abnormal result.

For diagnosis sensitivity analysis 1, abnormal imaging test results were those classified as 'outside normal limits' and 'borderline' and the corresponding reference standard definition of disease was a diagnosis of glaucoma in the 'worse' eye. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The flow of study participants according to sensitivity analysis 1 is shown in Figure 4 with the respective number of abnormal, normal and 'no result' cases given by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 158 were classified as disease positive, and 770 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 22. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity (94.9%, 95% CI (89.8,97.9)) but the second lowest specificity (43.9%,

95% CI (40.2,47.6)), GDx had the lowest sensitivity (60.4%, 95% CI (51.6,68.8)) but the highest specificity (82.8%, 95% CI (79.8,85.5)), and the other two tests providing intermediate results (HRT-GPS values were very similar to the HRT-MRA results though marginally lower, and OCT had a high sensitivity and moderate specificity in relation to the other tests). Sensitivity was higher for all tests than under the default analysis but with correspondingly lower specificity. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 7.36 for GDx to 14.62 for HRT-MRA.

Reference standard definition: glaucoma
Test abnormal: outside normal limits or borderline
Handling of no result: A-E excluded

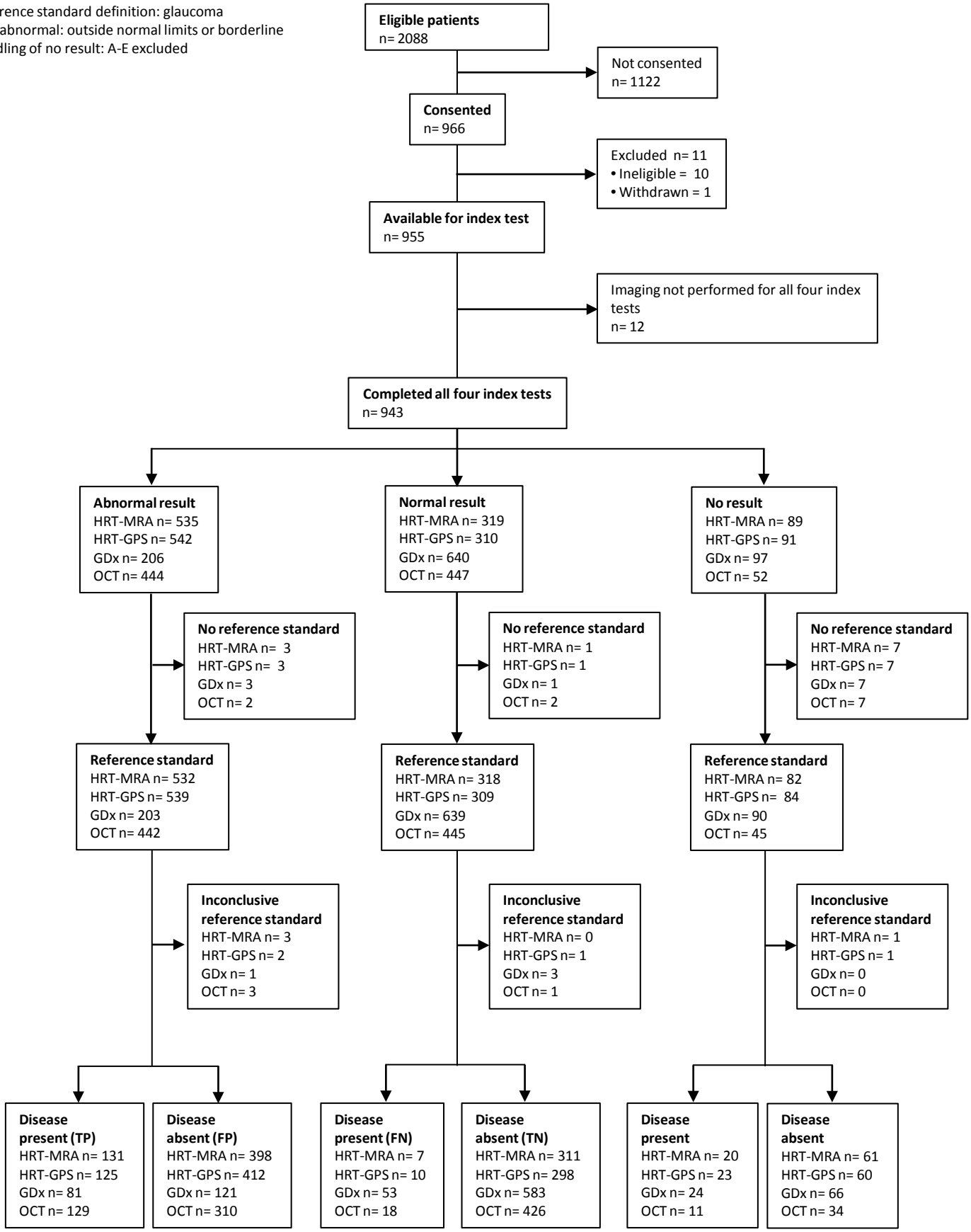


Figure 4 Flow diagram: diagnostic sensitivity analysis 1

Table 22 Diagnostic performance: diagnosis sensitivity analysis 1

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	94.9	89.8	97.9
	Specificity %	43.9	40.2	47.6
	Positive Likelihood Ratio	1.69	1.57	1.82
	Negative Likelihood Ratio	0.12	0.06	0.24
	DOR	14.62	6.74	31.73
HRT-GPS	Sensitivity %	92.6	86.8	96.4
	Specificity %	42.0	38.3	45.7
	Positive Likelihood Ratio	1.60	1.47	1.73
	Negative Likelihood Ratio	0.18	0.10	0.32
	DOR	9.04	4.67	17.51
GDx	Sensitivity %	60.4	51.6	68.8
	Specificity %	82.8	79.8	85.5
	Positive Likelihood Ratio	3.52	2.84	4.35
	Negative Likelihood Ratio	0.48	0.39	0.59
	DOR	7.36	4.95	10.96
OCT	Sensitivity %	87.8	81.3	92.6
	Specificity %	57.9	54.2	61.5
	Positive Likelihood Ratio	2.08	1.88	2.31
	Negative Likelihood Ratio	0.21	0.14	0.33
	DOR	9.85	5.89	16.49

Diagnosis sensitivity analysis 2

Diagnosis sensitivity analysis 2 differed from the default analysis in that the reference standard definition of disease incorporated all glaucoma suspects (irrespective of type). For diagnosis sensitivity analysis 2, abnormal imaging test results were those classified as 'outside normal limits' and the corresponding reference standard definition of disease was a diagnosis of glaucoma in the 'worse' eye. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The flow of study participants according to sensitivity analysis 2 is shown in Figure 5 with respective number of abnormal, normal and 'no result' cases by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 400 were classified as disease positive and 528 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 23. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity (74.0%, 95% CI (69.1,78.5)) but lowest specificity (76.5%, 95% CI (72.5,80.1)), GDx had the lowest sensitivity (16.5%, 95% CI (12.8,20.8)) but the highest specificity (98.2%, 95% CI (96.5,99.2)), and the other two tests providing intermediate results (HRT-GPS had lower sensitivity than HRT-MRA but a slightly higher specificity, and OCT had the second lowest sensitivity but the second highest specificity values). Sensitivity was lower for all tests than under the default analysis but with correspondingly higher specificity. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 5.44 for OCT to 10.51 for GDx.

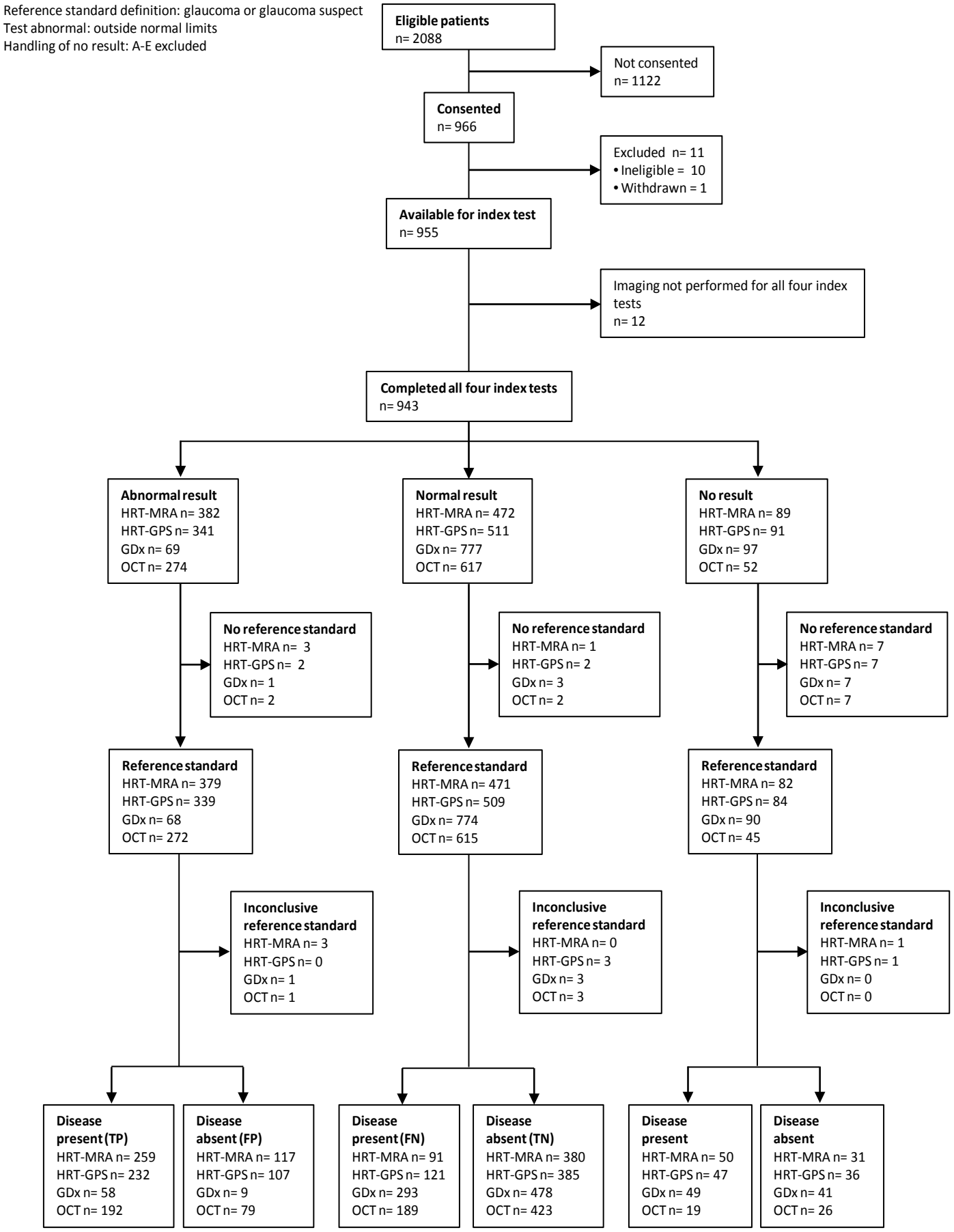


Figure 5 Flow diagram: diagnostic sensitivity analysis 2

Table 23 Diagnostic performance: diagnosis sensitivity analysis 2

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	74.0	69.1	78.5
	Specificity %	76.5	72.5	80.1
	Positive Likelihood Ratio	3.14	2.65	3.73
	Negative Likelihood Ratio	0.34	0.28	0.41
	DOR	9.24	6.74	12.68
HRT-GPS	Sensitivity %	65.7	60.5	70.7
	Specificity %	78.3	74.3	81.8
	Positive Likelihood Ratio	3.02	2.51	3.63
	Negative Likelihood Ratio	0.44	0.38	0.51
	DOR	6.90	5.08	9.38
GDx	Sensitivity %	16.5	12.8	20.8
	Specificity %	98.2	96.5	99.2
	Positive Likelihood Ratio	8.94	4.49	17.80
	Negative Likelihood Ratio	0.85	0.81	0.89
	DOR	10.51	5.13	21.54
OCT	Sensitivity %	50.4	45.3	55.5
	Specificity %	84.3	80.8	87.3
	Positive Likelihood Ratio	3.20	2.56	4.01
	Negative Likelihood Ratio	0.59	0.53	0.66
	DOR	5.44	3.98	7.44

Diagnosis sensitivity analysis 3

Diagnosis sensitivity analysis 3 differed from the default analysis in that a borderline finding on the imaging test was classified as an abnormal test result and the reference standard definition of disease incorporated all glaucoma suspects (irrespective of type).

For diagnosis sensitivity analysis 3, abnormal imaging test results were those classified as 'outside normal limits' or 'borderline' and the corresponding reference standard definition of disease was a diagnosis of glaucoma or glaucoma suspect in the 'worse' eye. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The flow of study participants according to sensitivity analysis 3 is shown in Figure 6 with respective number of abnormal, normal and 'no result' cases by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 400 were classified as disease positive and 528 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 24. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity 88.9%, 95% CI (85.1,92.0)) but the second lowest specificity (56.1%, 95% CI (51.6,60.6)), GDx had the lowest sensitivity (39.0%, 95% CI (33.9,44.4)) but the highest specificity (86.7%, 95% CI (83.3,89.5)), and the other two tests providing intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had very similar sensitivity and specificity values). Sensitivity was slightly higher for GDx, HRT-GPS and HRT-MRA than under the default analysis but with correspondingly lower specificity. OCT however had a slightly lower sensitivity and specificity than under the default analysis.

Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 4.04 for OCT to 10.21 for HRT-MRA.

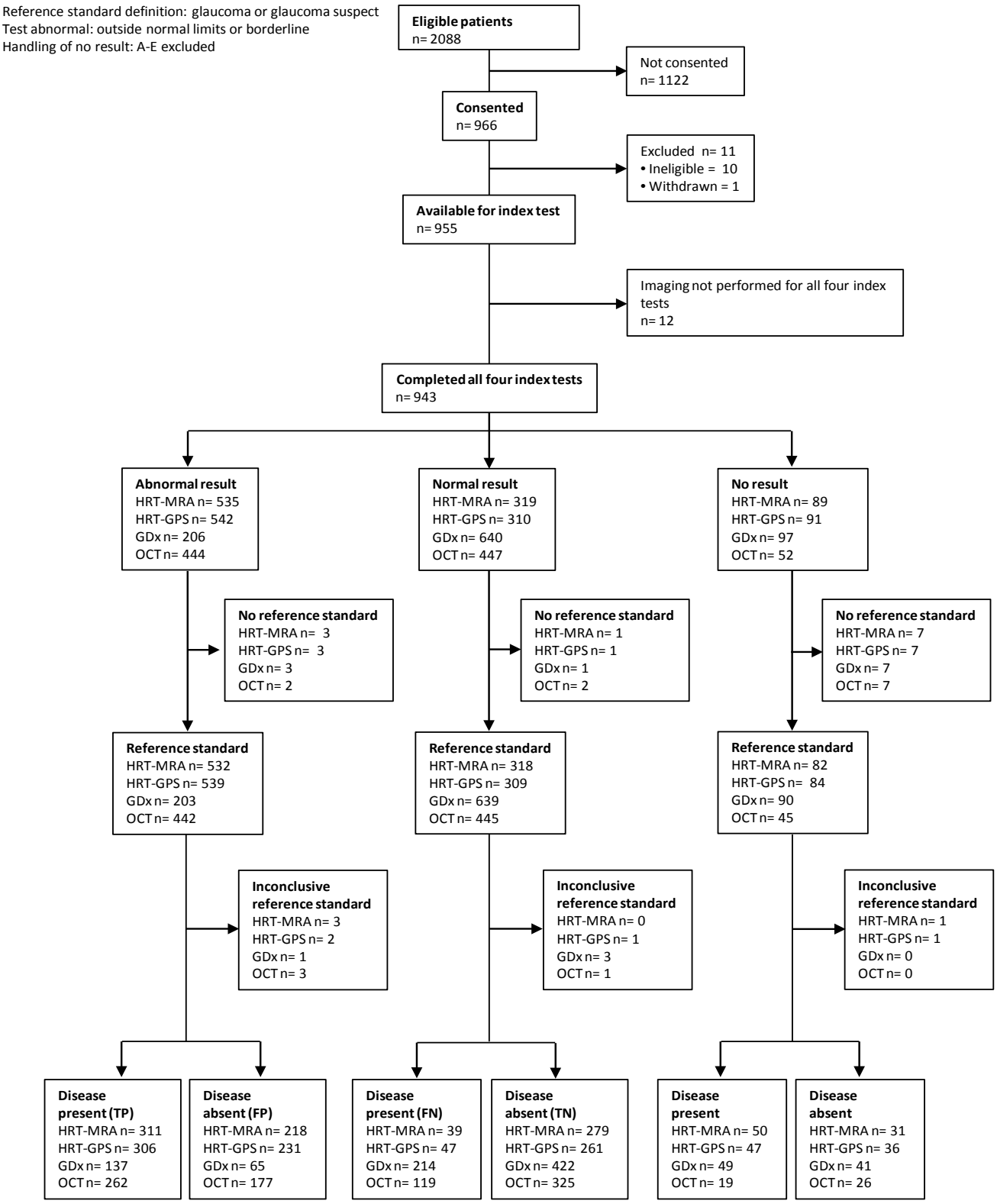


Figure 6 Flow diagram: diagnostic sensitivity analysis 3

Table 24 Diagnostic performance: diagnosis sensitivity analysis 3

Test	Diagnostic parameter	Point Estimate	95% CI Lower	95% CI Upper
HRT-MRA	Sensitivity %	88.9	85.1	92.0
	Specificity %	56.1	51.6	60.6
	Positive Likelihood Ratio	2.03	1.82	2.25
	Negative Likelihood Ratio	0.20	0.15	0.27
	DOR	10.21	7.00	14.88
HRT-GPS	Sensitivity %	86.7	82.7	90.1
	Specificity %	53.0	48.5	57.5
	Positive Likelihood Ratio	1.85	1.67	2.05
	Negative Likelihood Ratio	0.25	0.19	0.33
	DOR	7.36	5.16	10.49
GDx	Sensitivity %	39.0	33.9	44.4
	Specificity %	86.7	83.3	89.5
	Positive Likelihood Ratio	2.92	2.25	3.80
	Negative Likelihood Ratio	0.70	0.64	0.77
	DOR	4.16	2.96	5.83
OCT	Sensitivity %	68.8	63.8	73.4
	Specificity %	64.7	60.4	68.9
	Positive Likelihood Ratio	1.95	1.70	2.24
	Negative Likelihood Ratio	0.48	0.41	0.57
	DOR	4.04	3.04	5.37

Diagnosis sensitivity analysis 4

Diagnosis sensitivity analysis 4 has the same reference standard and definition of an abnormal imaging test as sensitivity analysis 3 differing by including the imaging test related 'no result' cases (the overall classification was used irrespective of the quality indicator and the types were all classified as abnormal).

For diagnosis sensitivity analysis 4, abnormal imaging test results were those classified as 'outside normal limits' or 'borderline' and the corresponding reference standard definition of disease was a diagnosis of glaucoma or glaucoma suspect in the 'worse' eye. The analysis included participants with a low quality imaging output if a classification was given; other imaging test results which did not provide an overall classification were included as abnormal .

The flow of study participants according to sensitivity analysis 4 is shown in Figure 7 with respective number of abnormal, normal and 'no result' cases, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 400 were classified as disease positive and 528 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 25.

Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity (89.2%, 95% CI (85.7,92.1)) but second lowest specificity (55.1%, 95% CI (50.7,59.5)), GDx had the lowest sensitivity (41.9%, 95% CI (37.0,47.0))) but the highest specificity (85.6%, 95% CI (82.3,88.5)), and the other two tests providing intermediate results (HRT-GPS values were similar to the HRT-MRA results, and OCT had similar sensitivity and specificity values). Sensitivity was higher for all tests than under the default analysis but with correspondingly lower specificity. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 3.89 for OCT to 10.19 for HRT-MRA.

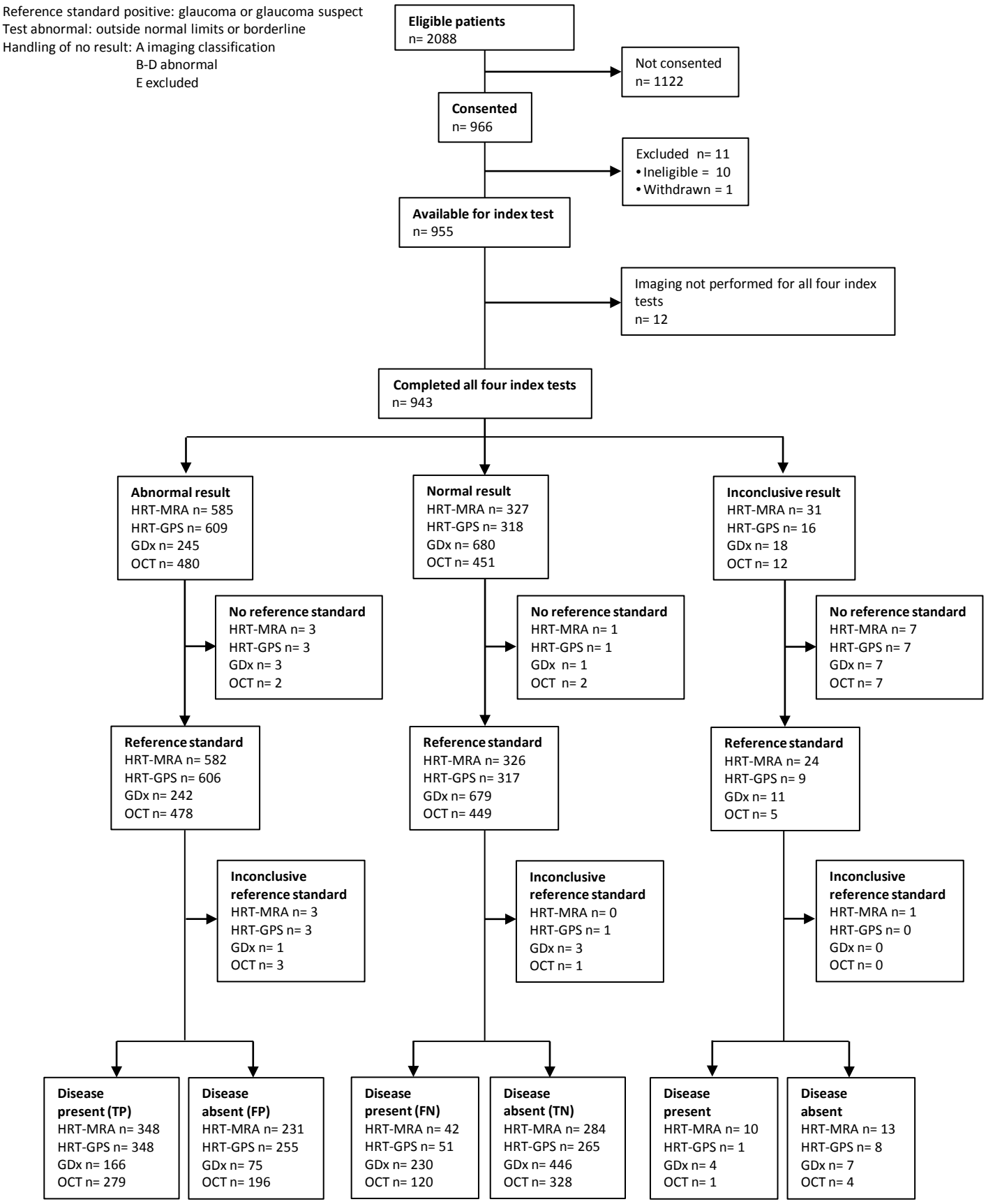


Figure 7 Flow diagram: diagnostic sensitivity analysis 4

Table 25 Diagnostic performance: diagnosis sensitivity analysis 4

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	89.2	85.7	92.1
	Specificity %	55.1	50.7	59.5
	Positive Likelihood Ratio	1.99	1.80	2.20
	Negative Likelihood Ratio	0.20	0.15	0.26
	DOR	10.19	7.08	14.66
HRT-GPS	Sensitivity %	87.2	83.5	90.3
	Specificity %	51.0	46.6	55.3
	Positive Likelihood Ratio	1.78	1.62	1.96
	Negative Likelihood Ratio	0.25	0.19	0.33
	DOR	7.09	5.04	9.97
GDx	Sensitivity %	41.9	37.0	47.0
	Specificity %	85.6	82.3	88.5
	Positive Likelihood Ratio	2.91	2.29	3.70
	Negative Likelihood Ratio	0.68	0.62	0.74
	DOR	4.29	3.13	5.89
OCT	Sensitivity %	69.9	65.2	74.4
	Specificity %	62.6	58.3	66.7
	Positive Likelihood Ratio	1.87	1.64	2.12
	Negative Likelihood Ratio	0.48	0.41	0.57
	DOR	3.89	2.95	5.14

Diagnosis sensitivity analysis 5

Diagnosis sensitivity analysis 5 differed from the default analysis in that the imaging test related ‘no result’ cases were all classified as an abnormal result.

For sensitivity analysis 5, abnormal imaging test results were those classified as ‘outside normal limits’ and the corresponding reference standard definition of disease was a diagnosis of glaucoma in the ‘worse’ eye. The analysis included participants with a low quality imaging output if a classification was given; other imaging test results which did not provide an overall classification were included as abnormal .

The flow of study participants according to sensitivity analysis 5 is shown in Figure 8 with respective number of abnormal, normal and ‘no result’ cases by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 158 were classified as disease positive and 770 as disease negative. The reference standard was missing and inconclusive for 11 and 4 participants respectively. The diagnostic performance for the four tests is given in Table 26. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA had the highest sensitivity (87.3%, 95% CI (81.0,92.0)) but lowest specificity (61.8%, 95% CI (58.2,65.3)), GDx had the lowest sensitivity (37.6%, 95% CI (30.0,45.7)) but the highest specificity (95.4%, 95% CI (93.7,96.8)), and the other two tests providing intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had very similar sensitivity and specificity values). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 8.96 for HRT-GPS to 12.47 for GDx.

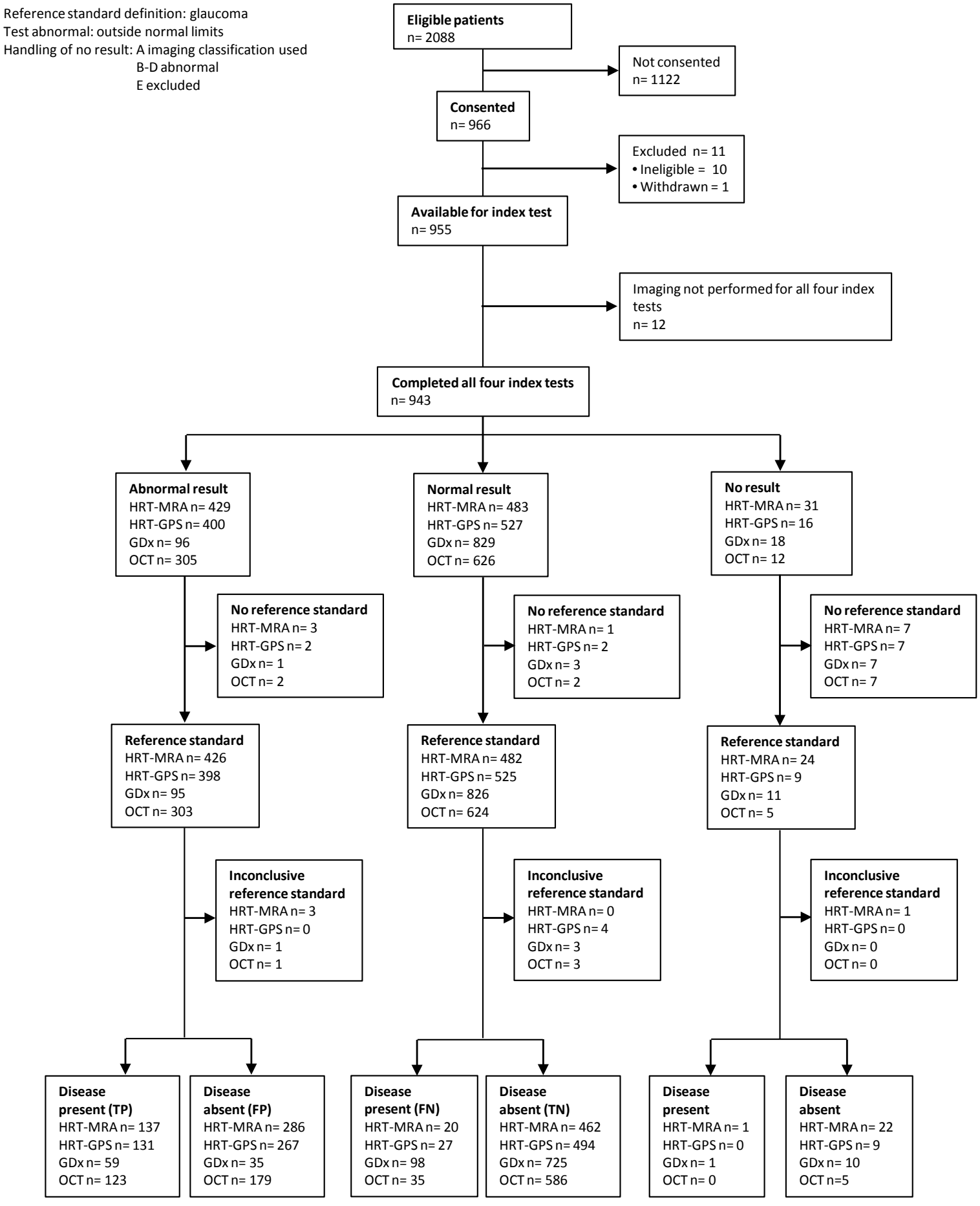


Figure 8 Flowdiagram: diagnostic sensitivity analysis 5

Table 26 Diagnostic performance: diagnosis sensitivity analysis 5

Test	Diagnostic parameter	Point Estimate	95% CI Lower	Upper
HRT-MRA	Sensitivity %	87.3	81.0	92.0
	Specificity %	61.8	58.2	65.3
	Positive Likelihood Ratio	2.28	2.05	2.54
	Negative Likelihood Ratio	0.21	0.14	0.31
	DOR	11.07	6.77	18.09
HRT-GPS	Sensitivity %	82.9	76.1	88.4
	Specificity %	64.9	61.4	68.3
	Positive Likelihood Ratio	2.36	2.10	2.66
	Negative Likelihood Ratio	0.26	0.19	0.37
	DOR	8.96	5.78	13.94
GDx	Sensitivity %	37.6	30.0	45.7
	Specificity %	95.4	93.7	96.8
	Positive Likelihood Ratio	8.16	5.57	11.95
	Negative Likelihood Ratio	0.65	0.58	0.74
	DOR	12.47	7.81	19.2
OCT	Sensitivity %	77.8	70.6	84.1
	Specificity %	76.6	73.4	80.0
	Positive Likelihood Ratio	3.33	2.86	3.88
	Negative Likelihood Ratio	0.29	0.22	0.39
	DOR	11.50	7.63	17.35

Diagnosis sensitivity analysis 6

Diagnosis sensitivity analysis 6 differed from the default analysis in that the diagnosis of the participants' 'best' eye according to the reference standard was used. Abnormal imaging test results were those classified as 'outside normal limits' and the corresponding reference standard definition of disease was a diagnosis of glaucoma. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The flow of study participants according to sensitivity analysis 6 is shown in Figure 9 with respective number of abnormal, normal and 'no result' cases by imaging test, and the corresponding reference standard finding shown. Of the 943 patients for whom all four tests were performed, 61 were classified as disease positive and 862 as disease negative. The reference standard was missing and inconclusive for 12 and 8 participants respectively. The diagnostic performance for the four tests is given in Table 27. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-GPS had the highest sensitivity (82.4%, 95% CI (69.1,91.6)) but also the second lowest specificity (67.8%, 95% CI (64.5,77.1)), GDx had the lowest sensitivity (26.9%, 95% CI (15.6,41.0)) but the highest specificity (96.7%, 95% CI (95.2,97.8)), and the other two tests providing intermediate results (HRT-MRA had a slightly lower sensitivity and specificity than HRT-GPS but a slightly higher specificity, and OCT had the second lowest sensitivity but the second highest specificity values). Sensitivity was slight lower for all HRT-MRA, GDx and OCT than under the default analysis but with a slightly higher specificity. HRT-GPS has very similar results to before. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). DORs ranged from 6.85 for MRA to 10.83 for GDx.

Reference standard definition: glaucoma in 'best eye'
Test abnormal: outside normal limits
Handling of no result: A imaging classification used
B-D abnormal
E excluded

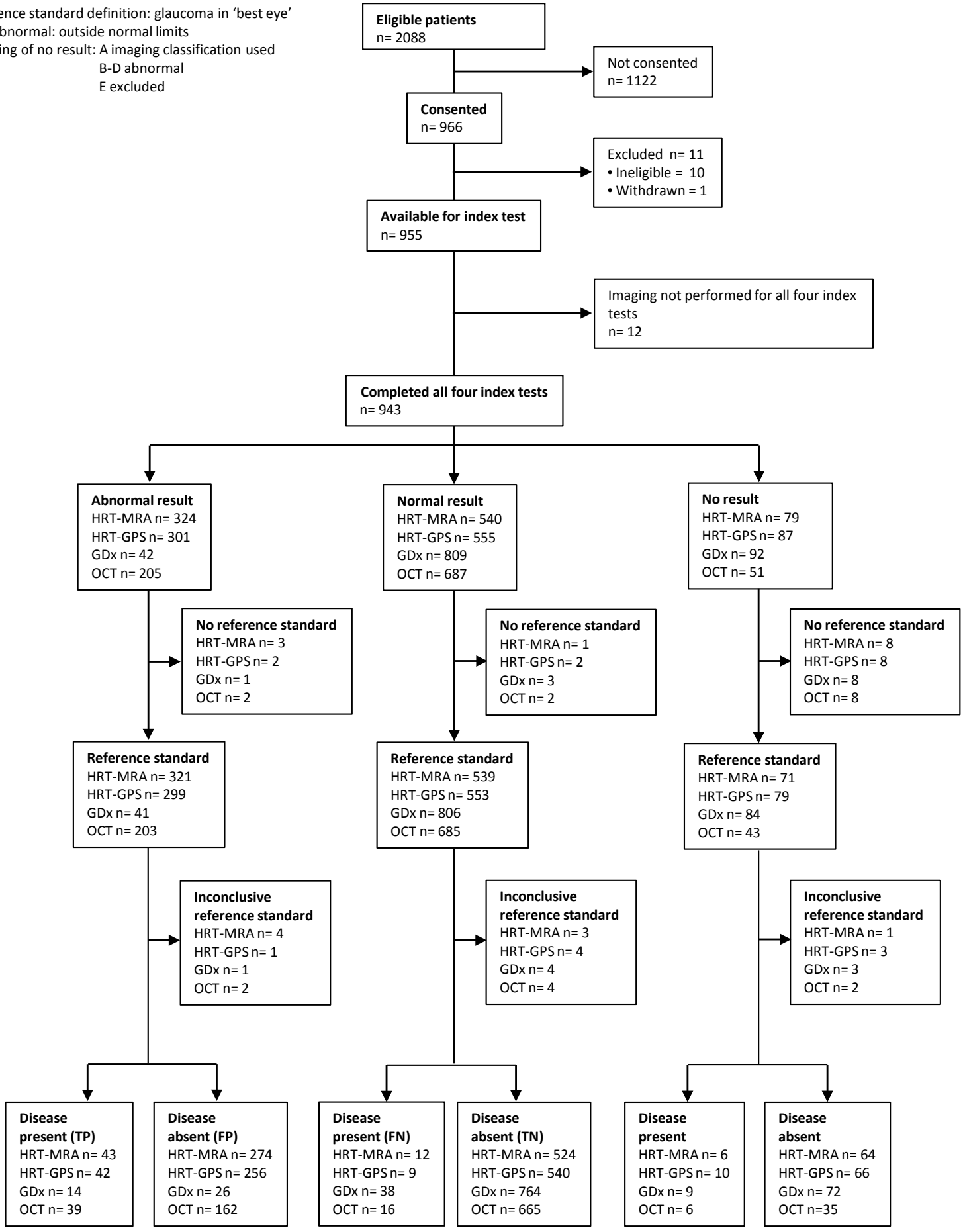


Figure 9 Flow diagram: diagnostic sensitivity analysis 6

Table 27 Diagnostic performance: diagnosis sensitivity analysis 6

Test	Diagnostic parameter	Point Estimate	95% CI Lower	95% CI Upper
HRT-MRA	Sensitivity %	78.2	65.0	88.2
	Specificity %	65.7	62.3	69.0
	Positive Likelihood Ratio	2.28	1.92	2.70
	Negative Likelihood Ratio	0.33	0.20	0.55
	DOR	6.85	3.55	13.21
HRT-GPS	Sensitivity %	82.4	69.1	91.6
	Specificity %	67.8	64.5	71.1
	Positive Likelihood Ratio	2.56	2.18	3.01
	Negative Likelihood Ratio	0.26	0.14	0.47
	DOR	9.84	4.72	20.53
GDx	Sensitivity %	26.9	15.6	41.0
	Specificity %	96.7	95.2	97.8
	Positive Likelihood Ratio	8.18	4.55	14.70
	Negative Likelihood Ratio	0.76	0.64	0.89
	DOR	10.83	5.23	22.39
OCT	Sensitivity %	70.9	57.1	82.4
	Specificity %	80.4	77.5	83.1
	Positive Likelihood Ratio	3.62	2.91	4.50
	Negative Likelihood Ratio	0.36	0.24	0.55
	DOR	10.01	5.45	18.35

Combinations of imaging tests

HRT-MRA was combined with the other imaging test to form three combined tests and the diagnostic performance assessed. The reference standard and the definition of an abnormal imaging test result was the same as for the default analysis (abnormal imaging test 'outside normal limits'; reference standard diagnosis of glaucoma in the 'worse' eye; and only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.). The corresponding flow of study participants is shown in Figure 10 with respective number of abnormal, normal and no results cases by combination imaging test, and the corresponding reference standard finding shown. The diagnostic performance for the four tests is given in Table 27. Results showed a trade-off between detection of glaucoma and correctly identifying non-glaucoma cases: HRT-MRA combined with OCT had the highest sensitivity (91.7%, 95% CI (85.7, 95.8)) but the second lowest specificity (53.8%, 95% CI (50.0,57.5)), HRT-MRA combined with GDx had the lowest sensitivity (89.5%, 95% CI (82.7,94.3)) but the highest specificity (62.8%, 95% CI (59.0,66.5)). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma all 3 combination imaging tests (CIs did not contain 1.0). DORs ranged from 11.34 for HRT-MRA combined with HRT-GPS, to 14.43 for HRT-MRA combined with GDx.

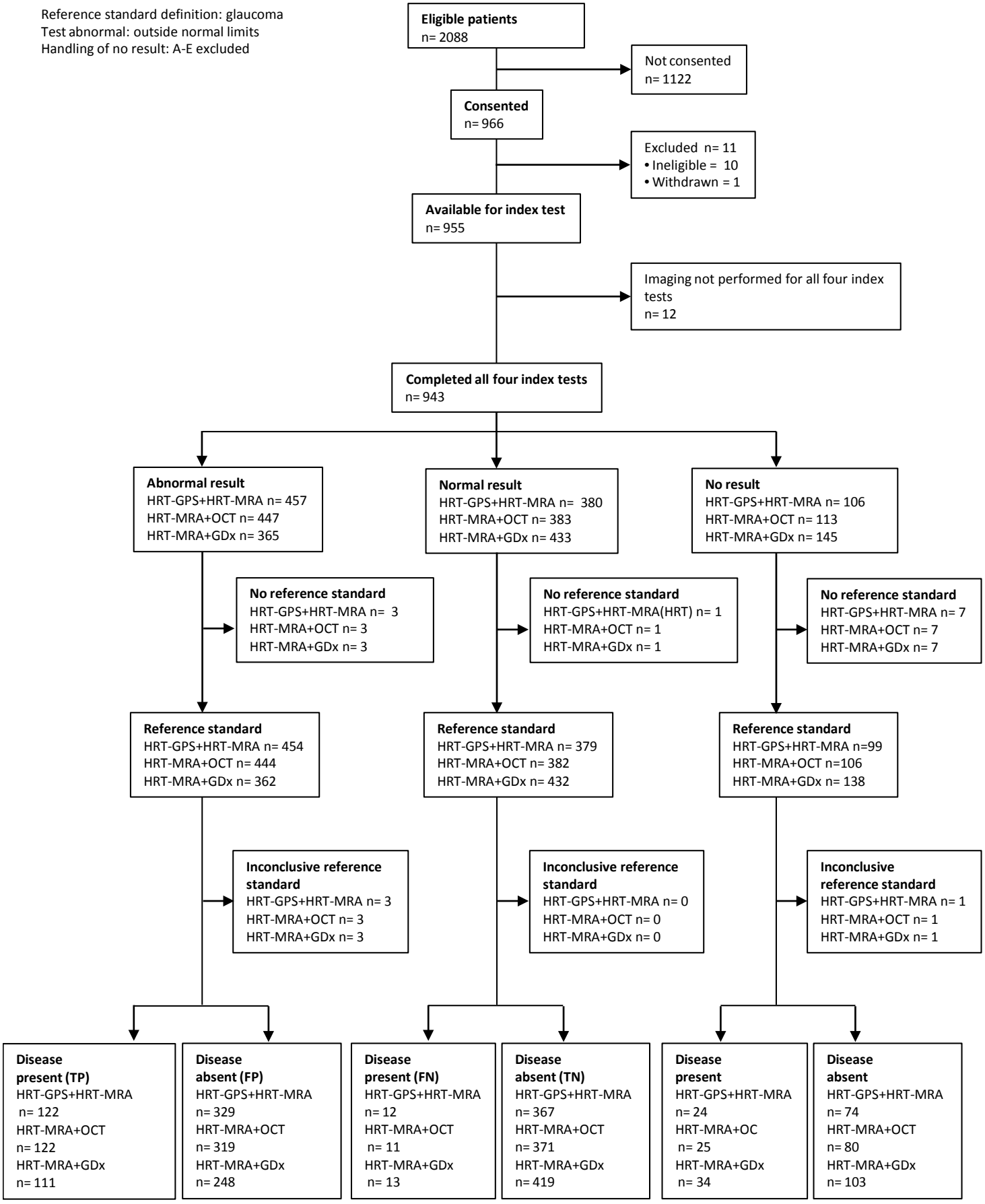


Figure 10 Flow diagram: combination of imaging tests

Table 28 Diagnostic performance: diagnostic performance of test combinations

Test	Diagnostic parameter	Point Estimate	95% CI Lower	95% CI Upper
HRT-MRA + HRT-GPS	Sensitivity %	91.0	84.9	95.3
	Specificity %	52.7	48.9	56.5
	Positive Likelihood Ratio	1.93	1.75	2.12
	Negative Likelihood Ratio	0.17	0.10	0.29
	DOR	11.34	6.15	20.90
HRT-MRA + GDx	Sensitivity %	89.5	82.7	94.3
	Specificity %	62.8	59.0	66.5
	Positive Likelihood Ratio	2.41	2.14	2.70
	Negative Likelihood Ratio	0.17	0.10	0.28
	DOR	14.43	7.95	26.17
HRT-MRA + OCT	Sensitivity %	91.7	85.7	95.8
	Specificity %	53.8	50.0	57.5
	Positive Likelihood Ratio	1.98	1.80	2.18
	Negative Likelihood Ratio	0.15	0.09	0.27
	DOR	12.90	6.84	24.34

Discussion

The diagnostic performance of four imaging tests (HRT-MRA, HRT-GPS, GDx and OCT) for the detection of glaucoma was compared for the GATE population of referrals to a glaucoma clinic in secondary care. The sensitivity and specificity of the

four imaging tests for the default diagnosis analysis and sensitivity analyses (see table 15 for details) are summarised in Figure 11 and 12 respectively.

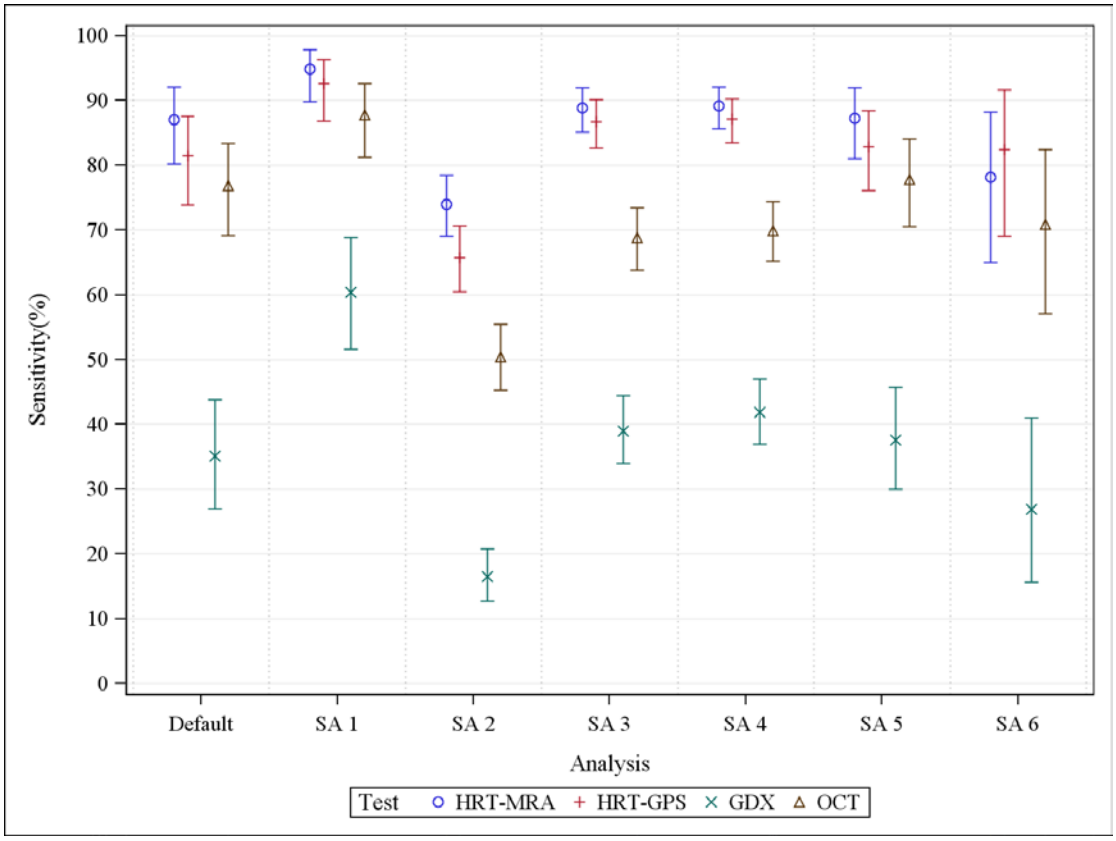


Figure 11 Summary of the sensitivity of imaging tests across all diagnosis analyses

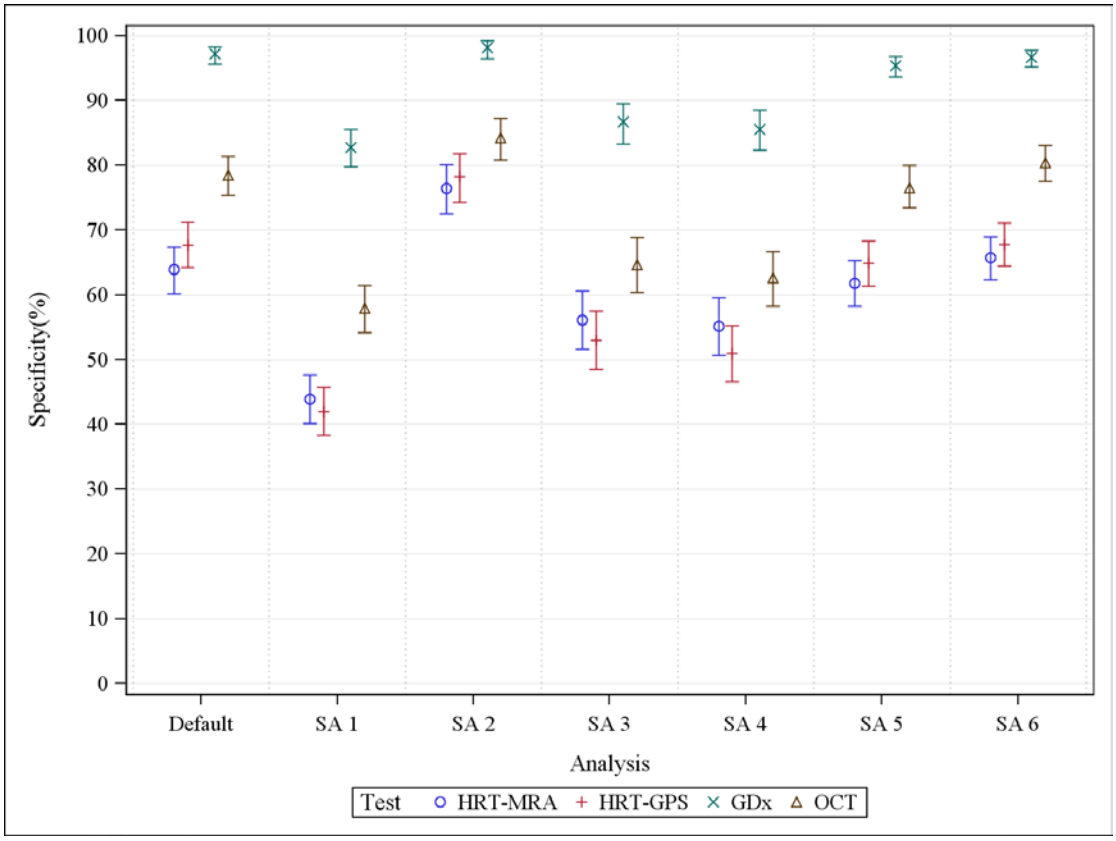


Figure 12 Summary of the specificity of imaging tests across all diagnosis analyses

All four imaging tests had some value in terms of ruling in and ruling out the presence of glaucoma. However, the diagnostic performance of the imaging tests differed in the ability to correctly diagnose glaucoma (sensitivity) and non-glaucoma cases (specificity). HRT-MRA had the highest sensitivity across analyses, except when the reference standard diagnosis was moderate and severe glaucoma only, when HRT-GPS was higher, but at a cost of lower specificity compared to other tests. In contrast GDx had consistently the best specificity though the lowest sensitivity. HRT-GPS results were typically similar to HRT-MRA as might be anticipated given that their analysis is based on the same imaging machine. Sensitivity for OCT was generally of a similar magnitude to its specificity. When the reference standard definition of disease excluded mild glaucoma, OCT provided higher diagnostic performance than the HRT-GPS and HRT-MRA with GDx providing the best specificity. The choice of which imaging test is to be preferred reflects the inherent trade off regarding diagnostic testing where the desire not to miss glaucoma when present must be balanced again the desire to correctly identify those who are without disease.

The non-diagnostic outcomes tended to favour OCT. OCT had the lowest number of low quality imaging results, with GDx having the highest. Average time taken to conduct the tests was lowest for OCT with the other tests taking a similar length of time. Less dilation was required for OCT followed by GDx than the HRT tests. Considering the time taken and need for dilation, patient preference tended to favour OCT followed by GDx although almost half of participants did not have a preference.

GATE was a large prospective paired diagnostic study and provided diagnostic tests in this desired setting. This is reflected in the precision in which the sensitivity and specificity were calculated with differences between every pair of tests identified for one if not both the sensitivity and specificity. McNemar's test was used to compare the sensitivity and specificity of the tests. Following the rationale of others in effectiveness studies, the paired comparisons were not adjusted for multi-comparisons. Even if such a correction were to have been applied such was the strength of evidence there would still be evidence of differences in the diagnostic performance of the different imaging tests.

A number of sensitivity analyses were carried out to assess the robustness of the findings of the default analysis. Varying the test definition of an abnormal imaging result by including the borderline category was carried out; this had the anticipated impact of improving the detection of glaucoma, though at the expense of more non-glaucoma cases being falsely classified to have glaucoma. This resulted in very high detection of glaucoma for HRT-MRA, HRT-GPS and OCT but with low to moderate diagnosis of non-glaucoma cases. GDx provided moderate performance for both detecting glaucoma and correctly diagnosing non-glaucoma cases. Additionally, the impact of also seeking to diagnose glaucoma suspects (based upon optic disc and/or visual field findings as described in Chapter 2) was assessed both with and without classifying borderline imaging findings as abnormal. When the test definition of abnormal incorporated the borderline category, the net impact was a slight increase in sensitivity for GDx, HRT-GPS and HRT-MRA, with OCT slightly reduced compared to the default analysis, suggesting that the OCT test deals less well with glaucoma suspect cases. The diagnostic performance upon the best eye as opposed to the worse eye gave similar results though with generally a lower sensitivity and slightly higher

specificity. HRT-GPS diagnostic performance for this data was remarkably similar to when the worse eye was used.

Finally, the impact of using a combination of tests was assessed. Given the findings of the default diagnosis analysis and associated sensitivity analyses, this was restricted to an assessment of whether using another imaging test in addition to HRT-MRA appeared to be beneficial. While the additional use of another test led to improved detection of glaucoma, the improvement was marginal and smaller than the loss in terms of the handling of non-diseased cases. While the use of two tests in combination did have some benefit in terms of reducing the number of no result cases, the change in diagnostic performance coupled with the additional practical and cost implications in terms of training and staff time and an additional requirement of equipment (for two of the three combinations) suggest that the use of a single test is to be preferred.

A number of assumptions underpinned the analysis and interpretation of the results. Most importantly the reference standard was assumed to be perfect though it is widely recognised that diagnosis of glaucoma is difficult and uncertainty exists even amongst specialists. While consensus was sought through structured training, some assessor differences may have remained between the sites. Additionally, the diagnosis and clinical management of patients with suspected glaucoma is uncertain; in particular the risk of conversion of such individuals is not known. Nevertheless, the findings provide evidence reflective of current clinical practice in NHS glaucoma clinics.
A number of areas for further research are clear. Further investigation of varying the results of the imaging tests beyond the standard options could be undertaken as the recommended classification may not the one best suited to the population which GATE recruited from. The definition and clinical management of glaucoma suspects is also an area in which further research is needed, in particular quantifying the proportion which will convert or will be discharged from clinical care over subsequent years. Finally, the diagnosis value of using an imaging test explicitly in a triage scenario with the additional use of an IOP measurement and visual acuity to form a composite triage test requires evaluation.

Chapter 5 Triage analysis results

Overview

The chapter reports the results of the triage analyses which aimed to assess the diagnostic performance of the four imaging tests in a triage setting. The specific diagnostic performance analyses covered in this chapter are the *default triage analysis* along with eight sensitivity analyses (*triage sensitivity analyses 1-8 – see Table 29*) for a list with definitions. A further set of three analyses specifically to inform the economic model are in Appendix 7. The default triage analysis was defined as one where the reference standard was the person level clinical decision ("not discharged" or "discharged"). The test was defined as categorising a patient as requiring to be referred on ("For referral") if any of the elements of the composite triage test (Imaging, IOP and/or visual acuity) were themselves "abnormal": imaging outside normal limits on the overall classification of the respective imaging test (*see Chapter 2*), IOP>21mmHg or visual acuity of 6/12 or poorer under the default triage analysis.

Cases where the imaging test did not produce an overall classification or where quality was poor were also defined as imaging test being abnormal and therefore "For referral". The eight sensitivity analyses assessed the impact of varying assumptions made in the default triage analysis relating to the definition of a positive test result, modifying or removing the IOP and/or visual acuity components of the triage test, and how cases where the test did not produce an overall classification were handled in the analysis.

The analyses in this chapter pertain to the 943 participants remaining in the study (*see Chapter 4*). The reference standard was available for 933 cases. For all analyses, a STARD diagram shows the flow of participants. The subset of participants who received all four tests and were considered in the statistical analyses are separated out into three groups according to whether each triage test result was "abnormal", "normal", or "no result" (the triage test result was not available whether because the test was inconclusive or the result was missing). For each of these three groups the group status according to the reference standard ("discharged" or "not discharged") for each participant is given or alternatively the reference standard was stated to be

missing or inconclusive. The final categorisations of the triage test result by reference standard status provides the four possible combinations (true and false positive, false and true negative) from which the diagnostic performance was assessed. Sensitivity, specificity, likelihood ratios and DOR are provided with associated 95% CIs for each analysis.

Table 29 Triage analyses

Analysis	Reference standard definition	Test abnormal	Handling of no result categories	Figure no.	Table no.
Default triage Analysis	Not discharged	Imaging (outside normal limits) or IOP>21mmHg or VA6/12 or poorer	A-D for referral E excluded	13	30, 31
Triage sensitivity analysis 1	Not discharged	Imaging (outside normal limits or borderline) or IOP>21mmHg or VA6/12 or poorer	A-D for referral E excluded	14	32
Triage sensitivity analysis 2	Not discharged	Imaging (outside normal limits) or IOP>21mmHg or VA6/12 or poorer	A use imaging classification B for referral C-E excluded	15	33
Triage sensitivity analysis 3	Not discharged	Imaging (outside normal limits or borderline) or IOP>21mmHg or VA6/12 or poorer	A use imaging classification B for referral C-E excluded	16	34
Triage sensitivity analysis 4	Not discharged	Imaging (outside normal limits) or IOP>21mmHg (Referred IOP) or VA6/12 or poorer	A-D for referral E excluded	17	35

Analysis	Reference standard definition	Test abnormal	Handling of no result categories	Figure no.	Table no.
Triage sensitivity analysis 5	Not discharged	Imaging (outside normal limits) or VA6/12 or poorer	A-D for referral E excluded	18	36
Triage Sensitivity analysis 6	Not discharged	Imaging (outside normal limits) or IOP>21mmHg	A-D for referral E excluded	19	37
Triage sensitivity analysis 7	Not discharged	Imaging (outside normal limits) or IOP>26mmHg or VA6/12 or poorer	A-D for referral E excluded	20	38
Triage sensitivity analysis 8	Not discharged	Imaging (outside normal limits) or IOP>21mmHg or VA6/18 or poorer	A-D for referral E excluded	21	39

No result categories

A: test performed and imaging report produced but quality is lower than manufacturer quality cut-off

B: test performed and imaging report produced but no overall classification generated by machine

C: test performed but there was a clear imaging artefact on the report

D: test attempted but no imaging could be acquired from the patient's eyes – no report generated

E: missing imaging (due to study related or data collection issues)

Default triage analysis

The results for the default triage analysis are presented in two sections:

- diagnostic performance of the triage tests, and
- paired comparisons of triage tests.

Diagnostic performance of the triage tests

For the default triage analysis, the triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' OR (b) IOP>21mmHg OR (c) visual acuity is 6/12 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to the default triage analysis is shown in Figure 13 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 576 were not discharged and 357 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 30. Results showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-GPS had the highest sensitivity (86.0%, 95% CI (82.8,88.7)) but lowest specificity (39.1%, 95% CI (34.0,44.5)), GDx had the lowest sensitivity (64.7%, 95% CI (60.7,68.7)) but the highest specificity (53.6%, 95% CI (48.2,58.9)), and the other two tests provided intermediate results (HRT-MRA values were very similar to the HRT-GPS results, and OCT had lower sensitivity 75.4%, 95% CI (71.9,78.9)) but higher specificity values than HRT-GPS and HRT-MRA). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.12 for GDx and OCT to 3.94 for HRT-GPS.

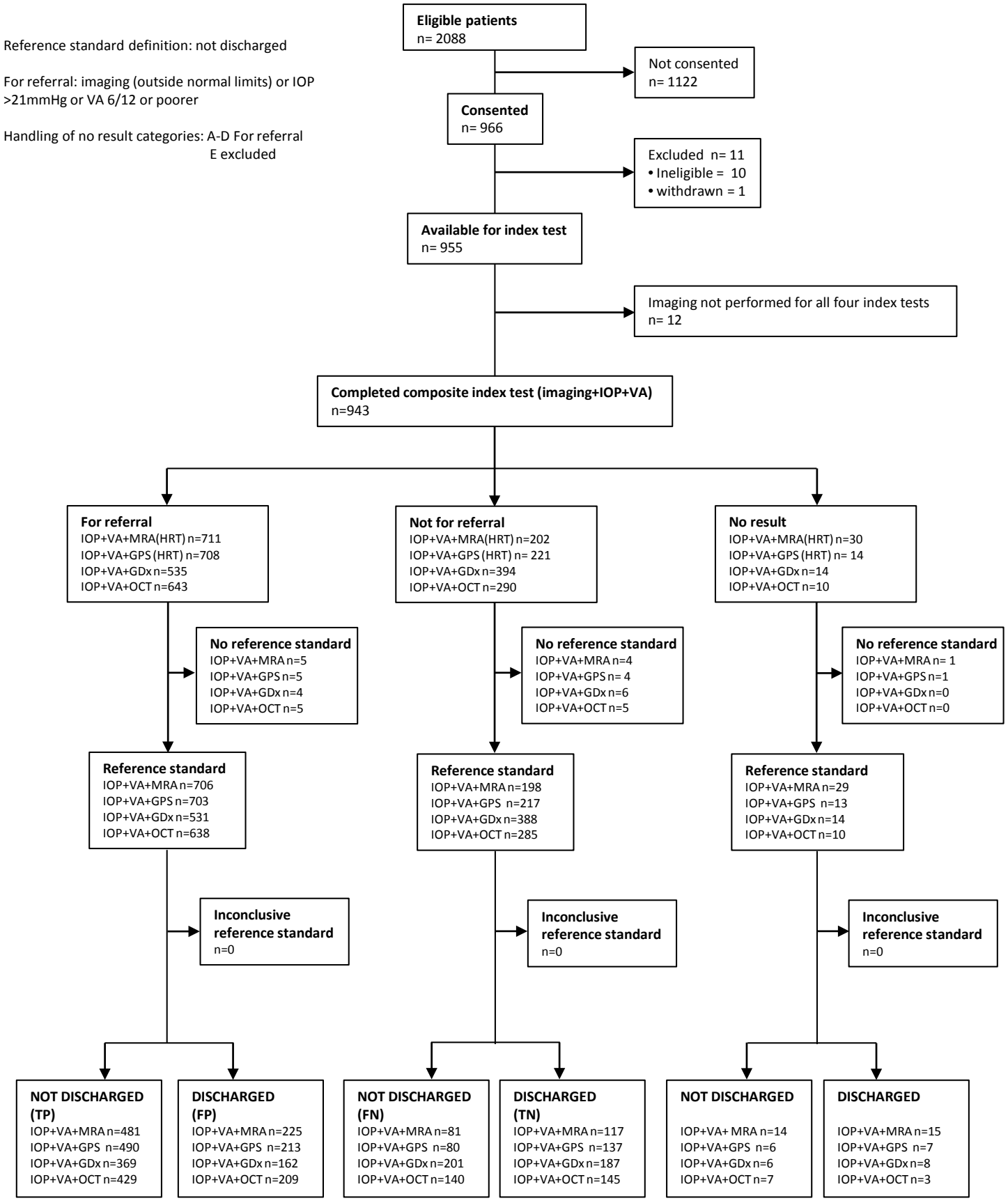


Figure 13 Flow diagram: default triage analysis

Table 30 Diagnostic performance: default triage analysis

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	85.6	82.4	88.4
	Specificity - %	34.2	29.20	39.5
	Positive Likelihood Ratio	1.3	1.20	1.41
	Negative Likelihood Ratio	0.4	0.33	0.54
	DOR	3.09	2.23	4.27
HRT-GPS	Sensitivity - %	86.0	82.8	88.7
	Specificity - %	39.1	34.0	44.5
	Positive Likelihood Ratio	1.41	1.29	1.55
	Negative Likelihood Ratio	0.36	0.28	0.46
	DOR	3.94	2.86	5.42
GDx	Sensitivity - %	64.7	60.7	68.7
	Specificity - %	53.6	48.2	58.9
	Positive Likelihood Ratio	1.39	1.23	1.59
	Negative Likelihood Ratio	0.66	0.57	0.76
	DOR	2.12	1.62	2.78
OCT	Sensitivity - %	75.4	71.6	78.9
	Specificity - %	41.0	35.8	46.3
	Positive Likelihood Ratio	1.28	1.16	1.41
	Negative Likelihood Ratio	0.60	0.50	0.73
	DOR	2.13	1.60	2.83

Paired comparisons of imaging tests

Table 31 shows the paired difference (with 95% CI) and corresponding McNemar's tests p-value for comparisons between pairs of tests. There was evidence that the sensitivity of all tests differed from each other except for HRT-GPS versus HRT-MRA.

HRT-GPS, HRT-MRA and GDx had the highest and lowest sensitivity respectively. Differences varied from 0.2% (HRT-GPS versus HRT-MRA) to 19.7 %. (HRT-GPS versus GDx). Similarly there was evidence that specificities for all the tests varied from each other (according to McNemar's test) except HRT-GPS versus OCT.

Table 31 Paired comparisons of sensitivity and specificity between the triage tests

Tests Compared	Parameter	Test	Value - % (95% CI)	p-value (McNemar's)
HRT-GPS vs	Sensitivity	HRT-GPS	85.8 (82.9 to 88.7)	
GDx		GDx	64.5 (60.6 to 68.5)	
		Difference	21.3 (17.7 to 24.9)	<.0001
	Specificity	HRT-GPS	39.6 (34.4 to 44.7)	
		GDx	53.8 (48.5 to 59.0)	
		Difference	-14.2 (-19.0 to -9.2)	<.0001
GDx vs	Sensitivity	GDx	64.8 (60.9 to 68.8)	
OCT		OCT	75.1 (71.6 to 78.7)	
		Difference	-10.3 (-13.5 to -7.0)	<.0001
	Specificity	GDx	53.4 (48.2 to 58.7)	
		OCT	41.1 (35.9 to 46.3)	
		Difference	12.4 (7.9 to 16.7)	<.0001
GDx vs	Sensitivity	GDx	64.9 (61.0 to 68.9)	
HRT-MRA		HRT-MRA	85.4 (82.5 to 88.4)	
		Difference	-20.5 (-24.3 to -16.7)	<.0001
	Specificity	GDx	53.3 (47.9 to 58.6)	
		HRT-MRA	34.3 (29.3 to 39.4)	
		Difference	18.9 (13.8 to 23.9)	<.0001
HRT-GPS vs	Sensitivity	HRT-GPS	85.7 (82.8 to 88.6)	
HRT-MRA		HRT-MRA	85.5 (82.6 to 88.4)	
		Difference	0.2 (-2.4 to 2.8)	0.8907
	Specificity	HRT-GPS	39.3 (34.1 to 44.5)	
		HRT-MRA	34.3 (29.3 to 39.3)	
		Difference	5.0 (0.3 to 9.6)	<.0001
HRT-MRA vs	Sensitivity	HRT-MRA	85.6 (82.7 to 88.5)	
OCT		OCT	75.2 (71.6 to 78.8)	
		Difference	10.4 (7.1 to 13.8)	<.0001
	Specificity	HRT-MRA	34.2 (29.2 to 39.2)	
		OCT	40.9 (35.7 to 46.1)	

Tests Compared	Parameter	Test	Value - % (95% CI)	p-value (McNemar's)
		Difference	-6.7 (-12.2 to -1.2)	0.0171
HRT-GPS vs	Sensitivity	HRT-GPS	86.1 (83.2 to 88.9)	
OCT		OCT	75.3 (71.8 to 78.9)	
		Difference	10.8 (7.4 to 14.2)	<.0001
	Specificity	HRT-GPS	39.1 (34.0 to 44.3)	
		OCT	41.1 (36.0 to 46.3)	
		Difference	-2.0 (-7.4 to 3.5)	0.4726

Triage sensitivity analysis 1

Triage sensitivity analysis 1 differed from the default triage analysis in that a borderline finding on the imaging test was also classified as an abnormal result.

For triage sensitivity analysis1, the triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' or 'borderline' OR (b) IOP>21mmHg OR (c) visual acuity is 6/12 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 1 is shown in Figure 14 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 576 were not discharged and 357 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 32. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-GPS had the highest sensitivity (94.0%, 95% CI (91.8,95.8)) but second lowest specificity (24.9%, 95% CI (20.4,29.7)), GDx had the lowest sensitivity (74.9%, 95% CI (71.1, 78.4)) but the highest specificity, and the other two tests provided intermediate results (HRT-MRA values were very similar though marginally inferior to the HRT-GPS results, and OCT had lower sensitivity (84.2%, 95% CI (80.9,87.1)) but slightly higher specificity than HRT-GPS and HRT-MRA). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in

and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.04 for OCT to 5.21 for HRT-GPS.

Figure 14 Flow diagram: triage sensitivity analysis 1

Table 32 Diagnostic performance: triage sensitivity analysis 1

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	92.7	90.2	94.7
	Specificity - %	24.0	19.5	28.9
	Positive Likelihood Ratio	1.2	1.14	1.30
	Negative Likelihood Ratio	0.30	0.21	0.43
	DOR	4.01	2.68	6.00
HRT-GPS	Sensitivity - %	94.0	91.8	95.8
	Specificity - %	24.9	20.4	29.7
	Positive Likelihood Ratio	1.25	1.17	1.33
	Negative Likelihood Ratio	0.24	0.17	0.35
	DOR	5.21	3.42	7.96
GDx	Sensitivity - %	74.9	71.1	78.4
	Specificity - %	45.0	39.7	50.4
	Positive Likelihood Ratio	1.36	1.22	1.51
	Negative Likelihood Ratio	0.56	0.46	0.67
	DOR	2.44	1.84	3.24
OCT	Sensitivity - %	84.2	80.9	87.1
	Specificity - %	27.7	23.1	32.7
	Positive Likelihood Ratio	1.16	1.08	1.51
	Negative Likelihood Ratio	0.57	0.44	0.74
	DOR	2.04	1.47	2.82

Triage sensitivity analysis 2

Triage sensitivity analysis 2 has the same reference standard and definition of abnormal test result as the default analysis but did not include all 'no result' cases (see Table 33).

For triage sensitivity analysis 2, the triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' OR (b) IOP>21mmHg OR (c) visual acuity is 6/12 or poorer. Poor quality imaging test results were included, and those where an image was acquired but no classification generated were included as abnormal. All other missing imaging results were excluded. The corresponding reference standard definition is a clinical decision not to discharge the patient.
The flow of study participants according to triage sensitivity analysis 2 is shown in Figure 15 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 33. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-GPS had the highest sensitivity (84.6%, 95% CI (81.6,87.5)) but the second lowest specificity (39.7%, 95% CI (34.6,45.1)), GDx had the lowest sensitivity (61.1%, 95% CI (59.9,65.1)) but the highest specificity (59.0%, 95% CI (53.7,64.2)), and the other two tests providing intermediate results (HRT-MRA values were very similar though slightly inferior to the HRT-GPS results, and OCT had the second lowest sensitivity (75.0%, 95% CI (71.3,78.5)) but the second highest specificity (42.1%, 95% CI (36.9,47.4)) values). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.19 for GDx to 3.61 for OCT.

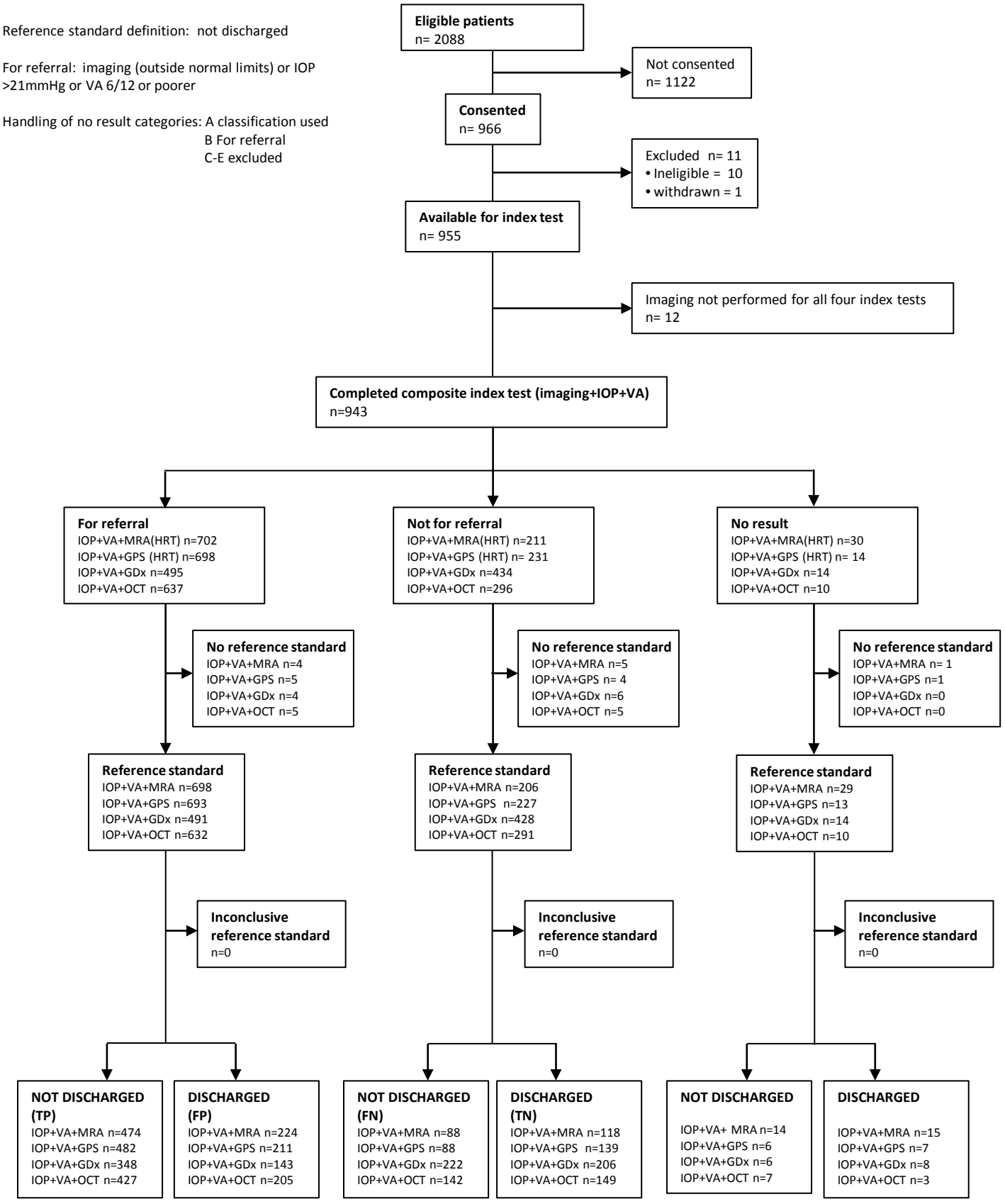


Figure 15 Flow diagram: triage sensitivity analysis 2

Table 33 Diagnostic performance: triage sensitivity analysis 2

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	84.3	81.1	87.2
	Specificity - %	34.5	29.5	39.8
	Positive Likelihood Ratio	1.29	1.18	1.40
	Negative Likelihood Ratio	0.45	0.36	0.58
	DOR	2.84	2.06	3.90
HRT-GPS	Sensitivity - %	84.6	81.4	87.5
	Specificity - %	39.7	34.6	45.1
	Positive Likelihood Ratio	1.40	1.28	1.54
	Negative Likelihood Ratio	0.39	0.31	0.49
	DOR	3.61	2.64	4.93
GDx	Sensitivity - %	61.1	56.9	65.1
	Specificity - %	59.0	53.7	64.2
	Positive Likelihood Ratio	1.49	1.29	1.72
	Negative Likelihood Ratio	0.66	0.58	0.76
	DOR	2.26	1.72	2.96
OCT	Sensitivity - %	75.0	71.3	78.5
	Specificity - %	42.1	36.9	47.4
	Positive Likelihood Ratio	1.30	1.17	1.43
	Negative Likelihood Ratio	0.59	0.49	0.72
	DOR	2.19	1.65	2.90

Triage sensitivity analysis 3

Triage sensitivity analysis 3 was the same as triage sensitivity analysis 2 except that 'borderline' test results were also classified as abnormal.

For triage sensitivity analysis 3, the triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' or 'borderline' OR (b) IOP>21mmHg OR (c) visual acuity is 6/12 or poorer. Poor quality imaging test results were included, and those where an image was acquired but no classification generated were included as abnormal. All other missing imaging results were excluded. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 3 is shown in Figure 16 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 34. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-GPS had the highest sensitivity (93.3%, 95% CI (91.0,95.2)) but second lowest specificity (24.9%, 95% CI (20.4,29.7)), GDx had the lowest sensitivity (72.3%, 95% CI (68.4,75.9)) but the highest specificity (49.0%, 95% CI (43.6,54.4)), and the other two tests providing intermediate results (HRT-MRA values were very similar to the HRT-GPS results though slightly inferior, and OCT had the second lowest sensitivity (84.2%, 95% CI (80.9,87.1)) but the second highest specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.12 for OCT to 4.63 for HRT-GPS.

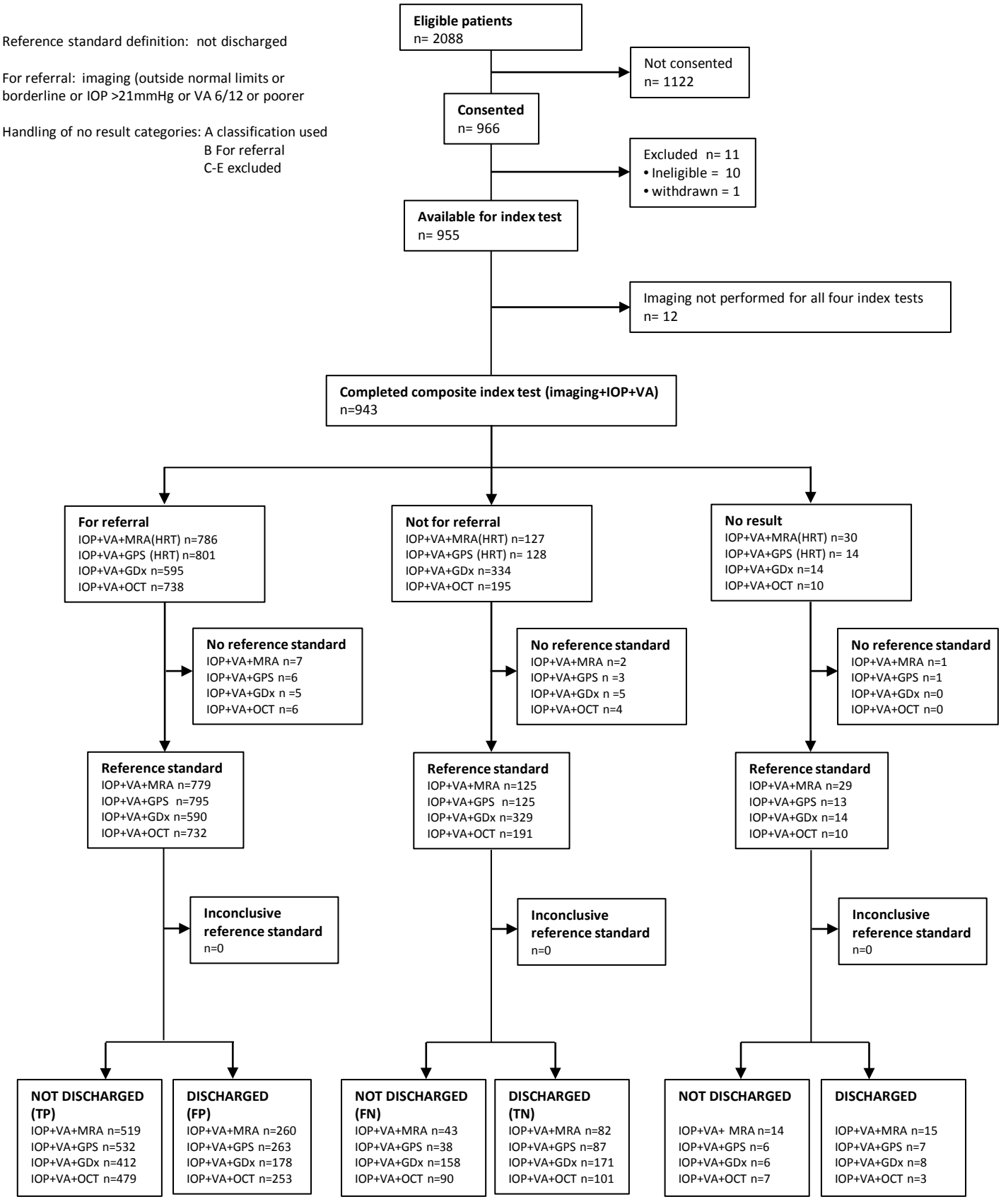


Figure 16 Flow diagram: triage sensitivity analysis 3

Table 34 Diagnostic performance: triage sensitivity analysis 3

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	92.3	89.8	94.4
	Specificity - %	24.0	19.5	28.9
	Positive Likelihood Ratio	1.21	1.14	1.03
	Negative Likelihood Ratio	0.32	0.23	0.45
	DOR	3.81	2.56	5.67
HRT-GPS	Sensitivity - %	93.3	91.0	95.2
	Specificity - %	24.9	20.4	29.7
	Positive Likelihood Ratio	1.24	1.16	1.32
	Negative Likelihood Ratio	0.27	0.19	0.38
	DOR	4.63	3.08	6.97
GDx	Sensitivity - %	72.3	68.4	75.9
	Specificity - %	49.0	43.6	54.4
	Positive Likelihood Ratio	1.42	1.26	1.59
	Negative Likelihood Ratio	0.57	0.48	0.67
	DOR	2.51	1.90	3.31
OCT	Sensitivity - %	84.2	80.9	87.1
	Specificity - %	28.5	23.9	33.5
	Positive Likelihood Ratio	1.18	1.09	1.27
	Negative Likelihood Ratio	0.55	0.43	0.71
	DOR	2.12	1.54	2.93

Triage sensitivity analysis 4

Triage sensitivity analysis 4 differed from the default triage analysis in that referral IOP >21mmHg rather than clinician IOP > 21mmHg was used to identify abnormal tests. The triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' OR (b) Referral IOP>21mmHg OR (c) visual acuity is 6/12 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 4 is shown in Figure 17 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 35. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-GPS had the highest sensitivity (86.5%, 95% CI (83.4,89.2)) but second lowest specificity (24.0%, 95% CI (19.6,28.8)), GDx had the lowest sensitivity (67.2%, 95% CI (63.2,71.0)) but the highest specificity (35.8%, 95% CI (30.8,41.1)), and the other two tests providing intermediate results (HRT-MRA values were very similar to the HRT-GPS results though slightly inferior, and OCT had the second lowest sensitivity (76.8%, 95% CI (73.1,80.2)) but the second highest specificity (27.7%, 95% CI (23.1,32.7)). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 1.14 for GDx to 2.02 for HRT-GPS.

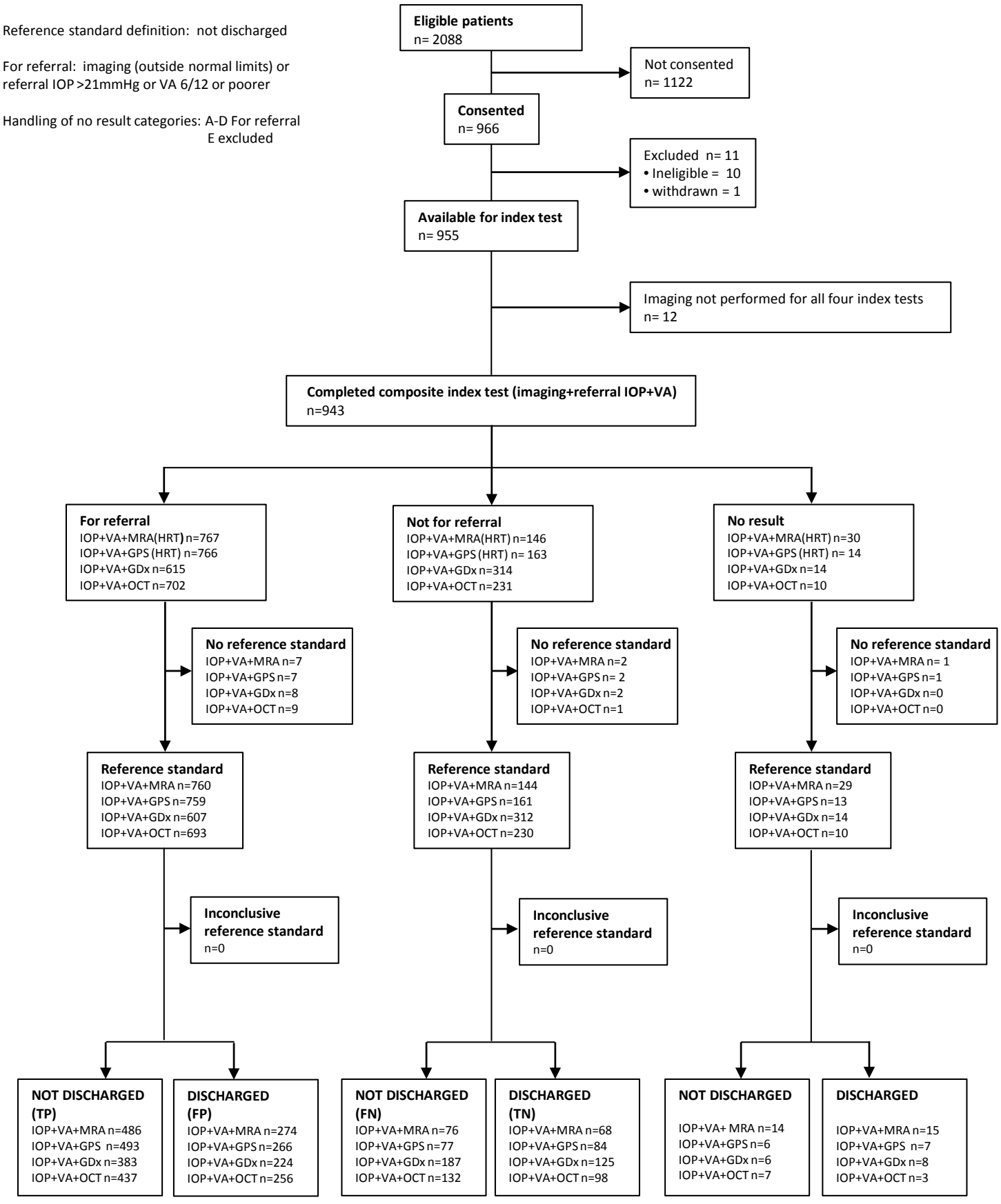


Figure 17 Flow diagram: triage sensitivity analysis 4

Table 35 Diagnostic performance: triage sensitivity analysis 4

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	86.5	83.4	89.2
	Specificity - %	19.9	15.8	24.5
	Positive Likelihood Ratio	1.08	1.01	1.15
	Negative Likelihood Ratio	0.68	0.50	0.92
	DOR	1.59	1.11	2.27
HRT-GPS	Sensitivity - %	86.5	83.4	89.2
	Specificity - %	24.0	19.6	28.8
	Positive Likelihood Ratio	1.14	1.06	1.22
	Negative Likelihood Ratio	0.56	0.43	0.74
	DOR	2.02	1.43	2.85
GDx	Sensitivity - %	67.2	63.2	71.0
	Specificity - %	35.8	30.8	41.1
	Positive Likelihood Ratio	1.05	0.95	1.15
	Negative Likelihood Ratio	0.924	0.76	1.10
	DOR	1.1	0.86	1.51
OCT	Sensitivity - %	76.8	73.1	80.2
	Specificity - %	27.7	23.1	32.7
	Positive Likelihood Ratio	1.06	0.98	1.15
	Negative Likelihood Ratio	0.84	0.67	1.05
	DOR	1.27	0.94	1.72

Triage sensitivity analysis 5

Triage sensitivity analysis 5 differed from the default triage analysis in that the IOP component was removed from the composite triage test. The triage test is classified as abnormal if the imaging test result is classified as 'outside normal limits' OR visual acuity is 6/12 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 5 is shown in Figure 18 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 36. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (68.9%, 95% CI (64.9,72.7)) but the lowest specificity (52.3%, 95% CI (463.9, 57.7)), GDx had the lowest sensitivity (32.8%, 95% CI (29.0,36.8)) but the highest specificity (81.1%, 95% CI (76.6,85.1)), and the other two tests providing intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had the second lowest sensitivity but the second highest specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 1.80 for OCT to 2.91 for HRT-GPS.

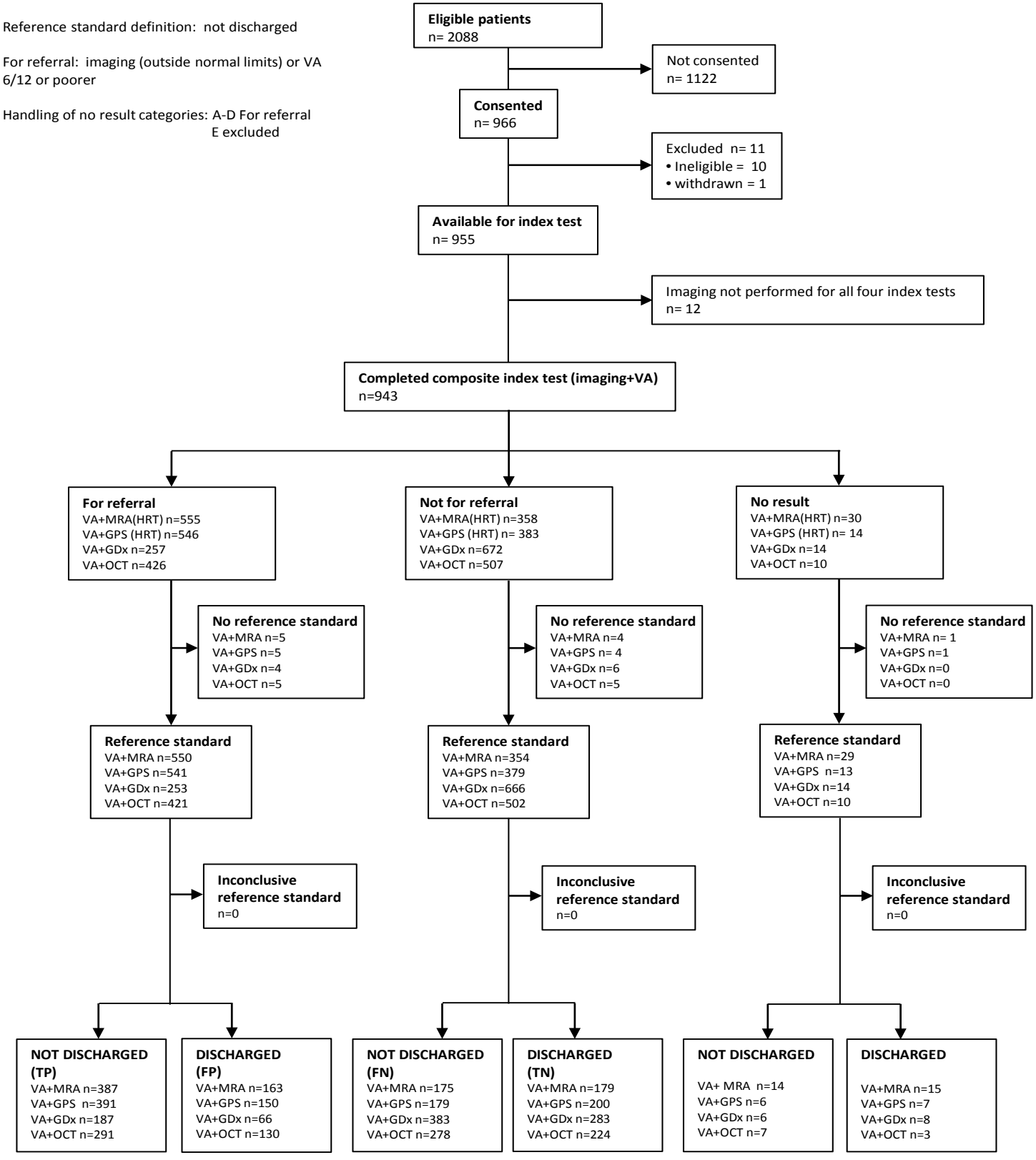


Figure 18 Flow diagram: triage sensitivity analysis 5

Table 36 Diagnostic performance: triage sensitivity analysis 5

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	68.9	64.9	72.7
	Specificity - %	52.3	46.9	57.7
	Positive Likelihood Ratio	1.44	1.28	1.64
	Negative Likelihood Ratio	0.59	0.51	0.70
	DOR	2.43	1.84	3.20
HRT-GPS	Sensitivity - %	68.6	64.6	72.4
	Specificity - %	57.1	51.8	62.4
	Positive Likelihood Ratio	1.60	1.40	1.83
	Negative Likelihood Ratio	0.55	0.47	0.64
	DOR	2.91	2.21	3.84
GDx	Sensitivity - %	32.8	29.0	36.8
	Specificity - %	81.1	76.6	85.1
	Positive Likelihood Ratio	1.73	1.36	2.22
	Negative Likelihood Ratio	0.83	0.77	0.89
	DOR	2.09	1.52	2.88
OCT	Sensitivity - %	51.1	47.0	55.3
	Specificity - %	63.3	58.0	68.3
	Positive Likelihood Ratio	1.39	1.19	1.63
	Negative Likelihood Ratio	0.77	0.69	0.87
	DOR	1.80	1.37	2.37

Triage sensitivity analysis 6

Triage sensitivity analysis 6 differed from the default triage analysis in that the visual acuity component was removed from the composite triage test. The triage test is classified as abnormal if the imaging test result is classified as 'outside normal limits' OR IOP >21mmHg. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 6 is shown in Figure 19 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 37. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (84.9%, 95% CI (81.9,87.7)) but second lowest specificity (37.4%, 95% CI (32.3,42.8)), GDx had the lowest sensitivity (60.5%, 95% CI (56.4,64.6)) but the highest specificity(57.6%, 95% CI (52.2,62.8)), and the other two tests provided intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had the second lowest sensitivity but the second highest specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.03 for OCT to 3.97 for HRT-GPS.

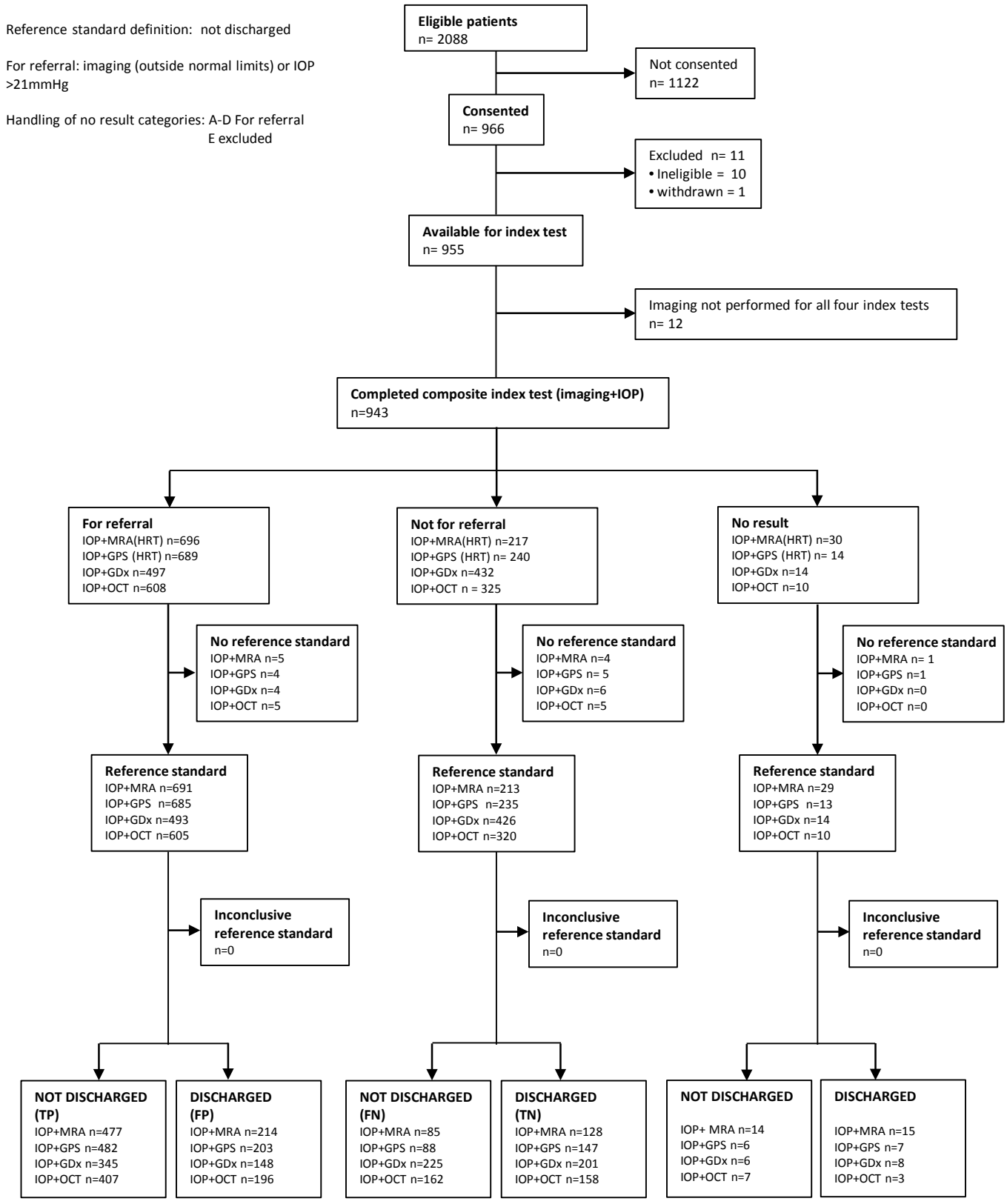


Figure 19 Flow diagram: triage sensitivity analysis 6

Table 37 Diagnostic performance: triage sensitivity analysis 6

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	84.9	81.6	87.7
	Specificity - %	37.4	32.3	42.8
	Positive Likelihood Ratio	1.36	1.24	1.48
	Negative Likelihood Ratio	0.40	0.32	0.48
	DOR	3.36	2.44	4.61
HRT-GPS	Sensitivity - %	84.6	81.3	87.4
	Specificity - %	42.0	36.8	47.4
	Positive Likelihood Ratio	1.46	1.32	1.60
	Negative Likelihood Ratio	0.37	0.29	0.46
	DOR	3.97	2.91	5.41
GDx	Sensitivity - %	60.5	56.4	64.6
	Specificity - %	57.6	52.2	62.8
	Positive Likelihood Ratio	1.43	1.24	1.64
	Negative Likelihood Ratio	0.69	0.60	0.79
	DOR	2.08	1.59	2.73
OCT	Sensitivity - %	71.5	67.6	75.2
	Specificity - %	44.6	39.4	50.0
	Positive Likelihood Ratio	1.29	1.16	1.44
	Negative Likelihood Ratio	0.64	0.54	0.76
	DOR	2.03	1.53	2.67

Triage sensitivity analysis 7

Triage sensitivity analysis 7 differed from the default triage analysis in that a higher IOP threshold of 26mmHg rather than 21mmHg was used to identify abnormal tests. The triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' OR (b) IOP>26mmHg OR (c) visual acuity is 6/12 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 7 is shown in Figure 20 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 38. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (77.2%, 95% CI (73.5,80.6)) but second lowest specificity (51.8%, 95% CI (46.3,57.2)), GDx had the lowest sensitivity (47.9%, 95% CI (43.7,52.1) but the highest specificity (79.1%, 95% CI (74.4,81.2)), and the other two tests provided intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had very similar sensitivity and specificity). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.61 for OCT to 4.03 for HRT-GPS.

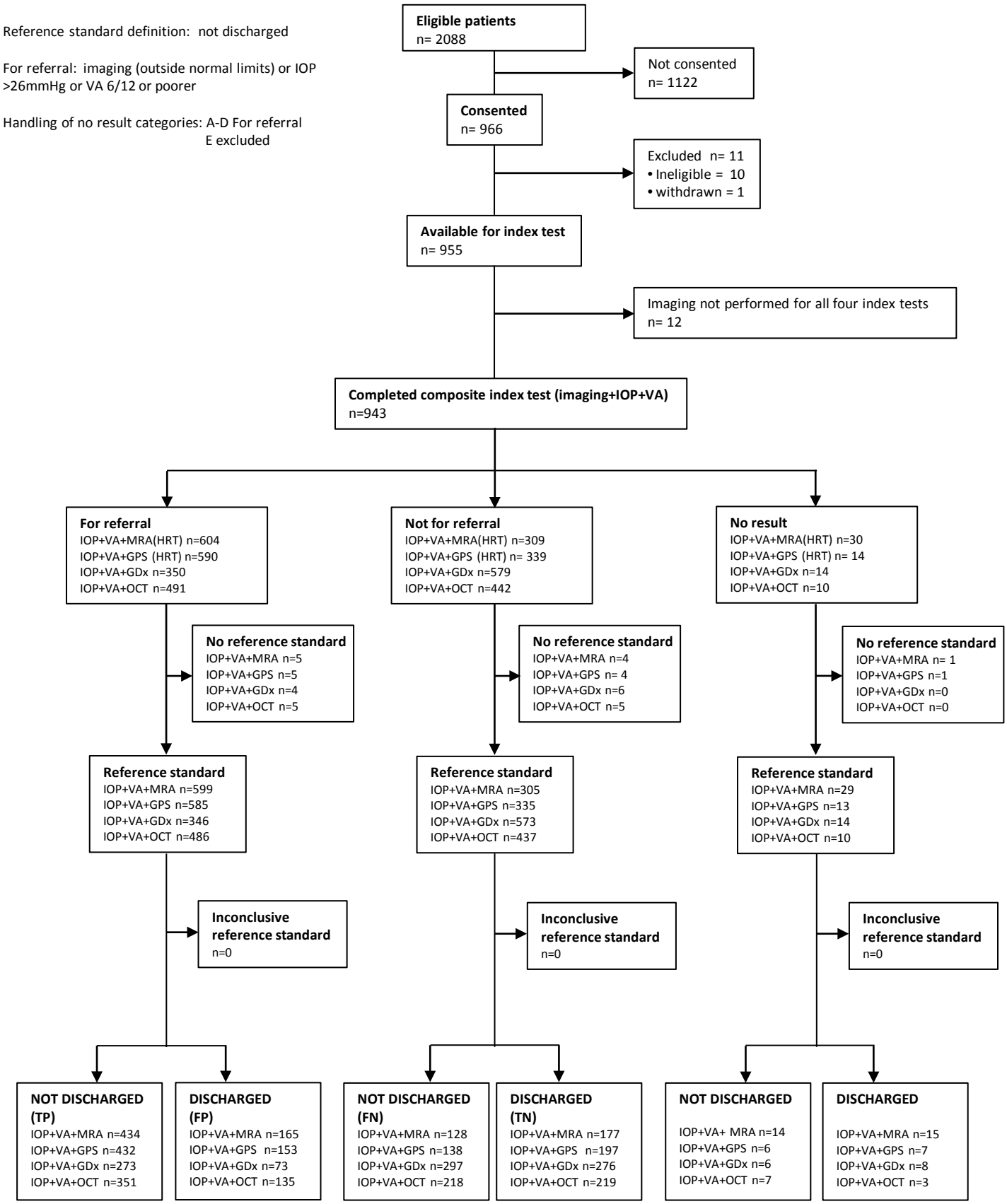


Figure 20 Flow diagram: triage sensitivity analysis 7

Table 38 Diagnostic performance: triage sensitivity analysis 7

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	77.2	73.5	80.6
	Specificity - %	51.8	46.3	57.2
	Positive Likelihood Ratio	1.60	1.42	1.80
	Negative Likelihood Ratio	0.44	0.37	0.53
	DOR	3.64	2.72	4.86
HRT-GPS	Sensitivity - %	75.8	72.1	79.3
	Specificity - %	56.3	50.9	61.6
	Positive Likelihood Ratio	1.73	1.53	1.97
	Negative Likelihood Ratio	0.43	0.36	0.51
	DOR	4.03	3.03	5.36
GDx	Sensitivity - %	47.9	43.7	52.1
	Specificity - %	79.1	74.4	81.2
	Positive Likelihood Ratio	2.29	1.84	2.86
	Negative Likelihood Ratio	0.66	0.60	0.72
	DOR	3.48	1.99	3.43
OCT	Sensitivity - %	61.7	57.6	65.7
	Specificity - %	61.9	56.6	66.9
	Positive Likelihood Ratio	1.62	1.40	1.87
	Negative Likelihood Ratio	0.62	0.54	0.71
	DOR	2.61	1.99	3.43

Triage sensitivity analysis 8

Triage sensitivity analysis 8 differed from the default triage analysis in that a higher visual acuity threshold of VA 6/18 or poorer was used to identify abnormal tests. The triage test is classified as abnormal if (a) the imaging test result is classified as 'outside normal limits' OR (b) IOP>21mmHg OR (c) visual acuity is 6/18 or poorer. Imaging test results which did not provide an overall classification were included as abnormal. The corresponding reference standard definition is a clinical decision not to discharge the patient.

The flow of study participants according to triage sensitivity analysis 8 is shown in Figure 21 with respective numbers of referral, not for referral and no result cases by triage test, and the corresponding reference standard finding shown. Of the 933 participants for whom all four tests were performed, 481 were not discharged and 562 were discharged. The discharge status was missing for 10 participants. The diagnostic performance for the four tests is given in Table 39. Results showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (85.1%, 95% CI (81.8,87.9) but lowest specificity (35.1%, 95% CI (31.8,40.4)), GDx had the lowest sensitivity (61.9%, 95% CI (57.8,65.9)) but the highest specificity (55.6%, 95% CI (50.2,60.9)), and the other two tests provided intermediate results (HRT-GPS values were very similar to the HRT-MRA results, and OCT had the second lowest sensitivity (72.9%, 95% CI (69.1,76.5)) but the second highest specificity (42.9%, 95% CI (37.7,48.3)). Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 2.03 for OCT to 3.80 for HRT-GPS.

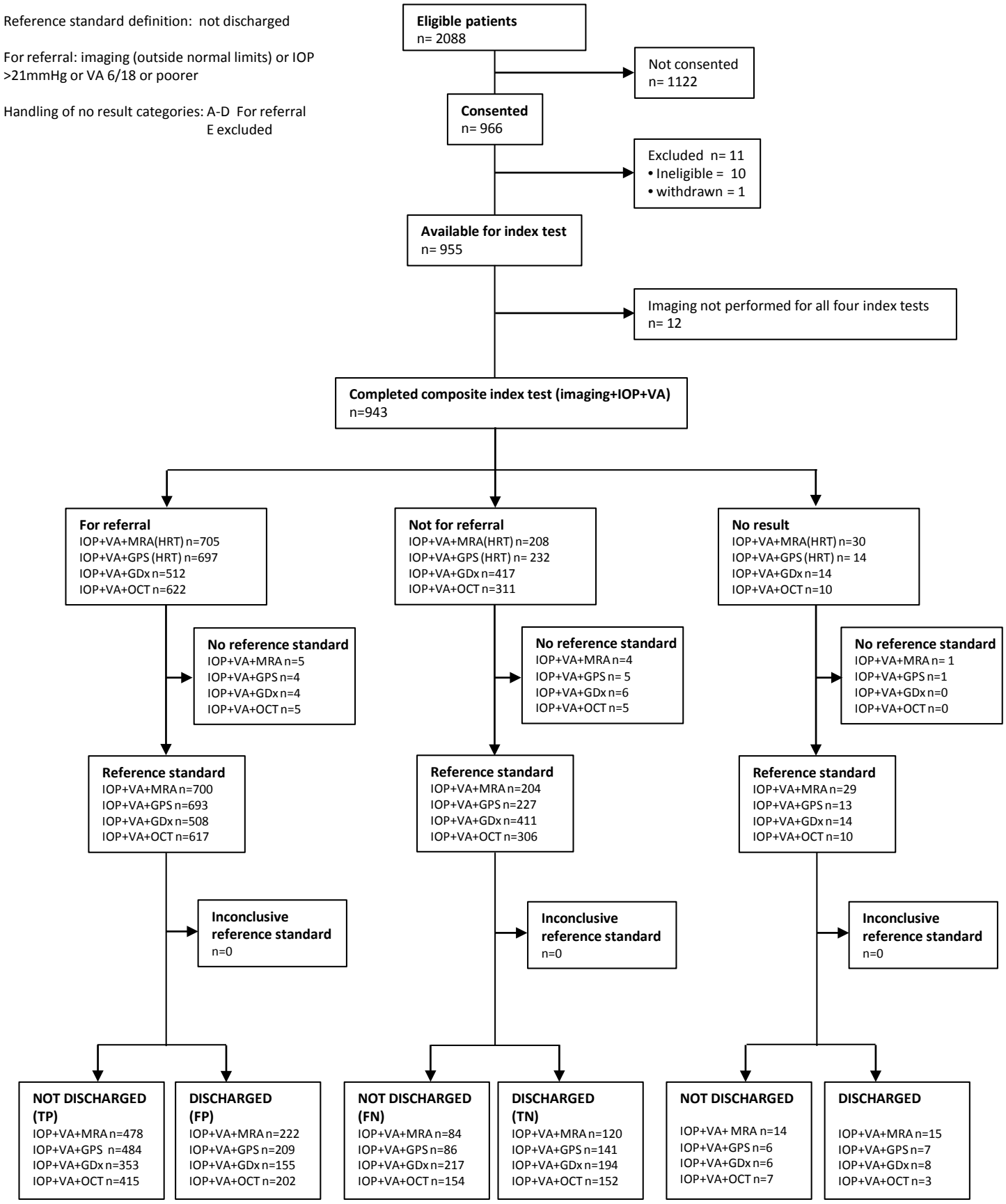


Figure 21 Flow diagram: triage sensitivity analysis 8

Table 39 Diagnostic performance: triage sensitivity analysis 8

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	85.1	81.8	87.9
	Specificity - %	35.1	30.0	40.4
	Positive Likelihood Ratio	1.31	1.20	1.43
	Negative Likelihood Ratio	0.43	0.33	0.54
	DOR	3.08	2.23	4.24
HRT-GPS	Sensitivity - %	84.9	81.7	87.8
	Specificity - %	40.3	35.1	45.6
	Positive Likelihood Ratio	1.42	1.30	1.56
	Negative Likelihood Ratio	0.37	0.30	0.47
	DOR	3.80	2.78	5.19
GDx	Sensitivity - %	61.9	57.8	65.9
	Specificity - %	55.6	50.2	60.9
	Positive Likelihood Ratio	1.39	1.22	1.59
	Negative Likelihood Ratio	0.68	0.60	0.79
	DOR	2.04	1.55	2.67
OCT	Sensitivity - %	72.9	69.1	76.5
	Specificity - %	42.9	37.7	48.3
	Positive Likelihood Ratio	1.28	1.15	1.42
	Negative Likelihood Ratio	0.63	0.53	0.76
	DOR	2.03	1.53	2.68

Discussion

Four composite triage (imaging, IOP measurement and visual acuity assessment) tests were compared with regards to their diagnostic performance for determining who should be referred for further assessment or discharged using the GATE population of referrals to a glaucoma clinic in secondary care.

The sensitivity and specificity of the four triage tests incorporating each of the imaging technologies along with IOP and VA for the default triage analysis and sensitivity analyses (see Table 29 for details) are summarised in Figure 22 and 23 respectively.

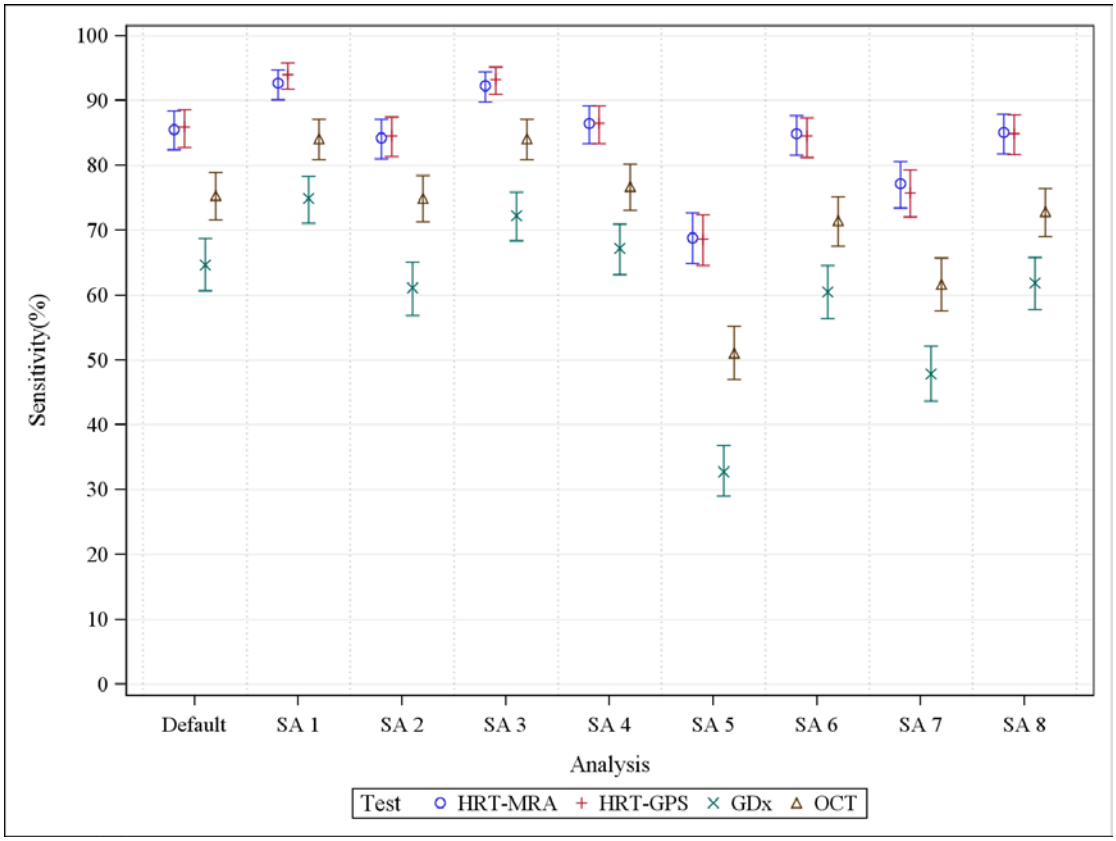


Figure 22 Summary of the sensitivity of the composite test across all triage analyses

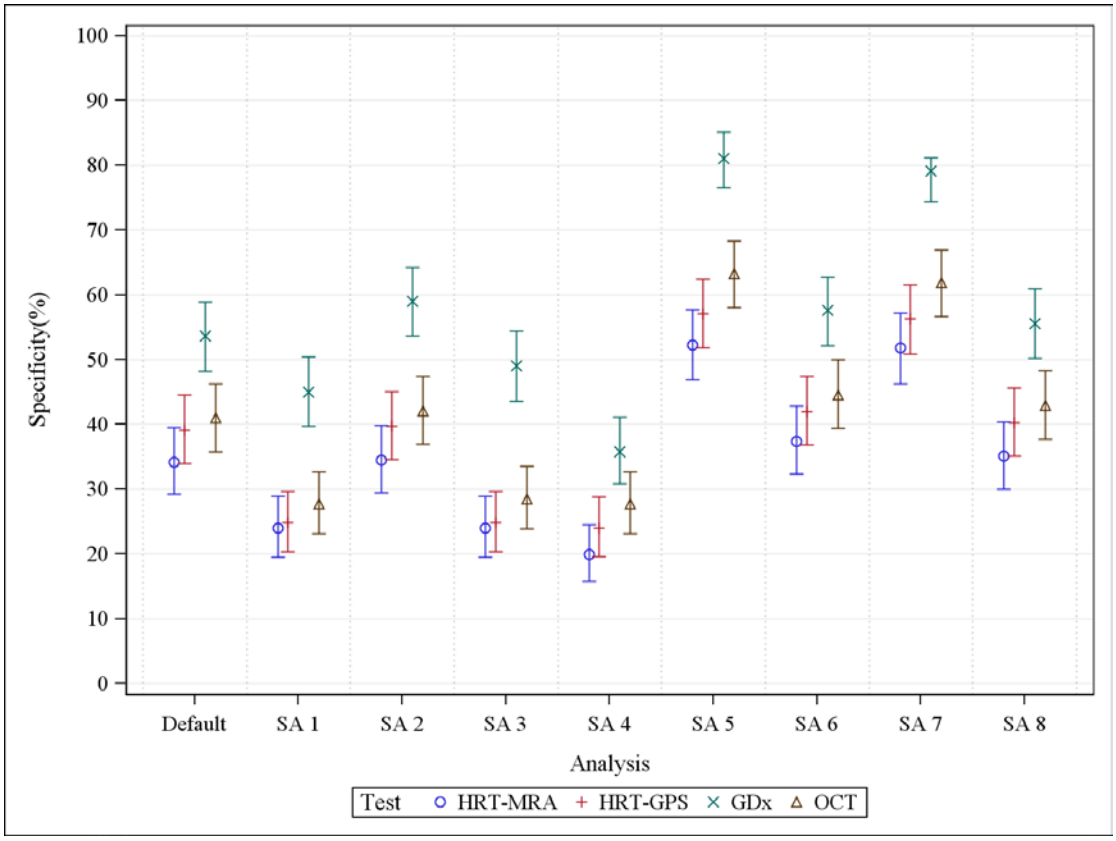


Figure 23 Summary of the specificity of the composite test across all triage analyses

All four triage tests had value in terms of ruling in and ruling out the need for referral on to a consultant ophthalmologist. The diagnostic performance of the triage tests differed with substantial differences in the ability to correctly detect those who need to be referred and those who do not. HRT-GPS and HRT-MRA had consistently the highest sensitivities across analyses but at a cost of lower specificity compared to other tests. HRT-GPS had the slightly higher specificity. In contrast GDx consistently has the best specificity though the lowest sensitivity. HRT-GPS results were typically similar to HRT-MRA as might be anticipated given they use the same machine. OCT generally had similar levels of sensitivity and specificity. The choice of which triage test is to be preferred reflects the inherent trade off regarding diagnostic testing where the desire to refer onwards when referral is needed must be balanced again the desire to discharge those who do not need a further assessment. A formal assessment of this trade-off and the consequences in terms of health outcome and costs is covered in Chapters 6 and 7.

The triage was formed from three components, an imaging test as evaluated in Chapter 4, a measurement of IOP and visual acuity measurement. The elements were combined in an additive manner where an individual was referred if any one of the three components met the respective referral criteria. A number of sensitivity analyses were carried out to assess the robustness of the findings of this default triage analysis. Varying the imaging test definition of a positive result by including the borderline category of imaging test result was carried out; this had the anticipated impact of improving the detection of glaucoma though at the cost of more non-glaucoma cases being falsely identified as having glaucoma. This resulted in very high detection of glaucoma for HRT-MRA, HRT-GPS and high sensitivities for GDx and OCT but the consequence of lower specificities (GDx had a higher specificity value than the other three triage tests). Additionally, the impact of using the classification from the imaging test when the quality criterion was not met was assessed. The impact was at most a small reduction in sensitivity with an increase in specificity (only GDx had more than a nominal change in values). The added value of the IOP and visual acuity components was assessed by dropping one of the respective components, varying the cut-off used to define abnormality, and for the IOP component, using the referral IOP measurement in place of the ophthalmologist's. Removal of the IOP component had a noticeable impact upon the diagnostic performance with exclusion leading to a reduction in sensitivity though a gain in specificity. Modifying the IOP cut-off value changed the balance in terms of sensitivity and specificity as anticipated. When the referral IOP was used in place of the ophthalmologist's IOP the specificity was reduced. Such an impact is unsurprisingly given the known variability in IOP measurements[41] and the use of an absolute cut-off will lead to a regression to the mean effect when another measurement is taken (in this case by a different observer). Removing the visual acuity component had very little impact upon the diagnostic accuracy with a slight reduction in the sensitivity and corresponding increase in specificity. This impact may have been limited by the method of data collection (referral letter quotation) as opposed to complete data capture of a new visual acuity measurement).

A number of assumptions underpinned the analysis and interpretation of these results in addition to those highlighted previously for diagnoses analyses. The reference

standard here was the clinical decision to discharge or not which will vary to some degree between individual clinicians and centres according to policies and practices (perhaps most noteworthy for individuals with suspected glaucoma). Components of the triage test were combined in an additive manner which reflects an implicit desire to favour sensitivity over specificity. No other options were assessed though arguably this approach reflects clinical practice. The use of the ophthalmologist's measurement does not reflect the reality of how a triage system would be implemented where, if a measurement was taken in hospital eye services, it would be by another individual (e.g. technician). Using the referral IOP did have a substantial impact though most if not all of this impact might be attributed to the inevitable variability between measurements taken at different times by different observers and the impact of regression to the mean. The finding does suggest there is value in taking a measurement upon referral to hospital eye services.

Chapter 6 Economic evaluation methods

The objective of this chapter is to present the economic evaluation of four automated optic nerve and retinal nerve fibre layer imaging tests (HRT-MRA, HRT-GPS, GDx and OCT), hereafter referred to as imaging technologies. These were evaluated in the GATE study as triage diagnostic stations in hospital eye services (secondary care), compared with current practice, for patients referred to hospital eye services for possible glaucoma. The triage diagnostic station included an imaging test, a visual acuity test, and an intraocular pressure measurement.

The model

The cost-effectiveness of the different imaging technologies and their subsequent care management pathways was assessed using a multi-state Markov model. As glaucoma is a chronic condition, which progresses slowly over time, the model reflects the timing of both diagnostic testing and disease progression. This approach allowed modelling of the logical and temporal sequence of events (e.g. diagnosis or monitoring visits) following the initial diagnostic strategy.

Typically, Markov models have states (e.g. Markov states) in which individuals stay for a period of time called a 'cycle'. The cycle must be a period relevant to the condition considered (e.g. 6 months, 1 year). At the end of each cycle, individuals can remain in the state in which they started the cycle or move to a different state. The probabilities of moving from one state to another are called transition probabilities. In each state, the model will assign costs and benefits for each individual according to different interventions and/or time spent in the state. In these models, there must be at least one absorbing state, typically death, from which the individual will not be able to leave. The sum of the cost in each year and the product of the utilities in each year were summed over fifty years of the simulated patient cohort to compute total cost and quality adjusted life year (QALY) for that cohort.

The purpose of this model was to compare and contrast different imaging technologies (used as part of a wider triage station) for the identification of patients who should be referred for a clinician led diagnostic examination. We can thus

compare and contrast these with standard care where all patients receive a clinician led diagnosis based on clinical examination and visual field (VF) assessment (automated perimetry). The model was constructed such that different sensitivities and specificities of each diagnostic strategy would determine if glaucoma was correctly identified or not, the health state patients would move to and the associated progression of any underlying glaucoma. The consequences could then be considered in terms of the monetary costs (of testing and subsequent management of the patient's condition) to the NHS and in terms of the effects on quality of life (by assigning utility weights). Combining these data with information of the probabilities of events occurring over time enabled cost, patient outcomes and quality adjusted life years (QALYs) to be estimated for a hypothetical cohort of patients undergoing each triage strategy.

The results of the model are presented in Chapter 7 and are presented as incremental cost per QALYs and incorporate: (i) costs (of testing) and diagnostic outcomes, (ii) costs (of testing and subsequent management) and (iii) QALYs.

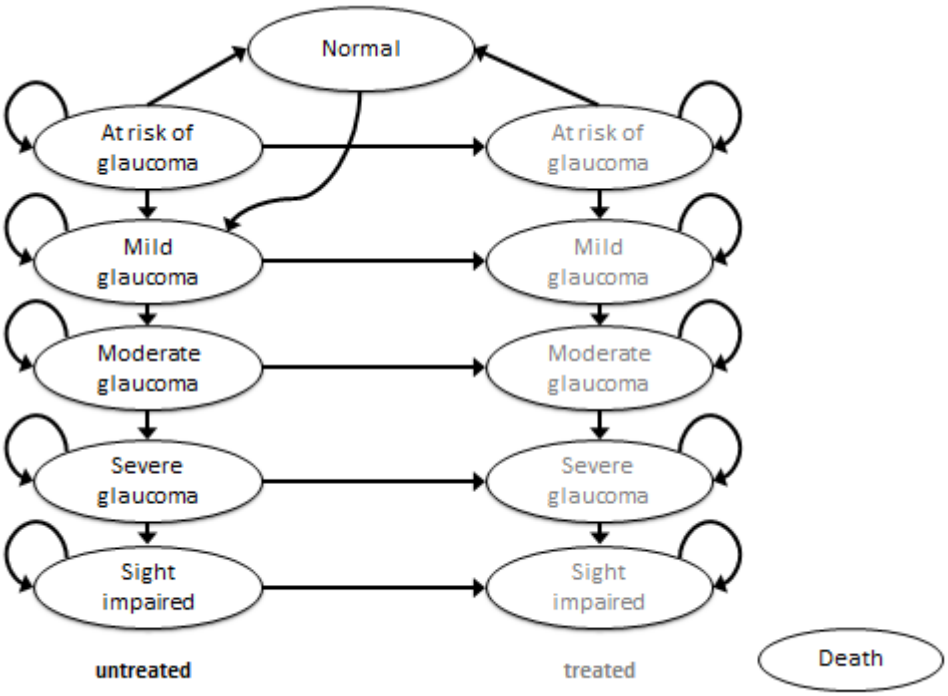


Figure 24 Schematic diagram of model states and possible transitions

Figure 24 shows the possible health states in ovals, while the arrows show the possible directions in which individuals can move at the end of each cycle, depending on the transition probabilities. The states considered in the model were those thought to reflect possible paths for normal, ‘at risk of glaucoma’ and glaucoma individuals of different stages (*see below*). Each state, other than normal and death, is divided into two categories. The treated states on the right hand side of Figure 24 represent those individuals whose condition is identified and being treated and the untreated states on the left hand side represent those individuals whose condition is not yet identified and thus they are not receiving treatment. The treatment health states refer to treated disease at each stage of glaucoma. The modality of treatment, either IOP-lowering eye drops, laser or surgery or any combination thereof, is not specified for a glaucoma related treatment state. A treatment state refers to any modality or combination treatment for each stage of glaucoma severity. There are three treatment states for the three stages of manifest glaucoma and a treatment state for sight impairment. The ‘at risk of glaucoma’ treatment state includes those individuals who are glaucoma suspects and those who have OHT and PAC. Among the “at risk of glaucoma” group we have assumed that all patients with OHT will be treated in the same way and incorporates annual outpatient appointments for observation with all OHT individuals receiving continuous Latanoprost (eye drops, once a day).

According to the individuals’ underlying condition, each individual will start in the model in a normal state, an untreated ‘at risk of glaucoma’ state or an untreated glaucoma disease state (mild, moderate, severe or sight impaired). Each individual will then enter a diagnosis process that will differ according to the compared strategies where their diagnosis will be made (i.e. for current practice in the form of consultant led diagnosis and care or a triage station including one of the imaging technologies under consideration, see above). The sensitivity and specificity of each diagnostic strategy determines the Markov state an individual will move to. In particular, it will determine if an individual enters a treated or untreated disease state and the possible transitions associated with these. In general, as time passes, the normal or ‘at risk of glaucoma’ individuals could develop glaucoma, while those with glaucoma could progress onto a more severe disease state until they eventually become visually impaired.

Glaucoma is not reversible and this is reflected in the model (*see Figure 24*). However, individuals can return to a normal state after a number of model cycles within the 'at risk of glaucoma' Markov state. The absorbing state in the model is death. Any individual can move into this state from any other state in the model.

The model allows for a cohort of the population, some with glaucoma, to pass through different diagnostic strategies. The intuitive idea behind the model is to identify the strategy that leads to the largest proportion of individuals with glaucoma correctly diagnosed and being in treatment to reduce disease progression and visual loss.

Definition of health states used in the model

Glaucoma states were defined in terms of severity of disease namely mild, moderate and severe glaucoma and sight impaired. The agreed glaucoma severity definitions used for the GATE study data collection was used for the economic model (*see Table 40*). Furthermore, an additional disease state defined as 'at risk of glaucoma' was included in the model to represent those individuals who do not have manifest glaucoma but have a higher risk of developing glaucoma (glaucoma suspects, those with OHT and those with primary angle closure).

Table 40 Definition of health states for the economic evaluation

"At risk of glaucoma" health state: glaucoma suspect, ocular hypertension, or primary angle-closure	
Glaucoma Suspect	Either the optic disc, or visual field, or both, have some features that are suggestive of glaucoma but may also represent a variation of normality (with or without high IOP)
Ocular Hypertension	Both the visual field and optic nerve appear normal in the presence of elevated pressure > 21 mmHg
Primary angle-closure	Closed anterior chamber angle (appositionally or synechial) in at least 270°, and at least one of the following two: IOP > 21 mmHg and/or presence of peripheral anterior synechiae. Both visual field and optic nerve appear normal
"Glaucoma": different health states according to mean deviation index (MD) of the visual field test	
"Mild glaucoma"	Evidence of glaucomatous optic neuropathy and a characteristic visual field loss. MD better than or equal to -6 dB
"Moderate glaucoma"	Evidence of glaucomatous optic neuropathy and a characteristic visual field loss. MD between -6.01dB and -12 dB
"Severe glaucoma"	Evidence of glaucomatous optic neuropathy and a characteristic visual field loss. MD worse than or equal to -12.01 dB
"Sight impaired" health state: sight impaired and severely sight impaired	
Sight impaired	Poor visual acuity (3/60 to 6/60) with full field of vision; or slightly reduced visual acuity (up to 6/24) and reduced field of vision or blurriness/ cloudiness in central vision; or relatively good visual acuity (up to 6/18) but significantly reduced field of vision
Severely sight impaired	Very poor visual acuity (less than 3/60) with full field of vision; or poor visual acuity (between 3/60 and 6/60) and severely reduced field of vision; or slightly reduced visual acuity (6/60 or better) and significantly reduced field of vision

Description of the healthcare diagnostic strategies and management pathways considered within the model

The care pathways modelled within the Markov model following diagnosis were developed in consultation with the study team and the independent steering committee members. The main study team for this element of the work included two ophthalmologists (AA-B, JB), and three health economists (RH, PM, JG), and a health services researcher (KB). Over a number of meetings, the group mapped out the sequence of events for patients potentially eligible for treatment or monitoring following the diagnostic strategies under consideration. Additional information came from our previous models in this area, notably our model comparing alternative screening strategies for open angle glaucoma[19] reviewed guidelines and expert opinion. These care pathways were then presented to the steering committee and revised to reflect the comments received.

Current practice care pathway

Patients enter the model as a cohort who have been identified with signs of e.g. possible glaucoma or OHT by a community optometrist or GP and have been referred to secondary care. Within hospital eye services, all individuals will see a nurse that will perform a visual acuity (VA) examination and a technician that will perform a visual field (VF) test. All individuals will then see a clinician (typically an ophthalmologist) who will measure IOP (using GAT), will look at the VF results and perform a fundus examination to examine the optic disc and the posterior retina. Figure 25 shows the care pathway.

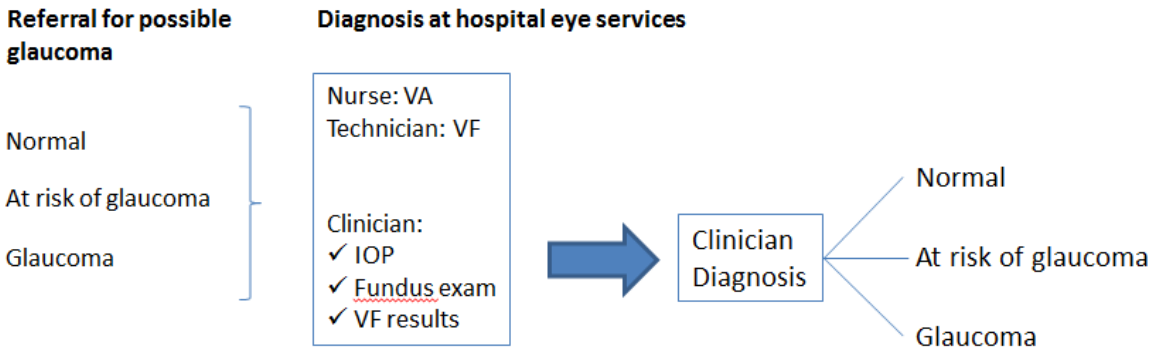


Figure 25 Care pathway: current practice

Considering all the clinical information the clinician will decide on a diagnosis as described in Chapter 2 (Methods). For the purpose of the model these diagnoses have been grouped into five health states described further in Table 40: Mild Glaucoma; Moderate Glaucoma; Severe Glaucoma; At risk of glaucoma; Normal. Furthermore, the 'at risk of glaucoma' health state includes those with a diagnosis of OHT or glaucoma suspect or PAC.

Individuals who are diagnosed by the clinician to be in the normal health state are discharged from secondary care. Individuals diagnosed with glaucoma remain in secondary care under treatment and enter the relevant glaucoma treated health state. Individuals diagnosed as 'at risk of glaucoma' also remain in secondary care and enter into the 'at risk of glaucoma' treatment state. The subset of 'at risk of glaucoma' patients with OHT are all assumed to be on treatment.

Triage care pathway

As described in Chapter 2, the triage pathway used IOP, imaging and visual acuity (VA) to identify patients who could be discharged from secondary care if all tests were normal. IOP and VA are routinely collected in primary and/or secondary care and used to inform the clinical decision-making process as to whether to discharge a patient or not. At HES the individuals will be seen by a nurse that will perform VA examination and IOP measurement. They will also be seen by a technician who will perform the index (imaging) test (HRT-MRA, HRT-GPS, GDx or OCT depending on triage strategy). Figure 26 shows the care pathway for the triage strategies.

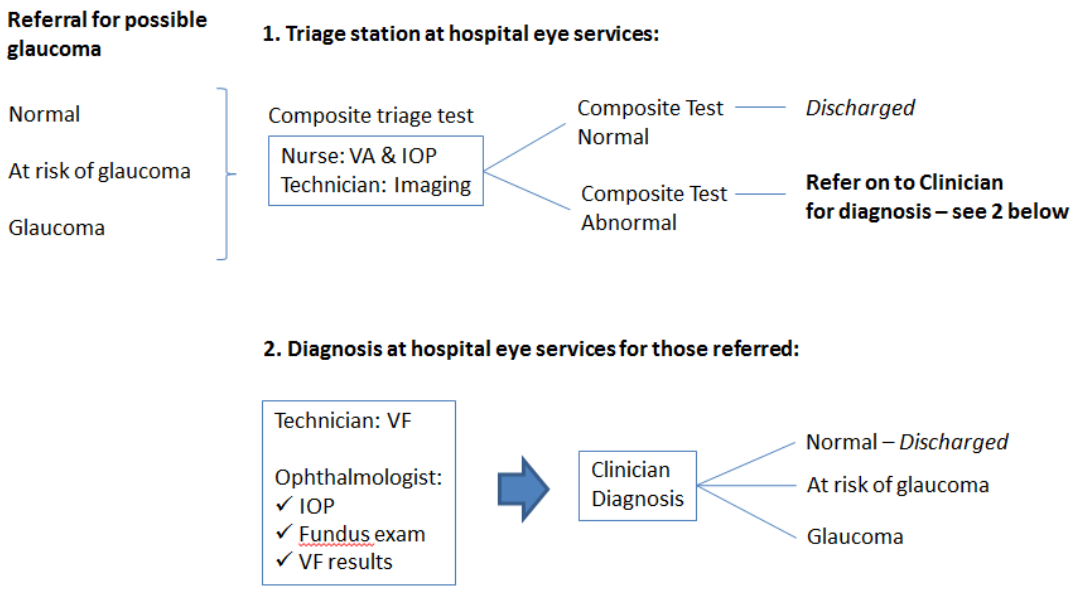


Figure 26 Care pathway: triage strategies

Results of these three examinations are combined into a composite triage test result as follows. If any of the VA or IOP or imaging test results are abnormal then the composite test is assumed to be positive or abnormal. Only if all three (VA, IOP and imaging test) are normal is the composite test result negative or normal. Individuals with normal (negative) composite triage test results are discharged from secondary care. Individuals with abnormal (positive) test results are referred to the clinician to make a diagnosis. Definitions of abnormal (positive) test results for the elements of the composite test are as follows: IOP>21mmHg, VA 6/12 or worse, imaging technology classification abnormal or borderline.

Individuals who have been discharged with normal (negative) composite triage test results can either be truly normal (true negative) or can be incorrectly diagnosed as normal when they do in fact have disease (false negative). Individuals with an abnormal (positive) composite triage test result are then referred to the clinician who will make a definitive diagnosis. Perfect information by the clinician is assumed in the model, therefore, individuals will be correctly identified as glaucoma (e.g. mild, moderate, severe or visually impaired), at risk of glaucoma or without any of these conditions (e.g. normal). Normal individuals are discharged while all others are kept under monitoring or observation. The perfect information assumption is explored in

sensitivity analysis with the possibility of misdiagnoses by the clinician (e.g. false positive and false negative results).

Model strategies

Five diagnostic strategies are explicitly considered in the model (Table 41). The comparator in the model reflects the current practice. In this strategy, all patients referred to secondary care for possible glaucoma see a clinician for diagnosis of their condition.

Four diagnostic imaging technologies (HRT-MRA, HRT -GPS, GDx, OCT) used as part of a composite triage test which includes an assessment of IOP and VA are evaluated within the model. The diagnostic strategies and associated care pathways used in the economic model are summarised in Table 41.

Table 41 Diagnostic strategies and associated care pathways

Strategy	Triage stage (composite test)	Diagnosis stage (clinician)	Treatment	Note
Current Practice / Standard Care	N/A	VA by nurse and VF by technician. Then IOP measured (GAT) and fundus exam conducted by a clinician who will make diagnosis decision (together with VF and VA information).	Treatment/monitoring according to NICE guidelines.[27] All glaucoma patients treated according to severity. Assume all OHT treated with Latanoprost. Glaucoma suspects monitored only.	
Triage 1: HRT-MRA	HRT-MRA test by technician; IOP and VA by nurse. If all 3 tests negative discharge. If any of HRT-MRA or IOP or VA test positive refer on to diagnosis stage (clinician exam).	VF test by technician, IOP measured (GAT) and fundus exam conducted by a clinician who will make diagnosis decision (together with VF and VA information).	Treatment/monitoring according to NICE guidelines.[27] All glaucoma patients treated according to severity. Assume all OHT treated with Latanoprost. Glaucoma suspects monitored only.	• MRA positive: 'outside normal limits' or 'borderline' classification • IOP positive: >21mmHg • VA positive: 6/12 or poorer

Strategy	Triage stage (composite test)	Diagnosis stage (clinician)	Treatment	Note
Triage 2: HRT-GPS	HRT-GPS test by technician; IOP and VA by nurse. If all 3 tests negative discharge. If any of HRT-GPS or IOP or VA test positive refer on to diagnosis stage (clinician exam).	VF test by technician; IOP measured (GAT) and fundus exam conducted by a clinician who will make diagnosis decision (together with VF and VA information).	Treatment/monitoring according to NICE guidelines.[27] All glaucoma patients treated according to severity. Assume all OHT treated with Latanoprost. Glaucoma suspects monitored only.	• GPS positive: 'outside normal limits' or 'borderline' classification • IOP positive: >21mmHg • VA positive: 6/12 or poorer
Triage 3: GDx	GDx test by technician; IOP and VA by nurse. If all 3 tests negative discharge. If any of GDx or IOP or VA test positive refer on to diagnosis stage (clinician exam).	VF test by technician; IOP measured (GAT) and fundus exam conducted by a clinician who will make diagnosis decision (together with VF and VA information).	Treatment/monitoring according to NICE guidelines.[27] All glaucoma patients treated according to severity. Assume all OHT treated with Latanoprost. Glaucoma suspects monitored only.	• GDx positive: NFI>35 • IOP positive: >21mmHg • VA positive: 6/12 or poorer

Strategy	Triage stage (composite test)	Diagnosis stage (clinician)	Treatment	Note
Triage 4: OCT	OCT test by technician; IOP and VA by nurse. If all 3 tests negative discharge. If any of OCT or IOP or VA test positive refer on to diagnosis stage (clinician exam).	VF test by technician; IOP measured (GAT) and fundus exam conducted by a clinician who will make diagnosis decision (together with VF and VA information).	Treatment/monitoring according to NICE guidelines.[27] All glaucoma patients treated according to severity. Assume all OHT treated with Latanoprost. Glaucoma suspects monitored only.	• OCT positive: 'outside normal limits' or 'borderline' classification • IOP positive: >21mmHg • VA positive: 6/12 or poorer

Estimation of parameters used within the model

This section summarises the parameter values used in the economic evaluation model.

Data regarding the cohort in terms of prevalence, incidence and progression are reported first, followed by diagnostic triage test performance data, with subsequent sections regarding data on cost and utilities reported respectively.

Cohort data: prevalence, incidence and progression data

Table 42 shows data on prevalence, incidence and progression of glaucoma used in the model.

Table 42 Cohort data: prevalence, incidence and progression

Probability	Value	Source
Cohort start age	40	Base-case assumption
Prevalence of glaucoma	0.17	GATE study
Proportion of normal	0.412	GATE study
Prevalence of 'at risk of glaucoma'	0.418	GATE study
Proportion of glaucoma mild	0.523	GATE study
Proportion of glaucoma moderate	0.302	GATE study
Proportion of glaucoma severe	0.174	GATE study
Incidence of glaucoma		
50 years old	0.0003	Burr 2007[19]
60 years old	0.0008	Burr 2007[19]
70 years old	0.00181	Burr 2007[19]
80 years old	0.00414	Burr 2007[19]
Progression to glaucoma mild from 'at risk of glaucoma'	0.002	Expert opinion
Progression to glaucoma moderate	0.129	Burr 2014[42]
Progression to glaucoma severe	0.048	Burr 2014[42]
Progression to sight impaired	0.042	Burr 2014[42]
Reduction in risk of progression from any medical treatment for glaucoma	0.65	Burr 2014[42]
Mortality	various	(*see Appendix 8 for interim life table*)

Prevalence data and proportion of glaucoma subjects by severity of disease was based upon the GATE study population (*see Chapter 3*). Incidence data and progression data as well as relative rate of progression between treated and untreated individuals were obtained from previous models of glaucoma management and surveillance.[19,41,42] The annual probability of having an eye test was informed by Burr[19] who used data on eye test, gender and age from the British Household Panel Survey[43] to estimate the annual probabilities of having an eye test by a community optometrist for different age groups. We utilised the average of 2 probabilities estimated in the report: 0.248 per year for those in the 40 to 59 year old range and 0.3769 per year for those in the 60 to 75 year old age range to give 0.312 visits per year.

Test performance data

Table 43 shows data on the test performances of each of the triage strategies that incorporated the different diagnostic technologies plus IOP and VA measurement and the current practice strategy in the form of clinician diagnosis. Although the imaging technology is used to define the strategy, all performance measures are calculated based on a composite test result which combines imaging, IOP and VA test results (*see Appendix 7*).

Table 43 Accuracy parameters of the triage test used in the model (*see Appendix 7*)

Probability	Value	Source
Sensitivity for all glaucoma individuals		
HRT-MRA	0.99	GATE Study
HRT-GPS	0.99	GATE Study
GDx	0.88	GATE Study
OCT	0.97	GATE study
Sensitivity for all 'at risk of glaucoma' individuals		
HRT-MRA	0.97	GATE Study
HRT-GPS	0.97	GATE Study
GDx	0.77	GATE Study
OCT	0.87	GATE Study
Specificity for all normal individuals		
HRT-MRA	0.30	GATE Study
HRT-GPS	0.28	GATE Study
GDx	0.51	GATE Study
OCT	0.35	GATE Study
Sensitivity and specificity of current practice (diagnosis by an ophthalmologist) for all individuals (glaucoma, 'at risk of glaucoma' and normal)		
Sensitivity	1	Assumption
Specificity	1	Assumption

For current clinical practice, diagnosis by a clinician was assumed to be 100% sensitive and specific. The remaining composite test performances for detecting glaucoma, 'at risk of glaucoma' and normal individuals was informed by statistical analysis of the GATE study specifically carried out to inform the economic model. Triage accuracy data for the four triage strategies (e.g. sensitivity and specificity) were calculated for glaucoma, 'at risk of glaucoma', and normal groups (*see Appendix 7*).

Estimation of costs used within the model

All costs were estimated based on resource-use inputs and unit costs for the 2012-2013 financial year and are reported in UK pound sterling. With the exception of treatment costs, which were taken from the literature, costs included in the model were estimated using a micro costing exercise or using NHS reference costs. The data used in this exercise were then subsequently checked by the steering committee members. Specific costs to the NHS relevant to the diagnostic strategies, subsequent treatment pathways, and events included diagnostic imaging, staff time, treatment, equipment and capital costs. With the exception of capital costs which were sourced from specific commercial providers, most unit costs were sourced from NHS reference costs,[23] Unit costs of health and social care[44] and Agenda for Change.[45] Where costs were not reported in 2012-13 values, they were inflated by the Hospital and Community Health Sector (HCHS) inflation index.[44]

All capital costs for each of the diagnostic imaging technologies were costed using current market prices obtained from various commercial providers to the NHS (*see explanations below*). These initial outlay costs were annuitised over the useful working lifespan of the piece of equipment (assumed to be 10 years for all equipment) applying an annual discount factor of 3.5%[46] to account for the opportunity cost of the investment over time.

The equivalent annual cost of each piece of equipment was divided through by its estimated maximum number of uses per annum (from NHS providing units and expert opinion) to give cost per use estimates.

Tables 44, 45 and 46 show the cost estimates used in the model for diagnosis by current practice, diagnosis by the triage strategies and treatment costs respectively.

Costs of diagnosis pathway: current practice

The costs of the current practice diagnostic pathway are presented in Table 44. At hospital eye services, all individuals see a nurse that will perform a VA examination and a technician who will perform a VF test. It was assumed that the VA test would take 10 minutes of a Grade 5 (midpoint scale) nurse's time and the VF test would take

15 minutes of a grade 3 (midpoint scale) technician's time. The unit costs for these were taken from Agenda for Change[45] and inflated to 2012-13 prices. All individuals will then see a clinician and the cost of this was based on the NHS reference cost (HRG WF01B) of a first consultant led ophthalmology outpatient appointment.

Table 44 Costs of current practice diagnosis used in the model

Costs	Value (£)	Source
Nurse led VA test	2.45	Agenda for Change [45]
Technician VF test	2.72	Agenda for Change[45]
Ophthalmology first outpatient appointment.	106	NHS Reference Costs[23]

Costs of diagnosis pathway: triage strategies

The costs of the GATE triage diagnostic strategies are specified in Table 45. All individuals will see a nurse who will perform a VA and an IOP test. It was assumed that this would take 10 minutes of a Grade 5 (midpoint scale) nurses' time. All patients would then go on to have one of the four index tests (diagnostic technologies). We assumed that these imaging tests would be performed by a band 3 technician (midpoint scale) and would take 15 minutes of staff time. As stated previously, the unit costs of staff time were calculated from Agenda for Change[45] and inflated to 2012-13 values.

Table 45 Costs of triage strategies used in the model

Costs	Value (£)	Source
Triage appointment costs		
Nurse led VA & IOP test	2.45	Agenda for Change[45]
Technician led Index test (e.g. OCT, GDx or HRT)	2.72	Agenda for Change[45]
Capital cost OCT diagnostic technology	1.32	Micro costed
Capital cost of HRTIII (GPS & MRA) and GDx diagnostic technologies	0.79	Micro costed
Appointment costs for those triaged and referred to the clinician		
Technician VF test	2.72	Agenda for Change[45]
Ophthalmology first outpatient appointment	106	NHS Reference Costs[23]

UK capital costs for the OCT Spectralis and HRT III diagnostic imaging technologies and associated installation and maintenance costs were obtained from Heidelberg Engineering Ltd (http://www.HeidelbergEngineering.co.uk) (personal communication, Tosh Vadhia, Regional Business Manager – South, August 2013). These initial outlay costs were annuitised over the useful working lifespan of the piece of equipment (assumed to be 10 years for all equipment) applying an annual discount factor of 3.5%[46] to account for the opportunity cost of the investment over time. The equivalent annual cost of each piece of equipment was divided through by its estimated maximum number of uses per annum (from NHS providing units and expert opinion) to give cost per use estimates. The expected number of uses per annum were based on a 253 working days per year with each use taking a fifteen minute slot over a seven and a half hour working day. This assumption was based on information provided by Moorfields Eye Hospital NHS Foundation Trust (personal communication, Edward White, Chief Ophthalmology Technician, September 2013). During the course of the study, we were unable to obtain data on capital cost of the GDx diagnostic technology. As such, we assumed that due to the competitive nature of the pricing from suppliers to the NHS, this technology had the same capital, installation and associated maintenance contract costs as the HRT III machine.

In each triage diagnostic strategy, patients who were diagnosed with a positive composite test result were referred for a first consultant led ophthalmology outpatient appointment, the cost of which was based on NHS reference costs (HRG WF01B). This outpatient visit would also involve VF testing by a technician (costs as for the standard care strategy detailed above). Thereafter, those who were identified by the ophthalmologist as being normal were then assumed to be discharged from secondary care.

Costs of treatment

Table 46 shows costs of treatment which are separated into two distinct categories: those related to glaucoma related states (mild, moderate, severe and sight impaired) and those for the 'at risk of glaucoma' state.

Table 46 Annual cost of treatment

Costs	Value (£)	Source
Glaucoma related treatment costs		
Glaucoma mild treatment	499.80	Burr 2007[19]
Glaucoma moderate treatment	562.87	Burr 2007[19]
Glaucoma severe treatment	447.44	Burr 2007[19]
Sight impaired annual cost	796.11	Burr 2007[19]
At risk of glaucoma state treatment costs		
Multi professional follow up ophthalmology outpatient appointment	87.00	NHS reference costs[23]
Latanoprost	23.64	British National Formulary[47]

Costs of treating the glaucoma related states (mild, moderate, severe, sight impaired) were taken from a related study[19] and inflated to 2013-14 prices. The authors used costs estimates based on Traverso and colleagues[48] which is a Europe-based study and includes data for the UK by severity of glaucoma. Treatment costs related to the 'at risk of glaucoma' state (i.e. individuals who are glaucoma suspects or diagnosed with OHT) were based on a number of assumptions and expert opinion and were micro costed to get an average annual cost per patient. It was assumed that all individuals in

the ‘at risk of glaucoma’ state would be given an annual multi professional follow up ophthalmology outpatient appointment, the cost of which was taken from NHS reference costs (WF02A).[23] Furthermore, it was assumed that all OHT individuals would be treated (based on advice from our expert advisory group) with Latanoprost for the rest of their lives or until their condition progressed with annual costs of £23.64.[47]

Estimation of utilities used within the model

QALYs are calculated by weighting life-years with utility values, to reflect individuals’ preferences for the health related quality of life that they experience. There are various methods and tools that can be used to elicit utility values. NICE recommends, in their methods guide,[46] the use of the EQ-5D.

Previous research by members of the study team used the EQ-5D to value quality of life states for those with mild, moderate or severe glaucoma and sight impaired and these data were used in the model to value time in these health states. The EQ-5D-3L data were obtained from responses from 640 participants with ocular hypertension and glaucoma sampled from a secondary glaucoma service.[41] Similar to the study by Burr and colleagues in 2012 who suggested that the degree of visual impairment for mild glaucoma is minimal, it was assumed that the score for those individuals in the ‘at risk of glaucoma’ state would be the same as the score for those with mild glaucoma. Table 47 shows the utility weights used in the model.

Table 47 Utility weights used in the model

Health state	Utility weight	Source/note
Normal	1	Assumption
Glaucoma mild	0.8371	Burr 2012[41]
Glaucoma moderate	0.7919	Burr 2012[41]
Glaucoma severe	0.7156	Burr 2012[41]
Sight impaired	0.5367	Burr 2012[41]
At risk of glaucoma	0.8371	Assumed equal to glaucoma mild individuals

Validation of the model

Our model was developed from that successfully used by Burr.[19] Developing the model from a pre-existing model meant that much of the structure had been previously validated. However, this approach also meant that there is no scope to make methodological changes to the way the previous model was implemented. Therefore, the Markov model was developed in TreeAge 2013 using the same core structures and transition probabilities as Burr 2007.[19] TreeAge is a frequently used tool for the type of model used in the economic evaluation and allows the documentation of our model and simplifies its' use by other researchers.

To validate the model structure where changes were made to that of Burr, a simple Markov model was developed in R in order to make comparisons with the model developed in TreeAge.

Base-case analysis

The base-case analysis was run for a cohort of 40 year-old males. Although the choice of this start age was arbitrary, it was felt that it covered the range over which diagnostic strategies for glaucoma might be considered, and would cover most of prevalent cases of glaucoma, which is an age-related disease. Gender-specific variables were not available for any of the model parameters except for mortality, and a decision was made to use male mortality rates in the base-case analysis, consistent with good modelling practice, as they are a conservative assumption for this enhanced case detection study. The model was run for a range of possible prevalence values and for a 50 year time horizon. Cycle length was set at 1 year. Costs are presented in 2012-13 UK pound sterling and effectiveness in quality adjusted life years (QALYs). A discount rate of 3.5% for costs and benefits was used following guidelines for technology assessment by the National Institute for Health and Clinical Excellence (NICE).[46] Results are presented in incremental cost effectiveness ratios (ICERs). This measure is a ratio of the difference in costs divided by the difference in the effectiveness between two alternative strategies. These data can be interpreted as how much society would have to pay for an extra unit of effectiveness. Central to the assessment of cost-effectiveness is the value that society would put on gaining an additional QALY. NICE states that "Below a most plausible ICER of £20,000 per

QALY, judgements about the acceptability of a technology as an effective use of NHS resources are based primarily on the cost-effectiveness estimate. Above a most plausible ICER of £20,000 per QALY, judgements about the acceptability of the technology as an effective use of NHS resources are more likely to make more explicit reference to factors including:

- the degree of uncertainty surrounding the calculation of ICERs
- the innovative nature of the technology
- the particular features of the condition and population receiving the technology
- where appropriate, the wider societal costs and benefits.

Above an ICER of £30,000 per QALY, the case for supporting the technology on these factors has to be increasingly strong.[46] In the absence of a more definitive statement this report focuses on a willingness to pay of £30,000 for a QALY.

Sensitivity analysis

We addressed uncertainty by conducting deterministic (e.g. one way) sensitivity analyses. In consultation with the independent advisory group, the following deterministic sensitivity analyses were considered:

1. The base-case analysis assumed that the annual probability of having an eye test is 31.2%. All patients who are discharged by the diagnosing clinician, or discharged by the triage station for the triage strategies, would therefore be expected to be picked up in the community and return to the secondary care triage station approximately every 3 years. In this analysis, based on clinical opinion, the impact of changing this probability and thus the diagnostic screening interval within a range of 1 to 10 years inclusive was explored.
2. In the base case, the diagnostic triage strategies were micro-costed and included staff time and capital costs of the diagnostic technologies. However, due to the relatively large cost differential of these triage strategies compared to current practice, it was deemed appropriate to explore the effects on cost-effectiveness of introducing an NHS reference cost for a non-consultant led first outpatient appointment (£85) to the costs of the triage strategies. This was further varied

from £10 to £85 in £5 intervals to explore if this changed either the diagnostic strategies that were deemed cost-effective or the magnitude of effect.

3. The base case analysis included a cohort of men with an age of 40 years to be modelled for 50 years. The impact of modelling older cohorts of men was explored by varying the start age from 45 to 70 years old in 10 year intervals.
4. The base case analysis was conducted on the basis that all glaucoma patients and those at risk of glaucoma (including glaucoma suspects, OHT and PAC) would be monitored and treated depending on their definitive diagnosis. It was discussed and agreed in a meeting between the study team and the independent steering committee that there was a need to explore the effects of a hypothetical secondary care service where those patients diagnosed as "at risk" would be discharged from the service thus potentially reducing the diagnostic, monitoring and or treatment costs.
5. The base case analysis assumed that clinicians were 100% sensitive and specific in their diagnosis of patients. The sensitivity and specificity was varied between 0.85 and 1 respectively to explore the impact for patients' not always being seen in secondary care by an ophthalmologist with glaucoma expertise and thus having 100% diagnostic accuracy.
6. A threshold analysis was conducted in order to explore the impact of increasing the costs of the triage strategies and discharging those patients that are given a diagnosis of "at risk" of glaucoma.
7. The base case analysis incorporated point estimates for the sensitivities and specificities of each of the imaging technologies that were estimated from the GATE study. We varied sensitivity and specificity of each triage strategy to create a best case diagnostic scenario (+ 10% sensitivity and + 5% specificity) and a worse case diagnostic scenario (-10% sensitivity and -5% specificity) for each of the imaging technologies as shown in Table 48 to explore the impact on the incremental cost effectiveness ratios. These values were decided upon by the research study team on the basis of variations in the confidence intervals in the base case analysis.
8. The base case analysis assumed the prevalence of glaucoma in the referred population which was estimated from the GATE study. However, no referral refinement schemes were in place during the GATE study. Other measures to

improve the accuracy of glaucoma referrals are constantly being explored, with a reduction in false positive rates. The impact of adding an imaging based composite triage system to a referred population with lower false positives rates was explored by decreasing the proportion of normal in the cohort from 0.412 to 0.212 and increasing the glaucoma prevalence from 0.17 to 0.27 and the "at risk" group from 0.418 to 0.518.

9. The base case analysis assumed that the utility weights for the "at risk" health state were the same as mild glaucoma in the absence of literature addressing this issue. We explored the impact of a utility weight for the "at risk" health state being the same as normals.

Table 48 Alternative best case and worst case sensitivity and specificity values used to explore uncertainties in point estimates

	'Glaucoma' sensitivity	**'At risk' sensitivity**	**'Normal' specificity**
HRT-MRA:			
Base Case	0.99	0.97	0.3
Best case	1	1	0.35
Worst case	0.89	0.87	0.25
HRT-GPS:			
Base case	0.99	0.97	0.28
Best case	1	1	0.33
Worst case	0.89	0.87	0.23
GDx:			
Base case	0.88	0.77	0.51
Best case	0.98	0.87	0.56
Worst case	0.78	0.67	0.46
OCT:			
Base case	0.97	0.87	0.35
Best case	1	0.97	0.4
Worst case	0.87	0.77	0.3

Chapter 7 Economic evaluation results

This chapter reports the results of the cost-utility analysis for four alternative triage strategies that incorporate each of the imaging tests evaluated in the GATE study (combined with IOP and visual acuity data) to identify appropriate referrals to hospital eye services, compared to current practice where all referrals receive assessment and diagnosis by a clinician in hospital eye services. Expected cost and expected QALYs, as well as incremental cost-effectiveness ratios (ICERs), are presented for the base case analysis and for sensitivity analyses conducted to explore uncertainties. Unless stated, ICERs are reported against the next least costly non-dominated strategy.

Base case analysis

The base case analysis was conducted for a male cohort with a starting age of 40 years old who were assumed to have an eye test approximately once every three years, and clinicians in hospital eye services were assumed to have perfect diagnostic ability. Table 49 shows the cost-effectiveness results for the base case analysis. All triage strategies were less costly than the current practice strategy but the triage strategies resulted in fewer expected QALYs than the current practice strategy where a perfect diagnosis by the clinician was assumed. Triage with GDx was the strategy with lowest expected cost, followed by triage with OCT, HRT-MRA and HRT-GPS, respectively. Triage with OCT was extended dominated (i.e. a combination of triage with GDx or HRT-MRA could, in theory, produce more QALYs at lower expected costs compared with triage only with OCT alone). Triage with HRT-GPS strategy was dominated by HRT-MRA (i.e. HRT-GPS was more costly but did not produce more QALYs compared with HRT-MRA). This is further illustrated in Figure 27.

Table 49 Incremental cost effectiveness ratios: base case

Intervention	Cost(£)	QALYs	ICER
GDx	2791	19.7701	
OCT	2917	19.7746	Ext Dominated[a]
HRT-MRA	2952	19.7771	22904
HRT-GPS	2961	19.7771	Dominated[b]
Current practice	3084	19.778	156985

[a] Extended dominated: a combination of a less costly and less effective intervention, and a more costly and more effective intervention would be more efficient.

[b] An intervention is more costly but equally as effective than an intervention

Incremental cost effectiveness ratios (ICERs) were calculated for all non-dominated strategies. The ICER reported for current practice (£156,985) represents the comparison between HRT-MRA and current practice strategies. It should be noted that the interpretation of this ICER is slightly different from the usual case. In moving from current practice to HRT-MRA, savings would be expected but at the expense of lost QALYs. ..

The usual willingness to pay (WTP) threshold value for an additional QALY has been stated around £30,000 for the UK.[46] However, it is not clear what decision rule should be applied when resources are saved in exchange for fewer QALYs. One possible interpretation is that of a similar threshold (e.g. £30,000 saved at the expense of a QALY) and this has been adopted in this chapter. Therefore, with this interpretation, adopting a triage with HRT-MRA strategy would be worthwhile (e.g. resources would be freed and could be used elsewhere in the health care system to obtain QALYs at the threshold value of £30,000 per QALY).

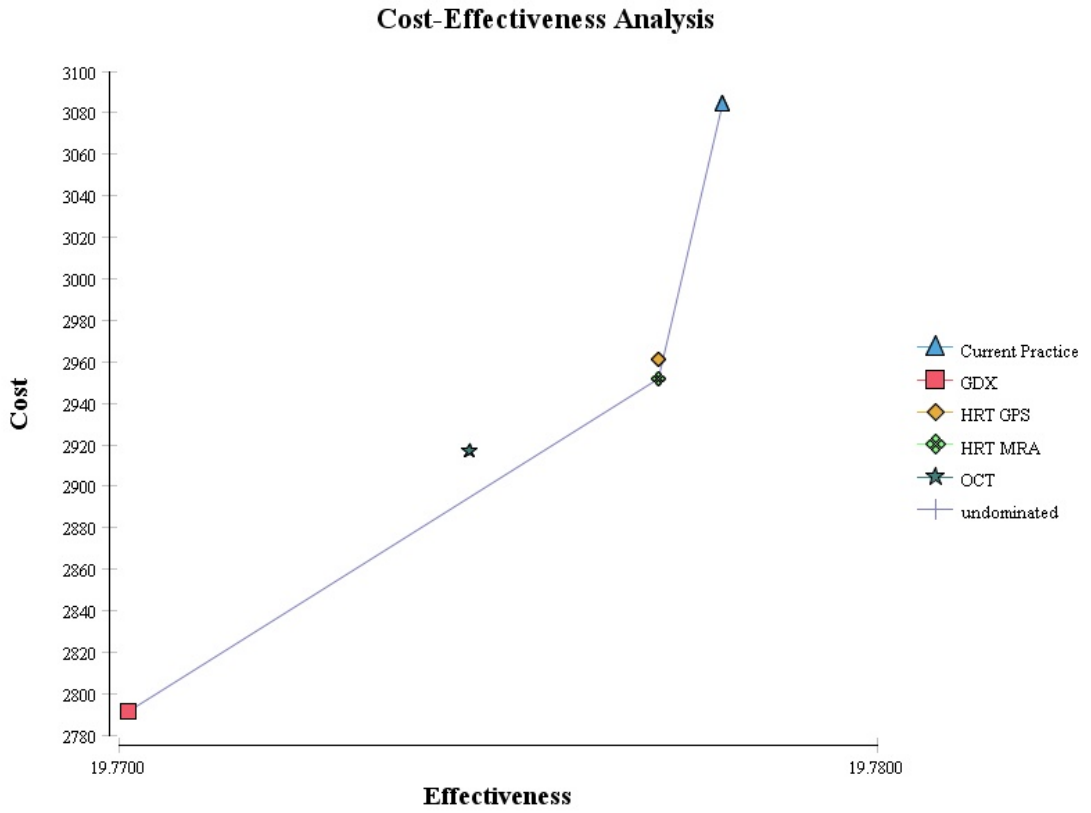


Figure 27 Base case cost effectiveness analysis results

As stated above, results show that GDx is the least costly, least effective strategy and OCT is extendedly dominated by GDx and HRT-MRA. This means that if it was possible to provide a mix of GDx and HRT-MRA, then a combination of provision of these two strategies would be dominant. Therefore, in economic evaluation we can disregard OCT from further consideration. In considering if it is worthwhile providing HRT-MRA in preference to GDx, we refer to incremental cost effectiveness ratio. Relative to GDx, the incremental cost effectiveness ratio of HRT-MRA is £22,904 and is below the typical £30,000 value considered to be cost-effective in the UK.[46] In other words, moving from a triage strategy with HRT-MRA to GDx would save only £22,904 but at the expense of a QALY. Given the £30,000 threshold, any saved resources would not be sufficient to allow the QALY lost to be regained elsewhere.

Sensitivity analyses

A number of sensitivity analyses were performed as described in the methods (*see Chapter 6).*

Changes to the annual probability of having an eye test

The base case analysis assumed that the annual probability of having an eye test by a community optometrist is 31.2%. All patients who are discharged by the diagnosing clinician, or by the triage station for the triage strategies (false negatives), would therefore be expected to be picked up in the community and subsequently referred back to the triage station at hospital eye services approximately every 3 years. We assumed that the community optometrist would identify a potential abnormality and subsequently refer the patient back to hospital eye services. In this sensitivity analysis, the impact of changing the annual probability of attending a community optometrist was explored. The annual probability was varied from 10% to 100% inclusive corresponding to a return period decreasing from 10 years to 1 year respectively. Note that as the annual probability increases, the time to return to community optometrist decreases. As shown in Table 50, as the annual probability increases, both costs and QALYs increase but the savings realised (at the expense of a QALY)decrease. For instance, for HRT-MRA changing from a 20% probability (once every 5 years) to 10% (once every 10 years), savings (at the expense of a QALY) decreased from £106,392 to £71,187. This is driven by a reduction in costs of the current practice strategy and since glaucoma progresses relatively slowly, there is only a small reduction in QALY for missed cases. Therefore, any reduction in total QALYs is more than offset by a reduction in costs.

Table 50 Incremental cost-effectiveness ratios for the 40 year old cohort by varying annual probabilities of having a community optometrist eye test

Probability	Strategy	Cost(£)	QALYs	ICER
10%	GDx	1853	19.7253	
	OCT	1960	19.7295	25407
	HRT-MRA	1989	19.7313	15503
	HRT-GPS	1992	19.7313	Dominated[a]
	Current practice	2038	19.7320	71187
20%	GDx	2451	19.7527	
	OCT	2564	19.7165	Dominated[a]
	HRT-MRA	2596	19.7596	20876
	HRT-GPS	2602	19.7596	Dominated[a]
	Current practice	2684	19.7604	106392
30%	GDx	2763	19.7686	
	OCT	2886	19.7731	27427
	HRT-MRA	2921	19.7756	13738
	HRT-GPS	2930	19.7756	Dominated[a]
	Current practice	3048	19.7765	150869
40%	GDx	2972	19.7794	
	OCT	3111	19.7837	32321
	HRT-MRA	3149	19.7862	15635
	HRT-GPS	3162	19.7862	Dominated[a]
	Current practice	3317	19.7870	208159
50%	GDx	3134	19.7873	
	OCT	3292	19.7913	39267
	HRT-MRA	3335	19.7936	18788
	HRT-GPS	3350	19.7936	Dominated[a]
	Current practice	3543	19.7943	282447

Probability	Strategy	Cost(£)	QALYs	ICER
60%	GDx	3271	19.7932	
	OCT	3449	19.7969	48375
	HRT-MRA	3497	19.7990	23231
	HRT-GPS	3516	19.7990	Dominated[a]
	Current practice	3746	19.7996	377466
70%	GDx	3393	19.7978	
	OCT	3593	19.8012	59823
	HRT-MRA	3647	19.8030	29027
	HRT-GPS	3668	19.8030	Dominated[a]
	Current practice	3937	19.8036	495605
80%	GDx	3505	19.8015	
	OCT	3728	19.8045	73741
	HRT-MRA	3788	19.8061	36177
	HRT-GPS	3813	19.8061	Dominated[a]
	Current practice	4119	19.8067	637178
90%	GDx	3611	19.8044	
	OCT	3858	19.8071	90141
	HRT-MRA	3924	19.8086	44594
	HRT-GPS	3952	19.8086	Dominated[a]
	Current practice	4297	19.8091	800615
100%	GDx	3713	19.8067	
	OCT	3983	19.8092	108907
	HRT-MRA	4057	19.8106	54140
	HRT-GPS	4088	19.8106	Dominated[a]
	Current practice	4471	19.8110	983958

HRT-GPS is always dominated regardless of the annual probability of having an eye test as this strategy is always more expensive and less effective than HRT-MRA. Moreover, when a higher proportion of the cohort comes back every year it is less clear which triage strategy should be adopted. The extreme case is for the cohort to come back every year (100% annual probability) where adopting a triage strategy with HRT-MRA would represent savings of £983,958, but moving from HRT-MRA

to OCT and from OCT to GDx would account for savings of £54,140 and £108,907 respectively but at the expense of a QALY. Therefore, at a WTP threshold value of £30,000 per QALY, GDx based triage should be adopted. It should be noted that this is an extreme case example.

Changes in the costs of the triage strategies

The costs of the triage strategies included in the base case analysis were estimated on the basis of a bottom up approach to costing and were therefore micro-costed. Due to the relatively large cost differential between the triage strategies and current practice and that NHS secondary care providers charge for a non-consultant led outpatient appointment, the effects of introducing an NHS reference cost for a non-consultant led first outpatient appointment (£85) was explored.[23] The results are presented in Table 51.

Table 51 Incremental cost effectiveness ratios for an increase to the unit costs of the triage strategies

Intervention	Cost(£)	QALYs	ICER
Current practice	3084	19.778	
GDx	3217	19.7701	Dominated[a]
OCT	3339	19.7746	Dominated[a]
HRT-MRA	3372	19.7771	Dominated[a]
HRT-GPS	3381	19.7771	Dominated[a]

[a] An intervention is more costly but less effective than an intervention that is more effective.

These data suggest that increasing the cost of the triage strategies by the inclusion of an NHS reference cost, renders them all dominated by current practice. On the basis of this result, a threshold analysis was performed to explore the maximum NHS reference cost which could be applied to the triage strategies for them to become un-dominated compared to current practice. The additional cost was varied from £10 to £85 in £3 intervals. The results are presented in Appendix 9 and suggest that as the cost of the triage strategies increases, the incremental cost per QALY of current practice decreases. Once the reference cost of the triage strategies reaches £61, all

triage strategies are dominated by current practice. The incremental cost effectiveness ratios of current practice relative to GDx or OCT are below the value typically considered to be cost effective in the UK[46] and HRT-GPS is always dominated. Triage (with HRT-MRA) is cost-effective if the NHS reference cost tariff lies below £22, given a WTP threshold of £30,000 per QALY.

Changes to the start age of the cohort

The base case analysis included a cohort of men with an age of 40 years to be modelled for 50 years. The impact of modelling older cohorts of men was explored by varying the start age from 40 to 70 years old in 10 year intervals for the same 50 years time horizon. The results are shown in Table 52.

Table 52 Incremental cost effectiveness ratios for changes in the age of the cohort at referral

Start age (years)	Intervention	Cost(£)	QALYs	ICER
40	GDx	2791	19.7701	0
	OCT	2917	19.7746	27904
	HRT-MRA	2952	19.7771	13896
	HRT-GPS	2961	19.7771	Dominated[a]
	Current practice	3084	19.7780	156985
50	GDx	2390	17.2356	0
	OCT	2503	17.2392	30995
	HRT-MRA	2535	17.2412	16016
	HRT-GPS	2544	17.2412	Dominated[a]
	Current practice	2647	17.2419	165616
60	GDx	1886	13.9949	0
	OCT	1983	13.9975	36940
	HRT-MRA	2011	13.9989	20152
	HRT-GPS	2018	13.9989	Dominated[a]
	Current practice	2098	13.9994	180864
70	GDx	1318	10.3259	0
	OCT	1395	10.3274	49717
	HRT-MRA	1419	10.3283	29376
	HRT-GPS	1423	10.3283	Dominated[a]
	Current practice	1478	10.3285	211668

[a] An intervention is more costly but equally as effective than an intervention that is less costly.

As the starting age of the cohort increases, the incremental cost per QALYs of all interventions increases. Incrementally, as the cohort ages, both costs and QALYs are decreasing but decreases in costs are outweighed by
decreases in QALYs. This can be explained by the fact that treating younger populations yields larger health gains.

Changes to the patients treated: not treating patients at risk

The base case analysis was conducted on the basis that all glaucoma patients and those "at risk of glaucoma" (including glaucoma suspects, OHT and PAC) would be monitored and treated depending on their definitive diagnosis. Due to the potential overload of hospital eye services, it was agreed there was a need to explore the effects of a hypothetical hospital eye service where those patients diagnosed as "at risk of glaucoma" would be discharged from the service thus potentially reducing the diagnostic, monitoring and or treatment costs. This analysis was conducted for all diagnostic strategies. The results are presented in Table 53.

Table 53 Incremental cost effectiveness for treating glaucoma patients only and discharging those "at risk"

Intervention	Cost (£)	QALYs	ICER
GDx	2673	19.7392	
OCT	2794	19.741	68362
HRT-MRA	2824	19.7414	83590
HRT-GPS	2833	19.7414	Dominated[a]
Current practice	2954	19.7415	752248

[a] An intervention is more costly but equally as effective than an intervention that is less costly.

Compared to base case, all strategies have lower expected costs and lower expected QALYs. This is explained by the lower proportion of individuals that are under treatment. Also, the incremental cost effectiveness ratios for all interventions have increased; in moving from current practice to HRT-MRA, HRT-MRA to OCT and OCT to GDx, savings are £752,248, £83,590 and £68, 362respectively but at the expense of a QALY.

The higher ICERs are due to less people "at risk of glaucoma" being referred to hospital eye services for re-diagnosis and further savings from the triage strategies are expected compared with base case analysis. In other words, there is not much benefit from referral to the clinician for the "at risk of glaucoma" group as the decision would always be to discharge them and wait until conversion to glaucoma in order to start

treatment. Given the value of all incremental cost effectiveness ratios, all the triage strategies except thedominated HRT-GPS can be considered cost effective given the typical thresholds used for decision making in the UK.[46]

Changes to the sensitivity and specificity of the clinician

The base case analysis assumed that clinicians were 100% sensitive and specific in their diagnosis of patients. In this sensitivity analysis, the sensitivity and specificity of clinicians was varied between 0.85 and 1 incrementally for all cohorts to explore the impact for patients of not always being seen in hospital eye services by a consultant ophthalmologist with glaucoma expertise, and thus having the possibility of reduced diagnostic accuracy. In the triage strategies, the diagnostic performance of the diagnosing clinician was not altered: for those referred (i.e. with a positive result of the triage testing) the clinician diagnosis was assumed to be perfect. The results are presented in Tables 54 and 55.

Table 54 Incremental cost effectiveness ratios for changes in sensitivity of clinicians

Sensitivity	Intervention	Cost (£)	QALYs	ICER
0.85	GDx	2791	19.7701	
	OCT	2917	19.7746	27904
	HRT-MRA	2952	19.7771	13896
	HRT-GPS	2961	19.7771	Dominated[a]
	Current practice	3025	19.7754	Dominated[a]
0.90	GDx	2791	19.7701	
	OCT	2917	19.7746	27904
	HRT-MRA	2952	19.7771	13896
	HRT-GPS	2961	19.7771	Dominated[a]
	Current practice	3046	19.7763	Dominated[a]
0.95	GDx	2791	19.7701	
	OCT	2917	19.7746	27904
	HRT-MRA	2952	19.7771	13896
	HRT-GPS	2961	19.7771	Dominated[a]
	Current practice	3066	19.7772	2068661

[a] An intervention is more costly and either equally or less effective than an intervention that is less costly

As the sensitivity of the clinician decreases from 1 to 0.95, the incremental cost per QALY of moving from HRT-MRA to current practice increases from £156,985 to £2,068,661. The incremental effect in terms of QALYs lost decreases as fewer patients are being correctly diagnosed. Similarly, incremental costs decrease; this is due to fewer patients being seen by a clinician which is only partially offset by cost increases due to more people being referred back for diagnostic testing with more expensive treatments. The incremental cost effective ratio decreases and is very sensitive to the performance of the clinician as the QALYs lost outweigh the cost gains. Once the sensitivity drops below 0.95, current practice along with HRT-GPS becomes dominated by HRT-MRA which is cheaper and either more or equally effective. This is because the cost savings realised by not being seen by a clinician are outweighed by the higher sensitivity of the alternative triage strategy (HRT-MRA).

The ICER of moving to any of the other triage strategies are below the values that are deemed acceptable in the UK to be cost effective (£30,000).[46]

Table 55 Incremental cost effectiveness ratios for changes in specificity of clinicians

Specificity	Intervention	Cost(£)	QALYs	ICER
0.85	GDx	3029	19.7706	
	OCT	3227	19.7752	42496
	HRT-MRA	3283	19.7778	22333
	HRT-GPS	3302	19.7778	1028309
	Current practice	3542	19.7789	221312
0.9	GDx	2952	19.7704	
	OCT	3126	19.7750	37961
	HRT-MRA	3176	19.7776	19709
	HRT-GPS	3191	19.7776	1278469
	Current practice	3395	19.7786	201885
0.95	GDx	2872	19.7703	
	OCT	3023	19.7748	33106
	HRT-MRA	3065	19.7773	16902
	HRT-GPS	3078	19.7774	2027006
	Current practice	3243	19.7783	177341

As the specificity of the clinician decreases from 1 to 0.85, the incremental cost per QALY of moving from current practice to another triage strategy increases from £156, 985 to £221,312. The incremental effect in terms of QALYs lost increases as more patients, although being incorrectly diagnosed, who would go on eventually to develop glaucoma or be at risk are already being monitored/treated. Incrementally, costs are also increasing due to more patients seeing a clinician and being subsequently monitored / treated. The costs are sensitive to clinicians' specificity as the cost increases are outweighed by the QALY gains. The values of incremental cost effectiveness ratios for current practice and HRT-GPS are above the acceptable threshold in the UK. That is, the savings, but with the loss of a QALY, of moving

from current practice to HRT-GPS and from this strategy to HRT-MRA exceed the WTP for a QALY and therefore a movement to HRT-MRA would be worthwhile.

Changes in the costs of the triage strategies and not treating patients "at risk"

A threshold analysis was conducted in order to explore the impact of increasing the costs of the triage strategies and discharging those patients that are given a diagnosis of "at risk" of glaucoma. Full results are presented in Appendix 9.

Adding an NHS reference cost of £85 to the cost of the triage station has the impact of current practice dominating all strategies. This prevails until the unit cost of triage station falls below £64 when both current practice and GDx become un-dominated. Reducing the reference cost to around £46, GDx becomes cost effective compared to current practice. OCT also becomes un-dominated when the unit cost of the triage strategy falls to £34. Adding a lower reference cost to the triage station makes the triage strategies with lower expected cost worthwhile. This is reflected in the values of the ICERs that, compared with the usual threshold value for cost effectiveness in the UK,[46] would render higher expected cost strategies to be not cost effective and therefore would make a triage with GDx worthwhile.

Changes to the diagnostic performance of the imaging technologies

The base case analysis incorporated point estimates for the sensitivities and specificities of each of the imaging technologies that were estimated from the GATE study. We explored the impact of changing these to a best case diagnostic scenario and a worst case diagnostic scenario for each of the imaging technologies (*see Chapter 6*) on the incremental cost effectiveness ratios. These figures were based on the confidence intervals of diagnostic performance measures used in the base case analysis and the results are presented in Table 56.

Table 56 Incremental cost effectiveness ratios for exploring triage performance best and worst case scenarios

Strategy	Cost	QALYs	ICER	Strategy	Cost	QALYs	ICER
GDx best case				**GDx worst case**			
GDx	2778	19.7717		GDx	2696	19.7683	
OCT	2917	19.7746	Ext Dominated[a]	OCT	2917	19.7746	Ext. Dominated[a]
HRT-MRA	2952	19.7771	31863	HRT-MRA	2952	19.7771	28988
HRT-GPS	2961	19.7771	Ext Dominated[a]	HRT-GPS	2961	19.7771	Ext Dominated[a]
Current practice	3084	19.778	156985	Current practice	3084	19.778	156985
OCT best case				**OCT worst case**			
GDx	2791	19.7701		GDx	2791	19.7701	
OCT	2928	19.7751	Ext Dominated[a]	OCT	2925	19.7746	Ext Dominated[a]
HRT-MRA	2952	19.7771	26326	HRT-MRA	2952	19.7771	26326
HRT-GPS	2961	19.7771	Ext Dominated[a]	HRT-GPS	2961	19.7771	Ext Dominated[a]
Current practice	3084	19.778	156985	Current practice	3084	19.778	156985
HRT-GPS best case				**HRT-GPS worst case**			
GDx	2791	19.7701		GDx	2791	19.7701	
OCT	2917	19.7746	Ext. Dominated[a]	OCT	2917	19.7746	Ext. Dominated[a]
HRT-MRA	2952	19.7771	26326	HRT-GPS	2921	19.7755	Ext Dominated[a]
HRT-GPS	2965	19.7773	89632	HRT-MRA	2952	19.7771	26326
Current practice	3084	19.778	172479	Current practice	3084	19.778	156985

HRT-MRA best case				**HRT-MRA worst case**			
GDx	2791	19.7701		GDx	2791	19.7701	
OCT	2917	19.7746	Ext. Dominated[a]	HRT-MRA	2905	19.7755	25658
HRT-MRA	2955	19.7773	26275	OCT	2917	19.7746	Dominated[b]
HRT-GPS	2961	19.7771	Dominated[b]	HRT-GPS	2961	19.7771	34269
Current practice	3084	19.778	186408	Current practice	3084	19.778	145579

[a] Extended dominated: a combination of a less costly and less effective intervention and a more costly and more effective intervention would be more efficient.

[b] An intervention is more costly but equally as effective than an intervention that is less costly.

The results show that in all scenarios, current practice always has the highest un-dominated ICER as it is always more costly and effective than any of the triage strategies. Furthermore, the order of the strategies, according to ascending cost, do not change, with GDx, even under a best case scenario, always having the lowest expected cost and the fewest expected QALYs.

When considering the performance of OCT in a best case scenario, it does not form part of the efficiency frontier and would never be considered as a triage strategy as it is always dominated by other strategies. The worst case scenario for OCT does not affect the ICER as OCT was not on the base case efficiency frontier.

Compared with base case analysis, when the best case diagnostic scenarios are applied to HRT-MRA and HRT-GPS technologies in turn, the particular triage technology either replaces the other as the dominant option or reinforces its position as the dominant technology. The results of the sensitivity analysis investigating the best case scenarios show that the choice of strategies, in order of willingness to pay, is sensitive to the relative performance of HRT-MRA and HRT-GPS. Given the assumptions in the model about consultant performance, no strategy displaces it as the most effective treatment.

When the worst case diagnostic scenarios are applied to all the imaging technologies in turn, with the exception of GDx and OCT which was dominated already, they all become dominated and are not cost effective. This can be explained by the lower cost of the GDx imaging technology. However, HRT-MRA was always un-dominated except in the worst case diagnostic scenario when it was replaced by HRT-GPS. This can be explained by the similarities in the diagnostic performance and confidence intervals of these two imaging technologies. Identical to the base case results, with the exception of reducing the diagnostic ability of HRT-MRA as stated above, GDx, HRT-MRA and current practice are all dominant strategies and have increasing incremental cost effectiveness ratios relative to each other.

In summary, in terms of GDx and current practice having the lowest and highest ICERs respectively, the base case results are not sensitive to changes in the diagnostic

accuracy of the imaging technologies. Similar to the base case analysis, current practice is not deemed cost effective in any scenario.[46] However, the results are sensitive to improvements in the diagnostic accuracy in all the imaging technologies. The corresponding ICERs rise and the best case triage strategy becomes cost effective. When the diagnostic accuracy of the imaging technologies are reduced, HRT-MRA remains the winning strategy. The exception to this is the worst case scenario for HRT-MRA where HRT-GPS becomes cost effective. This can be explained by the similarities in the diagnostic accuracy of these two technologies.

Changes to the prevalence of glaucoma and "at risk" groups in the referred population

The base case analysis assumed the prevalence of disease in the referred population was as found for the GATE study. We explored the impact of a more enriched referred population (with higher proportion of glaucoma and "at risk" patients and a lower proportion of normal) if the existing triage system was used alongside a referral refinement scheme to filter out normal cases before referral to secondary care. The results are reported in Table 57 and show higher expected costs and lower expected QALYs for all strategies compared with the base case analysis. This was anticipated as there are a higher proportion of glaucoma and "at risk of glaucoma" individuals entering the model compared with base case analysis. In addition, and also compared with base case analysis, triage strategies are less appealing (e.g. ICER for Current Practice compared with HRT-MRA of £156,985 for base case and £99,227 in Table 57); however, the ICER of £99,227 is still above the usual cost-effectiveness threshold.

Table 57 Incremental cost effectiveness ratios of increasing the prevalence of glaucoma and "at risk" groups in the referred population

Intervention	Cost (£)	QALYs	ICER
GDx	3991	19.1070	
OCT	4123	19.1131	Ext Dominated[a]
HRT-MRA	4158	19.1163	18152
HRT-GPS	4166	19.1163	Ext Dominated[a]
Current practice	4266	19.1174	99227

[a] Extended dominated: a combination of a less costly and less effective intervention and a more costly and more effective intervention would be more efficient.

Changes to the quality of life for the "at risk" health state

The base case analysis assumed a quality of life for the "at risk" health state equal to the mild glaucoma health state (QoL=0.8371). We explored the impact of assuming that the "at risk health state" would have a quality of life equal to the normal health state (QoL=1). As expected, Table 58 shows no changes in expected costs as well as higher values for expected QALYs for all strategies in the model. Moreover, there is no major impact in cost-effectiveness results with lower ICERs but close to the values observed for the base case analysis. Hence, base case results are robust to this sensitivity analysis.

Table 58 Incremental cost effectiveness ratios of changing the quality of life for the "at risk" health state

Intervention	Cost (£)	QALYs	ICER
GDx	2791	20.1788	
OCT	2917	20.1836	Ext Dominated[a]
HRT-MRA	2952	20.1864	21107
HRT-GPS	2961	20.1864	Abs Dominated[b]
Current practice	3084	20.1873	142873

[a]Extended dominated: a combination of a less costly and less effective intervention and a more costly and more effective intervention would be more efficient.

[b] An intervention is more costly but equally as effective than an intervention that is less costly.

Summary and discussion

This chapter reported the results of a cost utility analysis of alternative composite triage strategies using alternative diagnostic imaging technologies compared with current practice for patients referred to hospital eye services for possible glaucoma. The base case results suggest that HRT-MRA is the most cost effective strategy. Given that current practice represents standard care in the UK, large savings in costs (£156,985) could be made but at the expense of a QALY. Furthermore, the incremental cost effectiveness ratio for current practice relative to HRT-MRA would exceed the value that is deemed to be cost effective in the UK.

Another potential benefit is the release of clinicians' time which could be used to deliver other interventions.

Moreover, the sensitivity analysis results show triage strategies as a potential cost effective use of resources if the triage station cost does not reach £30 per triage visit. However, sensitivity analysis results were inconclusive in signalling a unique cost effective triage strategy. HRT-GPS was often dominated by HRT-MRA but the expected QALYs that these two strategies produce were almost identical with the difference in total expected costs of around £10 which is not surprising since the results are obtained from the same imaging machine.

Furthermore, on a cost effectiveness basis, GDx (or even OCT on a few occasions) could not be completely ruled out. GDx is highly specific and in a resource constrained health economy it could be an efficient use of resources. It should be noted, though, that clinically this strategy may not be acceptable to clinicians and/or patients due to its poor diagnostic performance (with low sensitivity). Determining a minimum level of diagnostic accuracy that is acceptable for clinical staff and patient was beyond the aims of this study and could be the subject of further research.

The QALY outcomes of all strategies depend only on the sensitivities of the tests to identify glaucoma and those at risk of glaucoma. The sensitivities of the different triage strategies for glaucoma are very close to each other with the exception of GDx, but there is a greater difference between the strategies in their ability to identify people at risk of glaucoma. The consequences, in QALY terms, of missing a diagnosis

of glaucoma are greater than those that result from missing a diagnosis of being "at risk of glaucoma". For these reasons the quality of life differences between triage strategies are small. The sensitivity of the triage strategies also means that the QALY differences between them and the base case scenario are small. This was to a certain extent anticipated for a study where triage strategies have similar diagnostic accuracies and a slow progression of disease. For example, this difference for the base case analysis between current practice and HRT-MRA triage strategy was 0.0008 QALYs representing less than 8 hours in full health. This small difference might make easier to accept a triage strategy where QALY would be lost in exchange of potential savings.

Furthermore, the incremental cost effectiveness of the triage strategies compared to current practice was very sensitive to costs included in the model. Unnecessary outpatient visits and associated treatment costs within current practice and, in particular, the costs of the actual triage strategies are model result drivers for the expected costs as well as the resulting incremental cost effectiveness ratios. The cost-effectiveness of any triage strategy is heavily dependent on the unit cost of the triage station. As such, all these strategies were dominated by the current practice under the plausible assumption that an NHS provider of care would charge, for the triage station, an NHS reference cost tariff corresponding to an outpatient appointment. Indeed, current practice becomes dominant when the cost of an outpatient appointment increases to £61 and above.

A key assumption used in the model was that clinicians are 100% accurate in their diagnostic ability. Relaxing this assumption further increased the incremental cost effectiveness ratios of current practice relative to other triage strategies above a level that would be deemed to be cost effective in the UK. [46] Even under extreme scenarios where the diagnostic accuracy of the triage strategies were reduced, current practice could not be deemed the most cost effective. Hence, in terms of diagnostic accuracy, no plausible scenarios rendered current practice to become the most cost effective. A probabilistic sensitivity analysis was therefore not warranted. Only when the costs of the triage strategies increased with an NHS reference cost did current practice become cost effective.

The strengths of this research are that an economic model has been developed and analysed using good modelling research practice.[49,50] The cost-effectiveness of the different imaging technologies and their subsequent care management pathways were assessed using a multi-state Markov model. This modelling approach is highly relevant as glaucoma is a chronic condition, which progresses slowly over time, allowing the model to reflect both the timing of both diagnostic testing and disease progression following the initial diagnostic strategy. Furthermore, we believe this is the first economic evaluation of these interventions to be conducted in this context.

There are limitations to this research. A key issue for the study is paucity of data regarding parameter inputs used in the model. As stated in the introduction (*see Chapter 1*), there is a lack of evidence regarding the diagnostic accuracy of imaging techniques in a triage setting and thus the parameter estimates regarding this have been based on the GATE study alone and not from multiple studies. Furthermore, the diagnostic accuracy of clinicians has been assumed to be perfect but explored in sensitivity analysis.

Only very limited data on the costs of diagnosis and treatment were available, and although efforts were made to identify the best data applicable to the UK, these were sparse. The model estimates would be more robust if further data were to become available and as previously stated by colleagues,[19] consideration should be given as to whether further primary research is needed. The model was very sensitive to the costs of the triage strategies and as stated above, adding additional costs to their unit costs renders triage not cost-effective compared to current practice.

The quality and usefulness of the economic model is dependent not only on the quality of the data, but also on the way in which the data are used. The data requirements and the use of the data were determined by the structure adopted for the model. The development of the economic model was, as described in Chapter 6, based on discussions with a number of key stakeholders. It then underwent a prolonged period of refinement during which the care pathways were critically examined and refined. The model structure applies to a UK context, and may not be relevant to other

country settings, although other strategies could be developed and readily added to the model.

As described in Chapter 6, the model structure was developed so that the assumptions made in the base case analysis could be explored in future work. For example, in the base case analysis it was assumed that the clinician would make a perfect diagnosis. The model structure has allowed for the possibility that this will not be the case and that the clinician might possibly initiate treatment when it is not required (a false positive) and fail to diagnose some cases of glaucoma(a false negative).

The model is a simplification of the care pathways that may follow. For example, the model structure does not include all possible health states that may be relevant in context, such as misdiagnosis of those at risk of glaucoma as true positives. A second simplification made in the model was the relatively small number of stages used to reflect the progression of this chronic condition. While this assumption may fail to represent the subtleties of disease progression, it was believed the health states were sufficient in number to reflect the relevant issues needed for this economic evaluation.

Estimates of the risk of progression between health states are based on data from one eye and do not necessarily represent the definition of the health states in the model, which is based on binocular visual field loss. The fellow eye may not have such advanced disease as the study eye and therefore the quality of life loss might be overestimated. While this is a limitation of the study, the alternative of using the better eye for the analysis would result in an under estimation of the risk of progressive binocular visual field loss. Furthermore, there were insufficient data to determine whether some of the parameter values varied between the stages of disease, for example the diagnostic performance of the diagnostic strategies. The model was, however, structured in such a way that should such data become available in the future, the model could be readily adapted and data incorporated.

A further simplification in the model structure was that rather than modelling the full variety of treatments available for glaucoma it has been assumed that the effect of treatment can be represented by a single relative effect size for treatment compared

with no treatment. Also, when interpreting the results of the economic evaluation it should be borne in mind that the estimates of cost-effectiveness relate to a male cohort. Gender-specific data were not available for any of the parameter estimates except for annual all-cause mortality.

Finally, there is no clear decision rule or willingness to accept threshold value to interpret cost-utility analysis results where savings are obtained at expense of QALY being lost. In this study a similar threshold value to the one often used as willingness to pay for a QALY gained was assumed (i.e. £30,000). While this is one value from many possible, in the great majority of the analysis the savings per QALY lost (ICERs) were well above this threshold. In other words, the adopted interpretation would be consistent with higher willingness to accept value should this become common practice.

Chapter 8 Discussion

GATE was a large multi-centre study designed to evaluate the performance of a triage test for patients referred to hospital eye services with possible glaucoma. The triage test would include visual acuity and intraocular pressure measurements, and one of four imaging tests from three different instruments (the HRT-3 confocal scanning laser ophthalmoscope (HRT-GPS and HRT-MRA), GDx scanning laser polarimeter, and a spectral domain optical coherence tomograph, OCT (Spectralis)). There were two diagnostic evaluations that consisted of (1) an estimation of the ability of imaging technologies to diagnose glaucoma at an eye level, and (2) an assessment of the performance of a triage test. All instruments are currently available in the NHS.

Regarding the diagnostic ability to detect and rule out glaucoma, all four imaging tests had some value but HRT-MRA had the highest sensitivity but lower specificity compared to other tests. In contrast GDx had the best specificity though the lowest sensitivity. HRT-GPS results were similar to HRT-MRA as might be anticipated given their analysis is based on imaging the same structure (i.e., the optic disc). Sensitivity for OCT was very similar magnitude to its specificity. OCT had the lowest percentage of low quality imaging results, with GDx having the highest, according to the image quality classification provided in the device software. Average time taken to conduct the tests was lowest for OCT. Patient preference tended to favour OCT followed by GDx although almost half of participants did not have a preference.

A number of sensitivity analyses were carried out to assess the robustness of the findings of the default analysis. Varying the test definition of an abnormal imaging result by including the borderline category had the anticipated impact of improving the detection of glaucoma, though at the expense of more non-glaucoma cases being falsely classified to have glaucoma. The impact of combining two imaging tests improved detection of glaucoma, but the improvement was marginal and smaller than the loss of specificity.

Regarding the triage analysis, four composite triage tests - which consisted of an imaging test, IOP measurement and visual acuity assessment - were compared with

regards to their performance for determining who should be referred onto a clinician for further assessment or discharged. All four triage tests had value in terms of ruling in and ruling out the need for referral to a clinician. The diagnostic performance of the triage tests differed substantially. HRT-GPS and HRT-MRA had consistently the highest sensitivities across analyses but at a cost of lower specificity compared to other tests. In contrast GDx consistently had the best specificity though the lowest sensitivity. OCT generally had similar levels of sensitivity and specificity. A number of sensitivity analyses were carried out that confirmed the robustness of the findings of this default triage analysis.

The economic analysis suggested that a composite triage test, introduced into the care pathway for patients referred from community with possible glaucoma, appears to be cost-effective compared with current practice where all referred patients are seen by a clinician. Our findings are based on a relatively inexpensive composite triage test (£<30) including an imaging technology, IOP and visual acuity testing.

Triage using HRT-MRA was the most cost effective strategy. Given that current practice in the model represented standard care in the UK, large savings in costs (£156,985) could be made for each QALY forgone. In the incremental cost effectiveness ratio current practice compared with HRT-MRA would largely exceed the value that is deemed to be cost effective in the UK. With the exception of GDx, the diagnostic accuracy of all the triage strategies and their unit costs are very similar. Using GDx in a triage test is the least costly and least effective diagnostic strategy but it was still cost effective compared with current practice for a number of analyses.

A variety of sensitivity analyses were conducted. The incremental cost effectiveness ratio of the triage strategies compared to current practice was very sensitive to costs included in the model. With the exception of increasing the costs of the triage stations to NHS commissioners, within the uncertainty analysis, triage was always more cost-effective than current practice. Furthermore, the present analysis is inconclusive on the decision about a particular imaging test to be included in a triage station. Further research on acceptability of the alternative imaging tests is warranted.

There are emerging models of eye care in the community that try to reduce the number of false positive referrals to hospital eye services.[51-53] Their effectiveness, efficiency and acceptability needs to be evaluated in primary research before implementing change. The GATE study provides robust data on how such services might be reconfigured.

Strengths and limitations

A number of strengths can be highlighted. GATE was a large prospective paired diagnostic study and evaluated diagnostic tests in the desired setting. The benefit of the large sample size is reflected in the precision in which the sensitivity and specificity were calculated with differences between every pair of tests identified for one if not both the sensitivity and specificity. McNemar's test was used to compare the sensitivity and specificity of the tests. Following the rationale of others in effectiveness studies, the paired comparisons were not adjusted for multi-comparisons. Even if such a correction had been applied, such was the strength of evidence there would still be evidence of differences in the diagnostic performance of the different imaging tests.

The population enrolled in GATE consisted of subjects without a known history of disease, which would reflect the potential clinical application of the triage test. Other reported studies evaluating the performance of diagnostic technologies have used patients already diagnosed with glaucoma, which has a risk of selection bias. This study recruited patients before diagnosis and the population tested had a broad spectrum of disease at presentation, from early through to severe glaucoma, and included a large percentage of healthy individuals. The healthy individuals from whom the test 'specificity' was determined were subjects referred from primary care with a possible glaucoma-related finding (either risk factor or suspected sign). Thus, the diagnostic performance reported here refers to a secondary care setting and may be different in an unselected population.

An intentional aspect of the study's design is the focus upon both the diagnostic performance of imaging tests for the identification of individuals with glaucoma, and the performance as a triage test where imaging tests would be used in conjunction

with other routine measurements (IOP and visual acuity). Both aspects are important for understanding the potential value of the imaging tests. We have also evaluated other important considerations for diagnostic technologies, such as interpretability, patient preference and time taken to perform the test.

The reference standard was provided by different ophthalmologists with glaucoma expertise. The ophthalmologists had been trained in the study protocols and agreed to a common set of criteria to define glaucoma and normality. By using different ophthalmologists working at different units the results of the study are more likely to be generalisable than results from studies performed in a single unit. The participating units are likely to be representative of the NHS practice, including two district general hospitals and three academic units of different size: relatively small (Aberdeen), medium (Liverpool) and large (Moorfields).

The economic model was developed and analysed using good modelling research practice.[49,50] The cost-effectiveness of the different imaging technologies and their subsequent care management pathways were assessed using a multi-state Markov model. This modeling approach is highly relevant as glaucoma is a chronic condition, which progresses slowly over time, allowing the model to reflect both the timing of diagnostic testing and disease progression following the initial diagnostic strategy.

Among the limitations, we recognise that diagnosing glaucoma during the very early stage of disease is challenging, and ideally a longitudinal follow-up would provide the best possible reference standard. This was proposed by Medeiros[35] who used optic nerve head progression on stereophotographic examination as the criterion for glaucoma diagnosis, but we could not contemplate this possibility in GATE as years of follow-up would have been required. The reference standard was assumed to be perfect though it is widely recognised that diagnosis of glaucoma is difficult in early disease, and uncertainty exists even amongst specialists. While consensus was sought through structured training, some assessor differences may have remained between the sites. Adding CCT information for patients referred for high IOP could potentially add valuable information and help further refine the referral pathway of such patients.

There was lack of evidence base regarding some parameter inputs used in the economic model. Only very limited data on the costs of diagnosis and treatment were available, and although efforts were made to identify the best data applicable to the UK, these were sparse. Data with respect to health utilities was available but it is unclear whether EQ-5D is sensitive enough to detect clinically significant changes in glaucoma. The model is a simplification of the care pathways that may follow, with a relatively small number of stages used to reflect the progression of this chronic condition. Estimates of the risk of progression between health states were based on data from one eye, and do not necessarily represent the definition of the heath states in the model, which is based on binocular visual field loss. A further simplification in the model structure is that rather than modelling the full variety of treatments available for OAG it has been assumed that the effect of treatment can be represented by a single relative effect size for treatment compared with no treatment.

Uncertainties

- The diagnosis, natural history and risk of conversion to glaucoma of untreated or treated patients classified as glaucoma suspects is unknown. It is likely this is a very heterogeneous group, as reflected in the categories of glaucoma suspect defined in GATE.
- The natural history and risk of conversion to glaucoma of untreated or treated patients with ocular hypertension undergoing standard care is unclear. Although there is evidence on the efficacy of treatment of OHT from large randomised controlled trials, the generalisability of their findings to routine clinical care in the NHS is ill defined.
- It is unclear how often people attend community optometrists for regular eye exams. If they have glaucoma missed by the triage it is unknown how quickly it would be detected by the optometrist and at what severity of disease. In our model we hypothesised that all those with a false negative diagnosis at the triage stage would return to hospital eye services within three years.
- The triage analysis used the IOP information provided by a consultant ophthalmologist. A triage system would rely on IOP measurements taken by a technician or a nurse, and it is uncertain whether such IOP measurements, possibly obtained with different tonometers, will be significantly different, and

what impact would it have in the performance of the triage test. The diagnostic accuracy of clinicians is uncertain. Glaucoma is diagnosed clinically, relying on the experience of the examiner, and it is likely the relative performance of the imaging technologies may be underestimated if the reference standard comparator consists of experienced glaucoma experts, as were used in GATE. Glaucoma in the NHS is diagnosed by a variety of health care professionals, including optometrists, specialist nurses, senior ophthalmologists with variable glaucoma expertise, and trainees.

- There are other OCT instruments in the market with glaucoma diagnostic capabilities and the results of this study using the Spectralis device may not be fully applicable to other OCT technologies.

Chapter 9 Conclusions

Implications for healthcare

Automated imaging technologies can be effective tests to aid in the diagnosis of glaucoma among individuals referred from the community to hospital eye services with possible glaucoma. A model of care incorporating a triage composite test for diagnosing patients referred from the community appears to be cost-effective compared with current practice. Our findings are based on a relatively non-expensive composite triage test (£<30) including an imaging technology, IOP and visual acuity testing. The most efficient strategy would include HRT-MRA imaging. However, a triage test would be associated with reduced health, and the acceptability of this option among users and clinicians has not been evaluated.

Recommendations for research

- Acceptability to patients and health care providers of implementing an efficient triage glaucoma diagnostic system but with reduced health should be explored. A qualitative or mixed methods study, e.g. including a discrete choice experiment, incorporating also public perspectives, would be suitable.
- Further data on the glaucoma disease progression under routine care, and specifically including patients classified as glaucoma suspects and ocular hypertension, associated utility, and cost of providing health care services and sight loss is needed. A long-term longitudinal cohort study would be ideal to address these issues.

- Further investigation of varying the thresholds for classification of the imaging tests beyond the standard options presented in the software could be undertaken as the standard classification may not be the one best suited to the population referred from the community to hospital eye services. Further analysis of GATE data or review of data from other relevant diagnostic studies would be able to answer this question.
- The effectiveness of implementing a triage test incorporating imaging, an IOP measurement and visual acuity requires evaluation. A longitudinal diagnostic impact study is needed.

Acknowledgements

Contributions of authors

Augusto Azuara-Blanco was the chief investigator of the study, had complete involvement and oversight of the study design, execution and data collection and provided clinical expertise, led the writing of all chapters with the exception of 4-7 and was responsible for the final report.

Katie Banister was responsible for the day-to-day management of the study, contributed to the writing of Chapter 2 and commented on all chapters, and was responsible for the production of the final report.

Charles Boachie conducted the statistical analysis and contributed to the writing of Chapter 4 and 5.

Peter McMeekin developed the structure of the Markov model and conducted the economic analyses and contributed to the writing of Chapters 6 and 7.

Joanne Gray led the economic analysis and led the writing of Chapters 6 and 7.

Jennifer Burr provided clinical advice and methodological support through all stages of the project and commented on the final report.

Rupert Bourne, David Garway-Heath and **Mark Batterbury** were clinical leads, provided expert advice on clinical aspects of the study, and commented on the final report.

Rodolfo Hernández had oversight of the health economic analysis, contributed to the writing of Chapter 6 and 7 and commented on the report.

Gladys McPherson provided technical data collection expertise throughout the study and reviewed the final report.

Craig Ramsay provided methodological oversight through all stages of the project and commented on the final report.

Jonathan Cook provided methodological oversight for the whole project, led the writing of Chapters 4 and 5, contributed to Chapter 2, and commented on the final report.

The authors would like to thank:

Pauline Garden for her data management and secretarial support, Lara Kemp for secretarial support in producing the final report, Cynthia Fraser for literature searches and referencing the final report; Kirsty McCormack for support in setting-up and initiating the study; Gladys McPherson and the programming team at the Centre for Healthcare Randomised Controlled Trials for providing and maintaining the study website; Luke Vale for invaluable advice and assistance in the development of the economic model; Ian Russell (user) and Russell Young (IGA representative) for reviewing the executive summary and lay summary; independent members of the steering committee; research and development departments at each research centre; particular thanks go to all the GATE study participants who gave of their time to take part in the study and staff who facilitated recruitment and data collection at Aberdeen Royal Infirmary, Bedford Hospital, Hinchingbrooke Hospital, Moorfields Eye Hospital and Royal Liverpool University Hospital.

Independent members of the steering committee:

Colm O'Brien (chair), Anthony King, Anja Tuulonen, Russell Young, David Wright

Project management group:

Augusto Azuara-Blanco, Katie Banister, Jennifer Burr, Jonathan Cook, Rodolfo Hernández, Kirsty McCormack, Gladys McPherson, Craig Ramsay.

Staff who facilitated recruitment and data collection:

Aberdeen: Augusto Azuara-Blanco (Principle Investigator), Jemaima Che-Hamzah, Manjula Kumarasamy, Vikki McBain, Sean Neville, Laura Park, Minimol Paulose, Patricia Peacock.

Bedford and Hinchingbrooke: Rupert Bourne (Principle Investigator), Lydia Chang Donna Gallagher, Shazia Hussein, Wendy Newsom, Paula Turnbull, Sheila Urquhart, Shalina Begum Charlotte Sullivan.

Moorfields Eye Hospital: David Garway-Heath (Principle Investigator) Ayse Barnes, Kanom Bibi, Jonathan Clarke, Cornelia Hirn, Poornima Rai, Gloria Roberti Nick Strouthidis, Ananth Viswanathan, Ed White.

Royal Liverpool University Hospital: Mark Batterbury (Principle Investigator), Anshoo Choudhary, Kathryn Nutter, Paula Burke, Jerry Sharp .

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health Directorate.

References

1 Tay E, Andreou P, Xing W, Bunce C, Aung T, Franks WA. A questionnaire survey of patient acceptability of optic disc imaging by HRT II and GDx. *Br J Ophthalmol* 2004;**88**:719-20.

2 Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006;**90**:262-7.

3 Bunce C, Wormald R. Leading causes of certification for blindness and partial sight in England & Wales. *BMC Public Health* 2006;**6**:8 Mar 2006.

4 Evans J. *Causes of blindness and partial sight in England and Wales 1990-1991*. London: Office of Population Censuses and Surveys; 1995.

5 Kelliher C, Kenny D, O'Brien C. Trends in blind registration in the adult population of the Republic of Ireland 1996-2003. *Br J Ophthalmol* 2006;**90**:367-71.

6 Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H. Causes of vision loss worldwide, 1990-2010: a systematic analysis . *The Lancer Global Health* 2013;**1**:e339-e349.

7 Bhargava JS, Patel B, Foss AJ, Avery AJ, King AJ. Views of glaucoma patients on aspects of their treatment: an assessment of patient preference by conjoint analysis. *Invest Ophthalmol Vis Sci* 2006;**47**:2885-8.

8 Fraser S, Bunce C, Wormald R, Brunner E. Deprivation and late presentation of glaucoma: case-control study. *Br Med J* 2001;**322**:639-43.

9 Maier PC, Funk J, Schwarzer G, Antes G, Falck-Ytter YT. Treatment of ocular hypertension and open angle glaucoma: meta-analysis of randomised controlled trials. *Br Med J* 2005;**331**:134-6.

10 Bonomi L, Marchini G, Marraffa M, Bernardi P, De F, I, Perfetti S et al. Prevalence of glaucoma and intraocular pressure distribution in a defined population: The Egna-Neumarkt study. *Ophthalmology* 1998;**105**:209-15.

11 Hollows FC, Graham PA. Intra-ocular pressure glaucoma and glaucoma suspects in a defined population. *Br J Ophthalmol* 1966;**50**:570-86.

12 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.

13 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.

14 Wensor M, McCarty C, Taylor H. Prevalence and risk factors of myopia in Victoria, Melbourne. *Arch Ophthalmol* 1999;**117**:658-63.

15 Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M et al. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol* 2002;**120**:1268-79.

16 Kass MA, Gordon MO. Intraocular pressure and visual field progression in open-angle glaucoma. *Am J Ophthalmol* 2000;**130**:490-1.

17 Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999;**106**:2144-53.

18 Grodum K, Heijl A, Bengtsson B. A comparison of glaucoma patients identified through mass screening and in routine clinical practice. *Acta Ophthalmol Scand* 2002;**80**:627-31.

19 Burr J, Mowatt G, Siddiqui MAR, Hernandez R, Cook JA, Lourenco T et al. The clinical and cost effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technol Assess* 2007;**11 (41)**:

20 Minassian DC, Reidy A, Coffey M, Minassian A. Utility of predictive equations for estimating the prevalence and incidence of primary open angle glaucoma in the UK. *Br J Ophthalmol* 2000;**84**:1159-61.

21 Reidy A, Minassian DC, Vafidis G, Joseph J, Farrow S, Wu J et al. Prevalence of serious eye disease and visual impairment in a north London population: population based, cross sectional study. *Br Med J* 1998;**316**:1643-6.

22 Tuck MW, Crick RP. The age distribution of primary open angle glaucoma.[. *Ophthalmic Epidemiol* 1998;**5**:173-83.

23 *NHS reference costs 2012-13 [website on the Internet]*. UK Department of Health; 2013 [accessed September 2014]. URL: https://www.gov.uk/government/publications/nhs-reference-costs-2012-to-2013.

24 Tuck MW, Crick RP. The projected increase in glaucoma due to an ageing population. *Ophthalmic Physiol Opt* 2003;**23**:175-9.

25 Claoue C, Foss A, Daniel R, Cooling B. Why are new patients coming to the eye clinic? An analysis of the relative frequencies of ophthalmic disease amongst new patients attending hospital eye clinics in two separate locations. *Eye* 1997;**11**:865-8.

26 Harrison RJ, Wild JM, Hobley AJ. Referral patterns to an ophthalmic outpatient clinic by general practitioners and ophthalmic opticians and the role of these professionals in screening for ocular disease. *Br Med J* 1988;**297**:1162-7.

27 *CG85 Glaucoma: diagnosis and management of chronic open angle glaucoma and ocular hypertension [document on the Internet]*. London: National Institute for Health and Clinical Excellence; 2009 [accessed September 2014]. URL: http://guidance.nice.org.uk/CG85/Guidance/pdf/English.

28 Bowling B, Chen SD, Salmon JF. Outcomes of referrals by community optometrists to a hospital glaucoma service. *Br J Ophthalmol* 2005;**89**:1102-4.

29 Patel UD, Murdoch IE, Theodossiades J. Glaucoma detection in the community: does ongoing training of optometrists have a lasting effect? *Eye* 2006;**20**:591-4.

30 Vernon SA, Ghosh G. Do locally agreed guidelines for optometrists concerning the referral of glaucoma suspects influence referral practice? *Eye* 2001;**15**:458-63.

31 Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D. The effectiveness of the Heidelberg retinal tomograph and the laser diagnostics glaucoma scanning system in detecting and monitoring glaucoma - systematic review. *Health Technol Assess* 2005;9**(46)**:

32 RCOPHTH Professional Standards Committee. *College statement on NICE glaucoma guidelines [webpage on the Internet]*. Royal College of Ophthalmologists; 2009 [accessed September 2014]. URL: http://www.rcophth.ac.uk/page.asp?section=594&sectionTitle=College+Statements+on+Glaucoma+Issues.

33 Tay E, Andreou P, Xing W, Bunce C, Aung T, Franks WA. A questionnaire survey of patient acceptability of optic disc imaging by HRT II and GDx. *Br J Ophthalmol* 2004;**88**:719-20.

34 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Fam Pract* 2004;**21**:4-10.

35 Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. Comparison of the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and stratus OCT optical coherence tomograph for the detection of glaucoma. *Arch Ophthalmol* 2004;**122**:827-37.

36 Alencar LM, Bowd C, Weinreb RN, Zangwill LM, Sample PA, Medeiros FA. Comparison of HRT-3 glaucoma probability score and subjective stereophotograph assessment for prediction of progression in glaucoma. *Invest Ophthalmol Vis Sci* 2008;**49**:1898-906.

37 McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;**12**:153-7.

38 Altman DG. Diagnostic tests. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence : confidence intervals and statistical guidelines.* London: BMJ Books; 2000.

39 Zhou XH, McClish DK, Obuchowski NA. *Statistical methods in diagnostic mediicne.* Oxford: Wiley-Blackwell; 2011.

40 Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Stat Med* 1998;**17**:2635-50.

41 Burr JM, Botello AP, Takwongi Y, Hernandez R, Vazquez-Montes M, Elders A et al. Surveillance for ocular hypertension: an evidence synthesis and economic evaluation. *Health Technol Assess* 2012;**16**:

42 Burr JM, Hernandez R, Ramsay CR, Prior M, Campbell S, Azuara-Blanco A et al. Is it worthwhile to conduct a randomized controlled trial of glaucoma screening in the United Kingdom? *J Health Serv Res Policy* 2014;**19**:42-51.

43 *British Household Panel Survey (BHPS) [website on the Internet]*. Institute for Social & Economic Research, University of Essex; 2006 [accessed September 2014]. URL: https://www.iser.essex.ac.uk/bhps/.

44 Curtis L. *Unit Costs of Health and Social Care 2013 [document on the Internet]*. Canterbury: Personal Social Services Research Unit.; 2013 [accessed September 2014]. URL: http://www.pssru.ac.uk/project-pages/unit-costs/2013/.

45 *Agenda for Change [wesite on the Internet]*. UK NHS Employers; 2014 [accessed September 2014]. http://www.nhsemployers.org/PayAndContracts/AgendaForChange/Pages/Afc-Homepage.aspx.

46 National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 [document on the Internet]*. London: National Institute for Health and Care Excellence; 2013 [accessed September 2014]. URL:http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9.

47 Joint Formulary Committee. *British National Formulary January 2014 [website on the Internet]*. British Medical Association and Royal Pharmaceutical Society of Great Britain; 2014 [accessed September 2014]. http://bnf.org/bnf/index.htm.

48 Traverso CE, Walt JG, Kelly SP, Hommer AH, Bron AM, Denis P et al. Direct costs of glaucoma and severity of the disease: a multinational long term study of resource utilisation in Europe. *Br J Ophthalmol* 2005;**89**:1245-9.

49 Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011;**342**:d1766.

50 Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3 Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *Value Health* 2012;**15**:812-20.

51 Azuara-Blanco A, Burr J, Thomas R, Maclennan G, McPherson S. The accuracy of accredited glaucoma optometrists in the diagnosis and treatment recommendation for glaucoma. *Br J Ophthalmol* 2007;**91**:1639-43.

52 Parkins DJ, Edgar DF. Comparison of the effectiveness of two enhanced glaucoma referral schemes. *Ophthal Physiol Optics* 2011;**31**:343-52.

53 Ratnarajan G, Newsom W, Vernon SA, Fenerty C, Henson D, Spencer F et al. The effectiveness of schemes that refine referrals between primary and secondary care--the UK experience with glaucoma referrals: the Health Innovation & Education Cluster (HIEC) Glaucoma Pathways Project. *BMJ Open* 2013;**3**:

.

Appendix 1 GATE study protocol v1.4

Versions of the study protocol

Version 1, 28th July 2010

Version 1.1, 31 January 2011 (minor typographical changes)

Version 1.2, 17 April 2012 (extension of recruitment timescale)

Version 1.3, 11April 2013 (extension of recruitment timescale)

Version 1.4 (current) 4 July 2013 (updated list of grant holders and TSC members)

Comparative study of new imaging technologies for the diagnosis of glaucoma: Protocol Approved by the Ethics Committee

HTA 09/22/111.

Applicants: Augusto Azuara-Blanco (CI), Jennifer Burr, , Rodolfo Hernández, Jonathan Cook, Kirsty McCormack, David Garway-Heath, Rupert Bourne, Mark Batterbury, Craig Ramsay.

Other participants: Professor Colm O'Brien (Dublin, Study Steering Committee Chair), and Russell Young (patient representative), Anja Tuulonen (independent steering committee member), Anthony King (independent steering committee member)

ACRONYMN: GATE (Glaucoma Automated Tests Evaluation)

INTRODUCTION

Approximately 4000 people are registered either blind or partially sighted each year because of glaucoma in the UK. Many more people have glaucoma not severe enough to be registered, but severe enough to reduce vision and quality of life.

Diagnosis of glaucoma is challenging for health professionals and many people are incorrectly diagnosed as having glaucoma by community optometrists. In fact, only 20-30% of those referred from optometric services have glaucoma and 45% of patients are discharged after their first visit. Secondary care services are very busy (accounting for nearly 10% of all outpatient attendances to the NHS) and glaucoma is a major part of the workload of any eye unit. The referral of so many healthy subjects

(less than 1/3 of referrals) is a waste of resources and causes unnecessary worry and distress to the patient. New diagnostic tests are available and are easy to perform. They are based on imaging the posterior part (fundus) of the eye where glaucoma damage can be observed. However, how well such tests perform and which is the best test to use is uncertain. This project will evaluate the performance of three new such tests to prioritise patients referred for possible glaucoma.
If one or more of the tests prove to be sufficiently accurate and easy to perform, the ophthalmologists would have more time and resources to treat patients with eye diseases.

AIM To compare directly the diagnostic performance of three automated imaging technologies within patients referred to secondary care with possible glaucoma and to explore patient test preferences.

METHODS

Study design – A within patient multi-centre comparative study of the diagnostic performance of three automated imaging technologies for glaucoma diagnosis.

The tests – 1) HRT-III: Confocal laser scanning imaging technology, employed by the Heidelberg Retina Tomograph exploits the principle of confocal laser scanning to allow quantitative structural information of the optic disc anatomy. 2) GDx-ECC: Scanning laser Polarimetry measures the retinal nerve fibre layer (RNFL) thickness. It is based on the birefringent properties of RNFL. 3) SD-OCT: Spectral Domain Optical Coherence Tomography is an optical imaging technique capable of providing high resolution, cross-sectional, imaging retina and quantifies the thickness of the RNFL. The Heidelberg Spectralis will be used in this study.

Inclusion and exclusion criteria

Inclusion criteria: Adult patients age 18 and over referred from community optometrists or General Practitioner to hospital eye services with glaucoma, or suspected glaucoma, including those with ocular hypertension, with or without associated ocular co-morbidity

Exclusion criteria: Patients referred to hospital eye services because of other ocular disease; children; patients who cannot give informed consent, patients already diagnosed with glaucoma in secondary care.

Sample and recruitment

Information about this study will be sent by post to eligible patients. The research nurse will discuss the study with the patient when they come to the clinic and will then take them through the consent process. Patients who agree to participate and sign the consent form will be enrolled. Of those patients who do not wish to participate, age and gender information will be collected.

Procedure

1. Each consented participant will undergo testing with the three technologies in both eyes. Each test produces a diagnosis of glaucoma (yes/no) without observer input. The order of testing will be randomly selected for each participant using a sheet with the randomisation order.
2. The research nurse will a) download the results to disks to be stored locally and sent to Aberdeen at the end of the project, ensuring all identifiers are removed, and b) print out the results of each test from the machines, remove the patient identifiers from the print out and write on the designated study number.
3. The participant will be asked to grade the tests in order of preference, using a standard form (see attached)
4. The participants will be examined by an experienced glaucoma clinician who will perform a comprehensive examination including intraocular pressure (IOP) measurement and biomicroscopic slit lamp examination of the optic nerve (with pupil dilated) and visual field testing (with Humphrey SITA 24-2 strategy) and provide the reference standard masked to the results of the imaging technologies. The clinician will complete a clinical data collection form.
5. The research nurse will collate the results for each participant including a copy of the VF test, complete forms for each participant, upload the information into the webpage, and post to the coordinating centre in Aberdeen. Information to be included in the webpage includes demographics (including non-participants), refraction (any

method), patient preference, need for pupil dilation, and visual field indices MD and VFI.

Data handling - Confidentiality will be maintained for all participants. All data collection sheets will have a unique study number and access to the data will be restricted to the study team.
The study data will be stored securely for a minimum of 10 years after study completion by the co-ordinating office at the University of Aberdeen (in line with current MRC guidelines).

Data analysis –
Primary diagnostic performance outcomes: sensitivity, specificity of the three imaging technologies HRT III (GPS and MRA output), GDx-ECC, and OCT will be compared using McNemar's test at the 5% significance level. Corresponding 95% confidence intervals for the paired difference will also be generated. For the primary analysis, the cut-off values for a positive diagnosis will be the respective manufacturer's recommended level. However, possible threshold effects will also be investigated in the study by varying the respective cut-off. The area under the ROC curve will also be formally compared between technologies.

Secondary diagnostic performance outcomes: ROC curves (thresholds varied for test positivity), diagnostic odds ratio, likelihood ratio, proportion of indeterminate tests, patient preference for test.

Primary and secondary diagnostic measures (area under ROC curve, likelihood ratios and diagnostic odds ratio) will be presented for each technology with appropriate 95% confidence intervals. We shall also measure the proportion of tests that are indeterminate. All analyses will follow a patient-based (referral eye) approach. For referrals where bilateral disease is suspected, a study eye will be selected at random at the analysis stage. Additionally, the level of diagnostic performance across the disease spectrum (as defined by the reference standard) will be explored.

The overall diagnostic performance of combinations of these three technologies will also be evaluated, and their relative performance. The diagnostic performance of the

tests (and corresponding combinations) will also be assessed according to the spectrum of glaucoma (mild, moderate, severe), as defined by the specialist ophthalmologist. Additionally, the impact of introducing a measure of IOP into the triage test will be explored.

A further analysis will consider the impact on diagnostic performance of using combinations of these technologies under three approaches: both positive, either positive or by generating a prognostic rule using multivariable logistic regression model. Furthermore, the impact of using these technologies in tandem with a measure of IOP to identify suspected cases of glaucoma and/or ocular hypertension for referral, as might occur in a clinic setting, will also be examined. For this analysis, disease will be defined as the presence of either condition. The analysis would follow the above comparative approach.

Economic outcomes: costs of providing the tests (initially retrieved from the four study sites), costs (of testing and subsequent management), QALYs, and incremental cost per QALY.

Economic evaluation

Modelling will be used to determine which test or combination of tests would be effective and cost-effective compared with current practice. The results of the modelling will be diagnostic performance. We will develop a new economic model which will estimate the costs and outcomes of diagnosing glaucoma in secondary care. We will consider the use of the diagnostic tests used alone or in combination in secondary care. The model will compare the diagnostic performance of the tests and effect of longer term outcomes (e.g. Quality Adjusted Life Years – QALYs). The costs of current management alternative diagnostic pathways will depend upon the findings of the diagnostic performance analyses, the literature and advice from the project team. Costs and outcomes following diagnosis (both for those with true and false diagnoses) will be derived from an updated version of an existing economic model. The perspective of the economic analysis will be that of the NHS. The results of the model will be presented in terms of (i) costs (of testing) and diagnostic outcomes, (ii) costs (of testing and subsequent management), (iii) QALYs and

incremental cost per QALY. The results will be presented as point estimates and cost-effectiveness acceptability curves (cost per QALY data). Deterministic sensitivity analysis will be combined with probabilistic sensitivity analysis to explore different types of uncertainty (e.g. impact of the scale of the service on cost-effectiveness).

Sample size considerations

The sample size calculation and analysis are based on standard McNemar diagnostic accuracy study methods. The sensitivity and specificity of each of the automated tests will be compared. A 5% significance level based upon a 2-sided test was used in the sample size calculations. A study of 897 individuals would have 90% power to detect a difference in accuracy of 9% for the primary outcome of diagnosis of glaucoma. This is based upon conservative assumptions of a probability of disagreement of 0.18 (maximum level possible), a glaucoma rate of 25% (as seen in similar populations) and a sensitivity of 86% as found in a systematic review for HRT II. Given this sample size, there would also be 80% power for detecting a 6% difference in accuracy should the sensitivity be 93% (the current best estimate from meta analyses of high quality diagnostic studies). For specificity, we would have over 90% power to detect a 5% difference. Based upon current available evidence, a rate of 6% indeterminacy of tests results was assumed which increased the sample size to 954 in total. A sample of this size would be of sufficient size for other measures of diagnostic performance (e.g. the sensitivity and specificity of individual technologies would be estimated to 95% confidence interval of width 10% and 5% respectively).

Dissemination

Participants will be offered a summary of the study findings once the study is complete. Papers will be submitted for publication to peer reviewed journals and submitted for oral/poster presentation at both international and national conferences.

We are also required to report these findings to the NIHR HTA who are funding this study.

Timelines

Pre-funding

Ethics

Start date

1 December 2010

October 22

First steering group meeting

Month 1-3

Study set-up, authorisations, protocol finalised

Month 4-32

Patient recruitment (from March 2011)

Month 12

Second steering group meeting

Month 19

Third steering group meeting

Month 32

Data cleaning and preliminary analysis

Month 32

Close down centre study processes

Month 33-36

Final analysis and reporting

Appendix 2 Information for patients

Evaluation of novel imaging techniques for the detection of Glaucoma.

Dear

Invitation to take part in a research study

I am currently undertaking a national study to evaluate the performance of three new automated eye tests for the diagnosis of glaucoma in conjunction with The Health Services Research Unit, University of Aberdeen. As you are coming to the eye out patients department for an appointment I would like to invite you to take part.

I have enclosed an information sheet about the study which will help you decide whether or not you would like to participate and would be most grateful if you could take a few minutes to read through the information.

If you agree to take part then you will be given three automated eye tests during your appointment which should take no longer than one hour to complete before being seen by the ophthalmologist as per a normal clinic appointment.

Although your involvement is very important to us we would like to stress that you are under no obligation to participate. We will be happy to discuss any aspect of the study with you at the clinic when we see you and if you have any questions about the study we will be pleased to answer them then.

Yours sincerely,

Augusto Azuara-Blanco
Consultant Ophthalmologist.

GATE study patient information leaflet

Evaluation of novel imaging techniques for the detection of Glaucoma (GATE study).

Information leaflet.

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully and discuss it with others if you wish. Ask us anything that is not clear or if you would like more information. Take time to decide whether you wish to take part. Thank you for reading this.

What is the purpose of the Study?

Glaucoma can reduce vision and quality of life but if diagnosed early it can be treated and reduction of vision prevented. The main risk factor for a reduction in vision due to glaucoma is being diagnosed late and damage to the eyesight has already begun. New promising diagnostic imaging tests are available and are easy to perform. They use a laser to explore and analyse the structure of the optic nerve head and surrounding tissues in the back of the eye. However, which test is the best to use is uncertain at present and this project will evaluate the performance of three new imaging tests. If one of the tests proves to be accurate and easy to perform, it could be implemented in the community to reduce the risk of reduced vision from glaucoma.

Why have I been chosen?

As you are attending the eye out-patient clinic for an eye examination from an ophthalmologist we would like to invite you to take part.

Do I have to take part?

No. It is up to you to decide whether you take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

What will happen if I take part?

The study is carried out alongside your appointment in the local Eye Out-Patient Clinic. Depending on routine practice at your local clinic this may be one or two appointments. You will be given a visual fields test along with three imaging tests and finally a basic eye examination by an Ophthalmologist. The addition of the imaging tests may extend your appointment time by approximately one hour.

What do the imaging tests involve?

There will be three imaging tests carried out by a research technician. The imaging tests are non-invasive and do not usually require eye drops. During the test you will have to look at a fixation light for a short period whilst a series of images of the optic nerve head at the back of the eye are acquired. The imaging tests are rapid and take approximately 10 minutes to complete. In a small number of cases we may have to place some eye drops in your eyes to dilate your pupils. This can make it easier to take the image.

What are the possible disadvantages and risks of taking part in this study?

Most people will not need their pupils dilated. If we need to dilate your pupils it can sometimes cause some temporary blurring of vision and sensitivity to light. This is, however, a routine procedure which would normally be performed as part of your eye examination

Are there any benefits to taking part in the study?

There will be no direct benefit to yourself in taking part in the study, however if any of the tests prove to be accurate and easy to perform, they could be implemented in the community.

Will my taking part in this study be confidential?

All information which is collected about you for the study will be kept strictly confidential. Information for all participants in the study will be kept for a minimum of ten years in line with current research governance arrangements and then destroyed. Only researchers involved with the study will have access to your information

What happens to the results of the study?

The results of the research will be published in relevant scientific journals and a report will be sent to the funder of the research, the NHS Health Technology Assessment programme. We would also be happy to send you a short report when the study when the research is complete. You will not be identifiable in any publications from this research.

How do I Complain?

If you wish to complain or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal NHS complaints mechanisms are available to you.

Who has reviewed the study?

This study has been reviewed and approved by the North of Scotland Research Ethics Committee.

Thank you very much for considering taking part in this research.

Contact details for further information

Central office:

GATE Study Office
Health Services Research Unit
University of Aberdeen
Health Sciences Building
Foresterhill
Aberdeen AB25 2ZD
Tel: 01224 438196
Fax: 01224 438165
Email:gate@abdn.ac.uk

Local contact details:

[Contact details]
[for local researchers]
[Affix sticker here]

GATE study consent form

(Form to be on headed paper)

Participant Study Number: ☐☐☐☐☐

CONSENT FORM

Comparative study of new imaging technologies for the diagnosis of glaucoma: the GATE study

		Please initial box
1	I confirm that I have read and understand the information sheet dated ..... (version .....) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	☐
3	I understand that relevant sections of my medical notes and data collected during the study, may be looked at by individuals from the University of Aberdeen, from regulatory authorities or from the NHS Trust/Health Board, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.	☐
4	I understand that my personal details collected during the study will be held in a secure central database, and may be subject to audit and monitoring by University of Aberdeen or NHS Trust/Health Board staff, without breaching data confidentiality	☐
5	I agree to take part in the above study.	☐

_____	_____	_____
Name of Patient	Date	Signature

I confirm that I have explained to the person named above the nature and purpose of the GATE study and the procedures involved

_____	_____	_____
Name of Person taking consent	Date	Signature

GATE study office, Health Services Research Unit, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD.
Tel: 01224 438196, Fax 01224 438165, Email: gate@abdn.ac.uk

Copies: Original to be returned to study office: 1 copy for patient; 1 copy to be filed with hospital notes

Appendix 3 GATE study case report forms

Inclusion criteria

- **Adult patients (aged over 18 years old)**
- **New referral from primary care to glaucoma clinic**

Clinic date

……..……

Study Number	Patient Name	Year of birth	Gender M/F	Info. sheet sent	Date info. sent	Consented	*If not consented, state reason (A, B, C or D)*	*Assigned Test Order*		
								HRT	GDX	OCT
		Y N		Y N						
		Y N		Y N						
		Y N		Y N						
		Y N		Y N						
		Y N		Y N						

								Y		N			Y		N					
								Y		N			Y		N					
								Y		N			Y		N					
								Y		N			Y		N					
								Y		N			Y		N					

Reasons for not including

A: Non attendance (DNA/CNA)

B: Refusal – record reason if possible

C: Missed

D: Equipment not working (please record which machine is not working)

Participant Study No

Glaucoma Automated Tests Evaluation

Research Officer Data Collection Form

CONFIDENTIAL

This study is funded by the NHS National Institute for Health Research Health Technology Assessment Programme

Research Officer Data Collection Form

Participant Study number | | | | | |

Date of Assessment | D | D | / | M | M | / | Y | Y | Y | Y |

SECTION A - PATIENT DETAILS

CHI number (Scotland only) or
NHS number | | | | | | | | | | |

Date of Birth | D | D | / | M | M | / | Y | Y | Y | Y |

Gender ☐ Male ☐ Female

ETHNIC ORIGIN

Please note the following are the main classification categories used by the Census 2001. Please ask the patient how they would describe themselves.

Black or Black British-Caribbean	☐	
Black or Black British-African	☐	
Other Black Background	☐	Please specify _____

Asian or Asian British-Indian	☐	
Asian or Asian British-Pakistani	☐	
Asian or Asian British-Bangladeshi	☐	
Chinese	☐	
Other Asian Background	☐	Please specify _____
Mixed – White and Black Caribbean	☐	
Mixed – White and Black African	☐	

Mixed – White and Asian ☐

White - British ☐

Other ☐ Please specify _____

Has patient been fully consented? Yes ☐

SECTION B – CLINICAL DATA

Referral Eye (please tick only one)

Right ☐ Left ☐ Both ☐

IOP on referral (mmHg)

Right ☐ Left ☐

Method of assessment (please tick only one)

NCT ☐

GAT ☐

Other ☐ Please specify _____

Refraction

	+/-	Sphere		+/-	Cyl		Axis
Right eye	☐	☐☐ . ☐☐	/	☐	☐☐ . ☐☐	x	☐
Left eye	☐	☐☐ . ☐☐	/	☐	☐☐ . ☐☐	x	☐

Best corrected visual acuity (Snellen)

Right eye	Left eye

Visual fields *(Humphrey 24.2)*

SITA standard or SITA fast**.** Record reliability information defined by the Humphrey

Right Eye: Reliable ☐ Unreliable ☐ Not done ☐

Fixation losses		False pos errors (%)	False neg errors (%)	+/-	MD (dB)	PSD (dB)	VFI (%)
	/						

Left Eye: Reliable ☐ Unreliable ☐ Not done ☐

Fixation losses		False pos errors (%)	False neg errors (%)	+/-	MD (dB)	PSD (dB)	VFI (%)
	/						

Printout of Visual Fields for research site file attached to CRF. Yes ☐

SECTION C – IMAGING DATA

Test order

The order that tests should be performed is found on the study website clinic log for this study number. Please record the order in which the tests were performed (1=1st,2=2nd,3=3rd)

HRT ☐ **GDX** ☐ **OCT** ☐

HRT

Start time (24hr clock) H H : M M **End time** (24hr clock) H H : M M

Were pupils dilated? Yes ☐ No ☐

Right Eye: Completed ☐ Not performed ☐ Reason _____

Left Eye: Completed ☐ Not performed ☐ Reason _____

Raw data filename _____ Raw data saved to disk ☐ Hard copy report printed MRA right eye ☐ MRA left eye ☐ GPS ☐

GDX

Start time (24hr clock) H H : M M **End time** (24hr clock) H H : M M

Were pupils dilated? Yes ☐ No ☐

Right Eye: Completed ☐ Not performed ☐ Reason _____

Left Eye: Completed ☐ Not performed ☐ Reason _____

Raw data filename _____ Raw data saved to disk ☐ Hard copy report printed ☐

OCT

Start time (24hr clock) H H : M M **End time** (24hr clock) H H : M M

Were pupils dilated? Yes ☐ No ☐

Right Eye: Completed ☐ Not performed ☐ Reason _____

Left Eye: Completed ☐ Not performed ☐ Reason _____

Raw data filename _____ Raw data saved to disk ☐ Hard copy report printed ☐
(RNFL basic report OU)

Has participant completed the GATE Participant Preference questionnaire? Yes ☐ No ☐

If No, why? _____

Participant Study No

Participant Preference Questionnaire

CONFIDENTIAL

This study is funded by the NHS National Institute for Health Research Health Technology Assessment Programme

Participant Study No

Participant Preference Questionnaire

Date of examination | | | / | | | / | | | | |

Now that you have had all three tests can you please give an order of preference from 1 for the most preferred test, to 3 for the least.

If you have no preference please tick the last box.

Optical Coherence Tomography

Scanning laser polarimetry – GDx-VCC

Heidelberg Retinal Tomography

I have no preference

Please note you may <u>not</u> have had your tests in the order above and may not remember which test is which. If you are unsure then ask the research nurse for help.

CLINICIAN CRF

CONFIDENTIAL

DO NOT LOOK AT IMAGING RESULTS

BEFORE COMPLETING THIS FORM

Participant Study number | | | | | |

Date of Assessment D D / M M / Y Y Y Y

Glaucoma Automated Tests Evaluation

Clinician Name (Capitals) _____

IOP (mmHg)

Today	Right		Left	

DIAGNOSIS **(tick only one category in each column)**	**Right**	**Left**
Glaucoma		
Disc suspect		
VF suspect		
VF+disc suspect		
OHT (normal disc and field)		
PAC (normal disc and field)		
PAC suspect (normal disc and field)		
No glaucoma-related findings		
Undetermined		

→ Severity of glaucoma

	R	**L**
Mild		
Moderate		
Severe		

Please specify reason

For glaucoma and suspects:		**R**	**L**
Please tick mechanism	Open angle		
	Angle closure		
	Other		

Co-morbidity – tick all that apply	**Right**	**Left**
AMD		
Cataract		
Neurological		
Other		

Please specify

ACTION (please tick)

Discharge? Yes ☐ No ☐

If NO please complete – tick only one box in each column

	Right	Left	**Comments**
Treat			
Monitor only			
Repeat assessment required			

Clinical diagnosis definitions

Glaucoma:

Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss**

Glaucoma severity: according to Humphrey SITA standard perimetry of a reliable VF ***:

Mild: MD better than or equal to -6 dB;

Moderate: MD between -6.01dB and -12 dB

Severe: MD worse than or equal to -12.01 dB

Mechanism:

Open angle: includes POAG, NTG,

Angle closure: includes evidence of glaucomatous optic neuropathy combined with a characteristic visual field loss, and a closed anterior chamber angle (appositionally or synechial) in at least 270°

Other: pigmentary glaucoma, pseudoexfoliation glaucoma or any other type of glaucoma

Disc suspect: appearance suggestive of glaucomatous optic neuropathy but may also represent a variation of normality, with normal visual fields (with or without high IOP)**.**

VF suspect: visual field loss suggestive of glaucoma, but may also represent a variation of normality, with normal appearance of the optic disc (with or without high IOP)

VF+disc suspect: both the optic disc and visual field have some features that resemble glaucoma but may also represent a variation of normality (with or without high IOP)

OHT: when both the visual field and optic nerve appear normal in the presence of elevated pressure,

> 21 mmHg

PAC: Closed anterior chamber angle (appositionally or synechial) in at least 270°, and at least one of the following two: IOP > 21 mmHg and/or presence of peripheral anterior synechiae. Both visual field and optic nerve appear normal

PAC suspect: Closed anterior chamber angle (appositionally without any synechiae) in at least 270°, with IOP ≤ 21 mmHg. Both visual field and optic nerve appear normal

The decision to monitor/treat will be defined in accordance with the NICE guidelines

* Evidence of optic nerve damage from any of the following: Optic disc or retinal nerve fibre layer structural abnormalities. Diffuse thinning, focal narrowing, or notching of the optic disc rim, especially at the inferior or superior poles. Documented, progressive thinning of the neuroretinal rim with an associated increase in cupping of the optic disc. Diffuse or localised abnormalities of the peripapillary retinal nerve fibre layer, especially at the inferior or superior poles. Disc rim or peripapillary retinal nerve fibre layer haemorrhages. Optic disc neural rim asymmetry of the two eyes consistent with loss of neural tissue.

** Reliable visual field abnormality considered a valid representation of the subject's functional status. Visual field damage consistent with retinal nerve fibre layer damage (e.g. nasal step, arcuate field defect, or paracentral depression in clusters of test sites). Visual field loss in one hemifield that is different from the other hemifield, i.e. across the horizontal midline (in early/moderate cases). Absence of other known explanations.

***A reliable visual fields is classified as: False positive error <15% and no evidence for learning effect or poor performance which could impact on MD value (clinical judgement). In patients with unreliable visual field, the severity of glaucoma will be based upon clinical judgement.

Appendix 4 Example imaging report outputs from the four imaging tests

HRT-MRA

Heidelberg Retina Tomograph
Regression Analysis

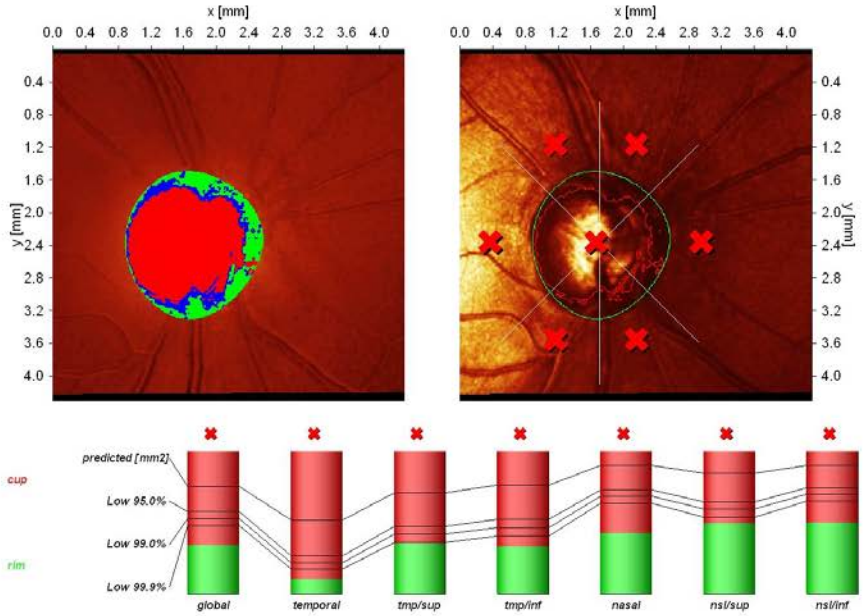
HEIDELBERG ENGINEERING

Patient:

Examination:

Scan:

OD



Moorfields Regression Classification: *Outside normal limits (*)*

(*) Moorfields regression classification (Ophthalmology 1998;105:1557-1563). Classification based on statistics. Diagnosis is physician's responsibility.

Rim Area	*global*	*temporal*	*tmp/sup*	*tmp/inf*	*nasal*	*nsl/sup*	*nsl/inf*
actual [mm2]	0.81	0.06	0.11	0.11	0.25	0.14	0.15
predicted [mm2]	1.77	0.29	0.22	0.24	0.52	0.24	0.26
low 95.0% CI lim. [mm2]	1.35	0.15	0.15	0.17	0.42	0.18	0.22
low 99.0% CI lim. [mm2]	1.24	0.12	0.13	0.15	0.40	0.17	0.20
low 99.9% CI lim. [mm2]	1.12	0.10	0.11	0.13	0.37	0.15	0.19
actual/disc area [%]	34.6	10.7	35.2	33.2	42.7	49.6	50.0
predicted [%]	75.6	51.9	71.0	76.1	89.9	84.9	89.8
low 95.0% CI lim. [%]	57.8	27.2	47.6	52.5	73.2	64.8	74.3
low 99.0% CI lim. [%]	53.1	22.1	42.0	46.7	68.6	59.6	70.0
low 99.9% CI lim. [%]	48.1	17.4	36.3	40.7	63.6	54.0	65.3

Date: 07/Nov/2011 Signature:

Software Version: 3.1.2/2688

HRT-GPS

Heidelberg Retina Tomograph
GPS Report

HEIDELBERG ENGINEERING

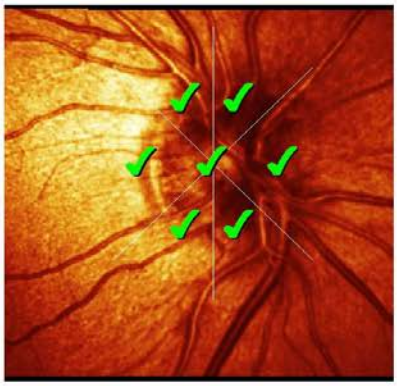
Quality: Very good (SD 13 µm)
Focus: 4.00 dpt
Operator: SN

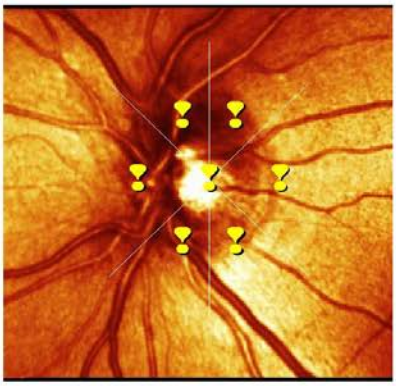
Initial Report

OD **OS**

Quality: Very good (SD 14 µm)
Focus: 2.00 dpt
Operator: SN

Glaucoma Probability Score (GPS)





global	temporal	tmp/sup	tmp/inf	nasal	nsl/sup	nsl/inf	Parameter	global	temporal	tmp/sup	tmp/inf	nasal	nsl/sup	nsl/inf
0.11	0.10	0.10	0.09	0.11	0.11	0.11	Glaucoma prob.	0.52	0.52	0.51	0.52	0.50	0.55	0.51
0.00	-0.21	-0.12	0.05	-0.14	0.14	0.03	Rim steepness	-0.44	-0.21	-0.31	-0.06	-0.74	-0.68	-0.52
0.36	0.17	0.04	0.06	0.07	0.04	0.04	Cup size [mm²]	0.52	0.14	0.08	0.09	0.11	0.10	0.07
0.50	---	---	---	---	---	---	Cup depth [mm]	0.60	---	---	---	---	---	---
-0.01	---	---	---	---	---	---	H. RNFL curv.	-0.02	---	---	---	---	---	---
-0.09	---	---	---	---	---	---	V. RNFL curv.	-0.07	---	---	---	---	---	---
✓	✓	✓	✓	✓	✓	✓		?	?	?	?	?	?	?

Glaucoma Probability Score Classification:
Within normal limits

Glaucoma Probability Score Classification:
Borderline

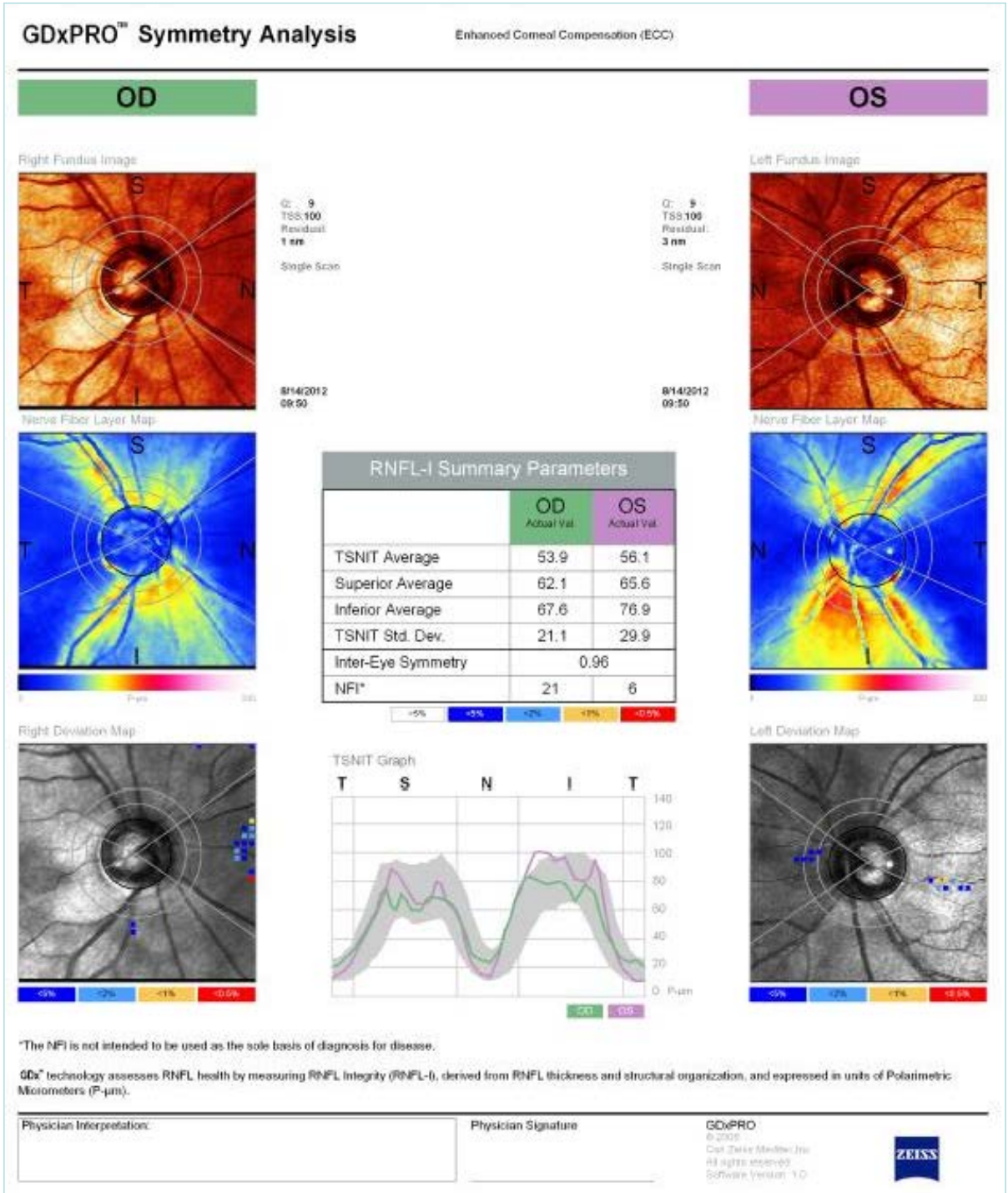
✓ Within normal limits
? Borderline
✗ Outside normal limits

Comments:

Signature:
Date: 14/07/2011

Software Version: 3.1.2/5473
www.HeidelbergEngineering.com

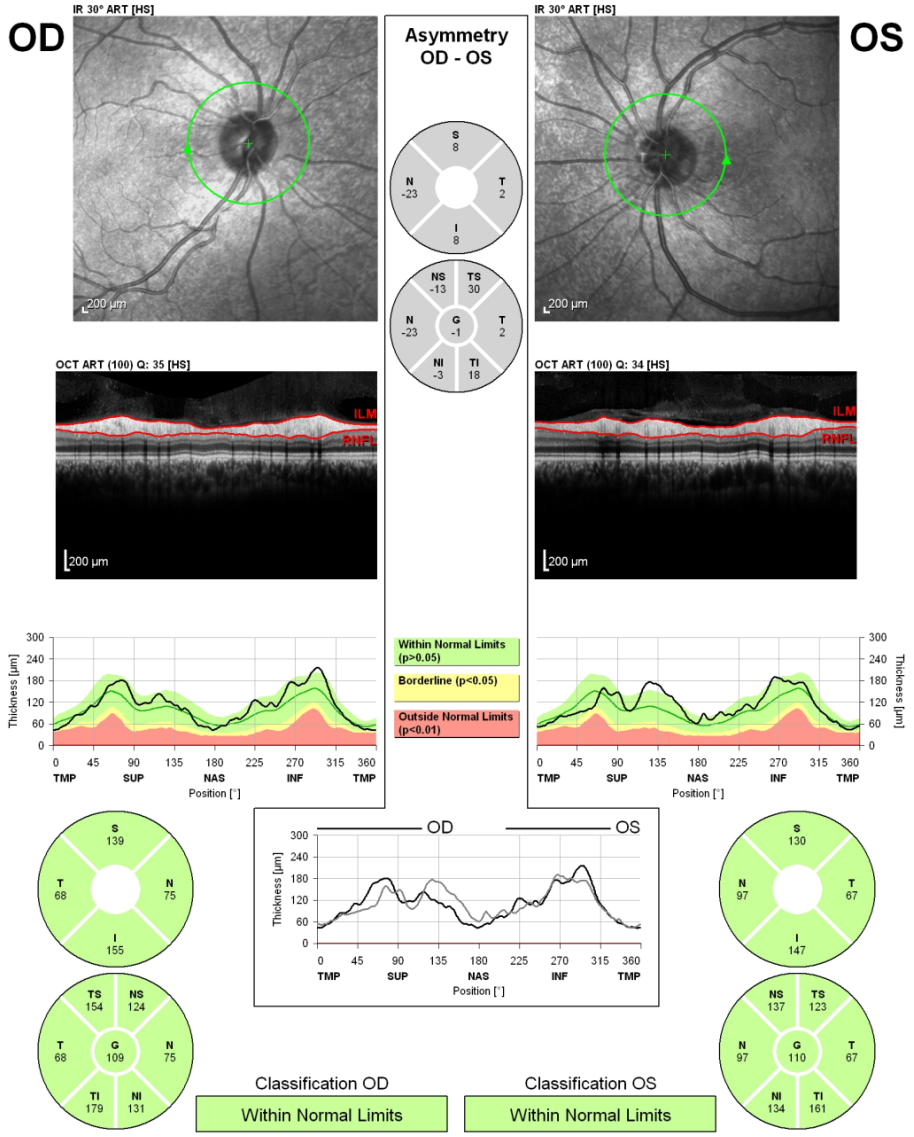
GDx


GDxPRO™ Symmetry Analysis
Enhanced Corneal Compensation (ECC)
OD
OS
Right Fundus Image
Left Fundus Image
Q: 9
TSS:100
Residual:
1 nm
Single Scan
8/14/2012
09:50
Q: 9
TSS:100
Residual:
3 nm
Single Scan
8/14/2012
09:50
Nerve Fiber Layer Map
Nerve Fiber Layer Map
RNFL-I Summary Parameters
OD
Actual Val
OS
Actual Val
TSNIT Average 53.9 56.1
Superior Average 62.1 65.6
Inferior Average 67.6 76.9
TSNIT Std. Dev. 21.1 29.9
Inter-Eye Symmetry 0.96
NFI* 21 6
Right Deviation Map
TSNIT Graph
T S N I T
Left Deviation Map
*The NFI is not intended to be used as the sole basis of diagnosis for disease.
GDx® technology assesses RNFL health by measuring RNFL Integrity (RNFL-I), derived from RNFL thickness and structural organization, and expressed in units of Polarimetric Micrometers (P-µm).
Physician Interpretation:
Physician Signature
GDxPRO
ZEISS

OCT

RNFL Single Exam Report OU with FoDi™
SPECTRALIS® Tracking Laser Tomography

HEIDELBERG ENGINEERING



Notes:

Date: Signature:

Appendix 5 Imaging standard operating procedures for the GATE study

STANDARD OPERATING PROCEDURES (SOP)

Image acquisition and storage

GATE: Glaucoma Automated Test Evaluation

Comparative study of new imaging technologies
for the diagnosis of glaucoma (HTA Reference Number: 09/22/111)

Sites:

1. Aberdeen Royal Infirmary
2. Hinchingbrooke Hospital
3. Moorfields Eye Hospital
4. St. Paul's Eye Unit, Liverpool
5. Bedford Hospital Trust

Instruments: HRT-III, GDx-PRO, Spectralis-OCT

Contents

Contact lens wear

There is currently no consensus as to whether a patient should be imaged with or without contact lenses. If a patient presents wearing contact lenses please follow local best practice.

Imaging Test order

Imaging should be performed in the random order allocated for each participant in the study. This can be found on the clinic log entry for that patient study ID on the GATE website.

Pupil dilation for imaging

Images should routinely be performed **without dilation** prior to clinician assessment and prior to visual field measurements. If pupil dilation is required to obtain an adequate quality HRT or OCT image then the GDx scan should be attempted prior to dilation (if not already performed). If an adequate quality GDx image is still not obtained prior to dilation the GDx scan should be repeated after pupil dilation. Whenever possible the random allocated test order should be used.

General indications for pupil dilation are media opacities and/or small pupils. However, the scan should always be attempted first to determine whether images are acceptable or if dilation is necessary.

Criteria for dilating the pupil are as follows:

- Unable to 'lock-on' to the pupil and save a scan
- Acceptability of best saved image is below requirements stated in SOP for that imaging technique

Acceptability criteria for each imaging technique are clearly detailed in the text for each technique below. Once an acceptable image has been obtained no further images should be acquired.

Acceptable quality criteria for imaging (summary)

HRT	Mean standard deviation ≤30, Image quality score: Good, Very Good, or Excellent
GDx	Q≥8
OCT	Q>15

Heidelberg Retina Tomograph (HRT-III)

Acquiring the image

Before imaging a subject on the HRT you should record their refraction (focimetry or auto-refraction. The focimetry/autorefraction is useful to guide the setting of the scan focus before image acquisition, but is not required to be input in the software.

Patients should not be imaged with their contact lenses in.

1. Ensure that IOP measurements (and other contact exams like gonioscopy) are done ***after*** HRT imaging.
2. Explain examination (method, time and requirements) to the patient.
3. Disinfect chin- and forehead supporting-stand.
4. Check optics for dirt or smudges, clean if required, with lint/oil-free lens paper moistened with a drop or two of photography quality lens cleaner.
5. Enter new patient details:

    a) Click on the new patient icon on the HEYEX tool bar to enter the subject's details and the operator initials. Enter the corneal curvature as an average of the two axes (i.e. 7.6 x 7.8 = 7.7) and enter the refractive error.

    The patient details need to be recorded as follows:
    - **Last Name:** 'GATE'
    - **First Name:** <site> e.g. 'Aberdeen'
    - **Title:** leave blank
    - **Date of birth:** enter patient date of birth
    - **Sex:** enter patient gender
    - **Patient ID:** enter < Participant study number>
    - **Ancestry:** enter the patient's ethnicity

6. Ensure that the table and the headrest are at the correct height for the subject. Adjust the chinrest height so that the patient's eyes are at the same level as the red canthus marks on the headrest posts. When the subject's details are entered the laser will activate and image acquisition can begin. As a starting point the focussing dial at the front of the HRT should be set to the subject's refraction.

7. Check that the subject is comfortable, and the scanning head is correctly aligned using the black adjustment handles on the HRT. The imaging distance between the HRT objective lens and the cornea is 1.5 cm; this can be judged by focussing the scanning beam on the iris (moving the scanning head towards or away from the eye). When the laser beam is focussed (sharp outline) on the iris, move the scanner sideways so the beam enters the pupil. All of the red light emitted from the HRT should be going straight into the pupil with little or none visible on the iris.

8. If cylinder correction exceeds ±0.75 diopter, place supplemental cylinder lens in front of the objective (image acquisition) lens. Note that the HRT will display a flashing alert on the refractive error correction dialog box if cylinder correction is recommended:

   a. click on the check box to display the recommended cylinder lens strength to be used. The lens should be oriented according to the axis provided by the autorefractometer.

   b. orientation of the cylinder may be adjusted manually during imaging in order to achieve the highest image quality on the screen (see point 9)

9. The HRT has an internal fixation point, a yellowy/green light which, when fixated by subject, should bring the optic disk into view on the screen. Once the scan is activated, the fixation point will appear on the subject's left for the right eye and on their right for the left eye (i.e. always towards the nose). Explain the fixation target (green light outside the red "carpet"). Once the subject is fixating, check the alignment of the laser in the pupil and make small adjustments to the focusing to optimise the image quality. The operator may also manually adjust the cylinder lens axis, if present, to achieve the best image quality.

   a. Images that are dark, have vignetting of the image corners or are grainy can be improved by dilation. If a good quality image cannot be obtained and dilation is required, delay the image acquisition until after clinical assessment as described in the flow chart in the GATE study procedures manual.

   b. In some cases where the internal fixation light cannot be seen by the patient, the external fixation device must be used. To use this, the opposite eye fixates on the green light which is manually positioned by the operator to display the optic nerve head in the centre of the screen.

10. Instruct the subject to blink as much as possible <u>before</u> you attempt to acquire a scan to avoid imaging a "dry eye" (drying of the corneal surface). The scan

duration is approx 6-8 seconds, during which time the subject should not blink and must maintain fixation. When ready, an image is acquired by pressing either the foot pedal or the grey button on the rear of the scan head.

11. Optimise image quality if necessary (artificial tears, elevate upper eyelids, re-adjust camera).

    a. Detector sensitivity should be as low as possible (<80)

    b. The image quality bar should be green for a good quality image and above 70%

12. The screen will display the progress of the scan; the HRT will take at least three scans of the optic nerve in one session.

    a. Monitor the progress of the scan to ensure that the subject's position and fixation is constant.

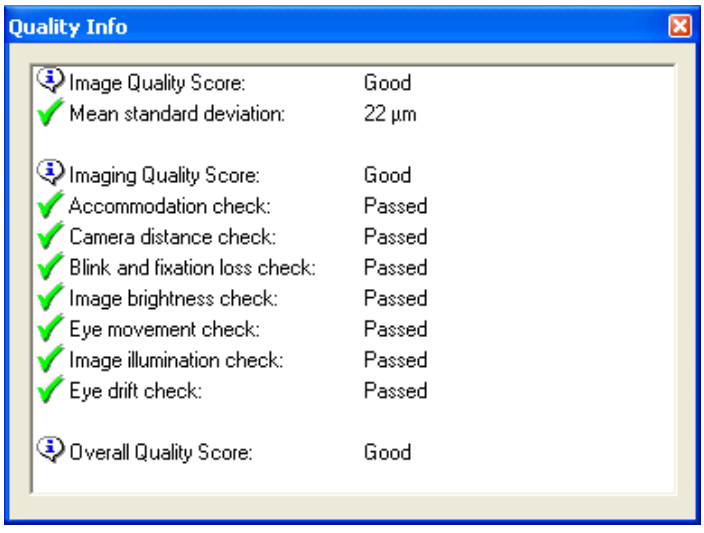
    b. The image series may be reviewed as a movie immediately after acquisition. HRT software can compensate for some movement but any scans containing large eye movements or blinks will have to be discarded and repeated.

    c. If you are satisfied with the scan select "save". Repeat to acquire another scan if image quality is not adequate.

    d. Move the machine over to image the fellow eye, the HRT software will recognise which eye you are scanning.

Checking image quality

Images acquired using the HRT III software allow the user to check the quality of the image by clicking on the **QC** icon in the right hand corner of the image:



a. Ensure that all elements are ticked. The overall quality score is given as one of Very Poor, Poor, Acceptable**, Good, Very Good, Excellent.** Ensure the overall quality score is Good or higher.

b. Check the **standard deviation** value displayed at the top of the topography. A value of **30 or below** is considered good image quality.

c. Repeat the image acquisition if required to obtain a good image quality.

d. If an acceptable image cannot be obtained after repeated attempts then the image acquisition should be attempted again after pupil dilation (please refer to study procedures document flow chart to ensure the order of clinical/imaging tests and visual field testing is correct)

Computing the topography

1. Once both eyes have been imaged, the topographies need to be computed.
   a. Exit from of the acquisition mode by clicking the "X" in the right hand corner of the acquisition window and you will be prompted to process the scans that have just been taken. When the scans are processed double click on the image to bring up the resulting topographical data in the examination results window

Drawing a contour line

1. Using the left mouse button select at least 3 points on the optic nerve rim to create a contour line and reposition to ensure the optic nerve rim is accurately located. Use the 3D viewer button to review your selection if required.
2. When you have located the optic nerve rim select 'Contour' then 'Accept Contour'
3. The contour and segment lines will appear.

HRT Report Printout:

Three printouts are required for the HRT scan,

1. Moorfields Regression Analysis (MRA) of right eye,
2. MRA of left eye and
3. GPS report

A hardcopy of each report should be filed in the study file for each participant.

An electronic copy of each report in **JPEG** format should also be saved to a memory stick and uploaded to the GATE study website.

The naming convention for filenames of any saved reports should be followed:

- MRA Right eye report: 'GATE<studyID>MRAright.jpg'
- MRA Left eye report: 'GATE<studyID>MRAleft.jpg'
- GPS report: 'GATE<studyID>GPS.jpg'

Printing a Moorfields Regression Analysis(MRA) Report

- Click the Moorfields classification tab
- Select 'Print'
- Select 'Examination report'
- Ensure the 'Moorfields report' is selected in the reports window
- Select 'preview'
- The Moorfields report for that eye will appear
- Select 'Save as'
- Enter the filename as 'GATE<studyID>MRAright.jpg' for right eye (or GATE<studyID>MRAleft.jpg for left eye)
- Select 'Save' to save the jpeg of the report
- Select 'print' to print a hardcopy report for the file
- Select other eye from Heidelberg Eye explorer window
- Repeat from start to print/save report from the other eye.

Printing a GPS report

- To print report select 'GPS classification' tab
- Select 'Print'
- Select 'Examination report'

To save as a jpeg file to upload to the GATE website

- Ensure GPS report is highlighted in the report window
- If more than one image is stored for the other eye, select the corresponding eye image for the GPS report
- Select Preview
- The GPS report will appear on the print preview screen
- Select 'Save as'
- Enter the filename as 'GATE<studyID>GPS.jpg'
- Select save to save the jpeg of the report

To print a hardcopy report

- Select 'Print'

HRT Data Export:

HRT imaging data should be exported on the same day every week. Export all the images since the last export.

1) From the main database screen select the patient you wish to export.
2) When the patient details appear on the right hand side of the screen double click the patient name.
3) Right click on the exam you wish to export and select the export option.
4) Select yes to export the 3D image series. Selecting this option could result in a prompt to retrieve the raw image data for that exam.
5) Select a folder to export using the browse option.
6) The file will export and appear as an *.E2E file in the selected export folder.
7) Save with filename ‘GATE<studyID>rawHRTOS.e2e’ for left eye, and ‘GATE<studyID>rawHRTOD.e2e for right eye

GDx ECC/GDx Pro

Acquiring the image

1. Patients should not be imaged wearing their contact lenses. Ensure that IOP measurements and other contact exams like gonioscopy are performed after GDx imaging

2. With the GDx VCC and peripherals properly connected, the Access card inserted and the optics unlocked turn the machine on using the power switch on the side.

3. Disinfect face rest.

4. When the warm-up test is complete, the logo screen will be displayed. Select “new patient”.

5. The patient details need to be recorded as follows:
   - **Patient ID** - Use <GATE Participant study number>
   - **Last name –** ‘GATE’
   - **First name** – <site>e.g. Aberdeen
   - **Middle name** – leave blank
   - **DOB** – enter patient date of birth
   - **Doctor** –use any identifier usually used or leave blank

Press √ button to continue.

6. Input ancestry and gender information as prompted. Press √ button to continue.

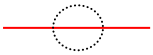
7. Check patient information for accuracy. If changes are needed press edit.

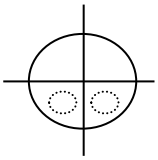
8. Select “full exam”.

9. Refraction input is needed for focusing purposes. Select Refraction. Press the **auto-refraction** button.

10. Position the subject in front of the GDx with the face placed comfortably against the face rest. For optimal positioning, ask the patient to place their brow bones on top of the upper rest “like wearing a mask”. Ask the subject to gaze at the blinking fixation target in the red field– located to the left hand side for the right eye and vice versa. The exam begins with the right eye as the default position.

11.

- Move the joystick , by pulling forwards and backwards, to vertically align the white focus dot on the horizontal red line 

- Centre the pupil in the target by moving the joystick up/down/ left/right and ensure that the 2 white dots are located in the bottom 2 quadrants of the target 

- Once aligned, ask the patient not to blink and press the **image acquisition button** on top of the joystick to scan the eye
- Once the refraction data is displayed on the top lhs select **image acquisition button** again to acquire refraction data from the other eye
- Click **image acquisition button** a 3rd time to display the ‘modify ellipse – measuring cornea’ screen

12. Following a cornea measurement, the “Modify Ellipse – Measuring Cornea” screen appears. The macular ellipse should be centred directly over the macula “bowtie”-pattern.

13. To change the macular ellipse position, use the arrow buttons. Do not change the size of the macular ellipse.

14. If the macula “bowtie” is not well defined, press the “Irregular Pattern” button to use an alternative cornea calculation based on the macula area within the dotted square which does not require macular ellipse placement (press the “Macular Ellipse” button to re-enable the macular ellipse placement options).

15. Press √ button to go to the “Modify Ellipse – Measuring Cornea” screen for the next eye. Optimise the placement of the cornea measurement ellipse in the same manner as for the first eye. Press √ button again when complete.

16. When ellipse modification is completed, the system displays the “Image Check-Measuring cornea” screen. The scan **quality score should be 8 and above**. If the scan quality is less than 8 then retake the image for that eye.

17. If image quality is acceptable then select “Accept”. The acquisition screen will now appear. If after repeating the GDx scan the image quality is still not acceptable then the image acquisition should be attempted again after pupil dilation (please refer to GATE study procedures flow chart to ensure the order of clinical/imaging tests is correct)

18. Move the joystick to vertically align the white focus dot on the horizontal red line and centre the pupil in the reticule. Once aligned, press the image acquisition button on top of the joystick to scan the eye. After the first image is captured the system will automatically move to the other eye.

19. Repeat step 18 for the left eye.

20. At the “Image Check” screen verify that the ellipses for both eyes are the correct size, shape and centred on the Optic Nerve Head (ONH). You can both change the ellipse diameter and shape using the arrow keys. (Note: While it is helpful to align the ellipse with the ONH margin, accurate centration is more important than perfect ellipse size).

21. When ellipse modification is completed, verify that the scans quality scores are 8 or above. If quality is less than 8, re-scan the patient. Proper ONH placement can influence image quality scores. Verify that placement is correct before deciding to retake an image.

22. If image quality is acceptable then select “Accept”.
    If not acceptable then the image acquisition should be attempted again after the clinician has dilated pupils for their routine clinical assessment (please refer to GATE study procedures flow chart to ensure the order of clinical/imaging tests is correct)

23. Press “print” or “save only” button.

GDx Printout

A printout of the GDx ECC report is required for the study file. Ensure that ECC is reported in the middle text box (white).

A hardcopy of each report should be filed in the study file for each participant.

A scanned electronic copy of each report in **JPEG** format should be uploaded to the GATE study website.
The naming convention for filenames of any saved reports should be followed:

- GDx ECC report: 'GATE<studyID>GDX.jpg'

After printing out the report you should select 'Save' to save the file for this patient

GDx Data Export

Data export is done on a floppy disk.

1. Select "existing patient".
2. Enter the patient ID number in the "Patient ID".
3. Press the "review" button
4. Use "Previous" and "Next" buttons to move through the list
5. Then using the "Select/Deselect" button highlight the chosen exam
6. Repeat 2 and 3 to select more then one exam
7. Press "review" button and then "export" button
8. Choose 'Export raw data'
9. You will be presented with a folder: My Computer. Press the tab key to get into the folder.
10. Using the arrow keys highlight the Floppy A Folder. Then press enter.
11. Name the folder (although the software will automatically save the data with Patient name and ID).
12. Press "√" button or the "enter" key.
13. You will be presented with the message: "Exporting Data" and when finished with the message "Export complete".

Optical Coherence Tomography Using: Heidelberg Spectralis OCT

Patients should not wear contact lenses

Acquiring the image

1. Explain examination (method and requirements) to the patient
2. Ensure that IOP measurements (and other contact exams like gonioscopy) are done **after** OCT imaging
3. Disinfect chin-and forehead-supporting stand
4. Check optics for dirt or smudges, clean if required/with lint/oil free lens paper moistened with a drop or two of photography quality lens cleaner
5. Ensure that the table and headrest are the correct height for the subject. Adjust the chinrest height so that the patient's eyes are at the same level as the red canthus marks on the headrest posts.

5. Create a new patient record by clicking on the ***New Patient*** button.

In the ***Patient File*** window you should enter the following patient details

- **Patient ID** - Use < Participant study number>
- **Patient Name** (Surname= 'GATE', Forename=<site> e.g. 'Aberdeen')
- **DOB** – enter patient date of birth
- Enter **Gender** and **Ethnicity** information

Examination Data Window

The ***Examination Data*** dialog opens before each exam, but can also be opened at any later stage using the ***Examination*** button in the patient file.
The respective ***Device Type*** for the examination must be selected from the dropdown menu; all other data is optional.

1. Select Spectralis OCT
2. Enter operator initials
3. Enter Study name as GATE

Eye Data Window

This window enables the entry of the eye parameters for both eyes. **Please Note**: Do NOT enter any data into this window.

1. Wait for the Laser On/Off button on the Control Panel to turn from Red to Yellow.
2. Press the Yellow On/Off button on the Control Panel to activate the Laser/OCT. Make sure that the OCT button is selected. (Note - On the Control Panel, Inactive/unselected buttons are Red; Active/Selected buttons are Blue).
3. Select the IR + OCT button.
4. Make sure that the Volume button is selected.
6. Field button should be at 30 degrees.
7. IR Intensity button will default to 100% but should be adjusted for patient media, typically 50% - 75%.
8. Always activate the ART Mean function when performing an OCT-Scan
9. Select **RNFL** preset on the monitor screen
10. Ensure HR for high resolution imaging is selected (not HS – high speed)
11. The OCT has a blue internal fixation point which when fixated by the subject should bring the optic disc into view on the screen
    a) slowly bring the camera towards the patient's eye,
    b) encourage the patient to blink just before a scan, since maintaining a good tear film is important for OCT image quality. In cases where the patient suffers from dry eye, or when the cornea cannot be kept moist enough by blinking alone, artificial tears may be used.
    c) Using the joystick (up, down, right or left) move the camera to the center of the pupil and adjust the distance between the objective and the examined eye to approx. 14 mm between the front edge of the objective and the cornea.
    d) Use the OCT Acquisition Window on the monitor to align the camera with the Optic disc Image on the left side of the window.
    e) Fine tune brightness and sharpness of the image using the focus knob. The optimum camera position is reached when no dark corners and overexposed areas are visible.

11. The bar above the OCT image will appear red if the OCT image touches the upper border. Move the camera further away from the patient if the OCT image is shown inverted. If the OCT image is tilted in a horizontal direction, move the camera slightly left/right (if capturing a horizontal scan) or up/down (if capturing a vertical scan). (Note in patients with moderate myopia, the scan can be tilted).

12 . The blue ***Quality bar*** in the lower part of the image indicates the signal strength. The quality score range is **0 (no image) to 40 (excellent quality)**. **Acceptable quality is**

>15. If the score is 15 or less, the quality bar turns red. If an acceptable quality image cannot be obtained, imaging should be repeated after pupil dilation (see flowchart in study procedures manual to determine test order)

13. To achieve optimum image quality, position the OCT image in the upper half of the ***Acquisition*** window. Using the joystick, move the camera slightly up/down and sideways until the optic disc and OCT image appear brightest and most evenly illuminated.

14. To acquire images, press the foot switch, the ***Acquire*** button on the control panel or the central button on the joystick. After acquiring images, save them using the ***Save images*** option in the top left corner of the ***Acquisition*** window. To end the acquisition session, exit the ***Acquisition*** window. The camera will automatically turn off.

OCT RNFL Basic Report OU Printout

- A hardcopy RNFL Basic Report OU should be filed in the study file for each participant.
- Add an image from each eye to the lightbox
- Select both images in the lightbox then select Print from the context menu
- The 'print spectralis report' window will appear
- Select the RNFL Basic Report option
- Select Preview
- Select Save to save a JPEG format then select Print to printout a hardcopy
- An electronic copy of each report in **JPEG** format should also be saved to a memory stick and uploaded to the GATE study website. The naming convention for filenames of any saved reports should be followed: OCT report: 'GATE<studyID>OCT.jpg'

OCT Data Export

OCT imagining data should be exported on the same day every week. Export all the images since the last export.

1) From the main database screen select the patient you wish to export
2) To export images and other data in an examination as an E2E file, select the desired thumbnail image(s) from the ***Patient File*** window, and select the item ***Export ► asE2E*** from the ***Context Menu*** in the ***Patient File***.

3) Save with filename ‘GATE<studyID>rawHRTOS.e2e’ for left eye, and ‘GATE<studyID>rawHRTOD.e2e for right eye
4) The *Batch* ► *Export E2E* feature in the *Database* window enables export of multiple patient records at once.

Appendix 6 Further assessment of threshold effects under diagnosis analysis using individual parameter from the imaging tests

As for default analysis, abnormal imaging test results were those classified as 'outside normal limits' and the corresponding reference standard definition of disease was a diagnosis of glaucoma of the worse eye. Only participants with an imaging test output with an overall classification which met the manufacturer quality cut-off were included in the analysis.

The HRT-MRA parameters for which a ROC curve was produced and the AUC calculated were the Global, Temporal, Temporal superior, Temporal inferior, Nasal, Nasal superior and Nasal inferior areas respectively. For HRT-GPS and OCT the respective probabilities and the retinal fiber nerve layer thickness values were used for the same segments of the eye. For GDx, the TSNIT parameters (NFI, TSNIT Average, Superior average, Inferior average, TSNIT standard deviation were used).

The corresponding ROC curves are shown in Figures 6.1 to 6.4 below with the corresponding AUC with 95% CIs in Table 6.1. From visually assessment it can be seen that the OCT and GDx curves differed the most between parameters with the HRT tests, MRA and particularly GPS showing less variation in the curve shape between parameter. The point estimates for the AUC differed by only 0.02 for GPS compared to GDx for 0.1 and 0.13 for OCT.

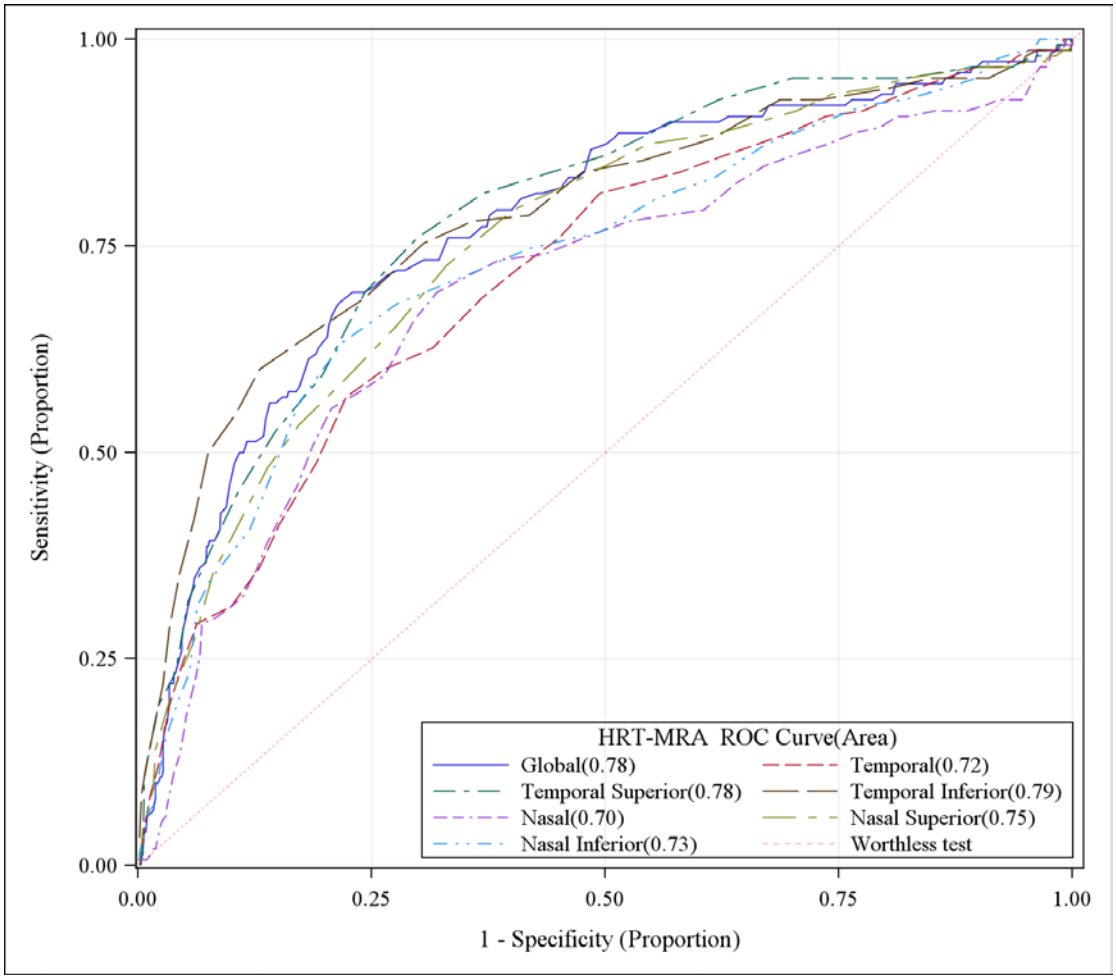


Figure 6.1 ROC curve for HRT-MRA parameters

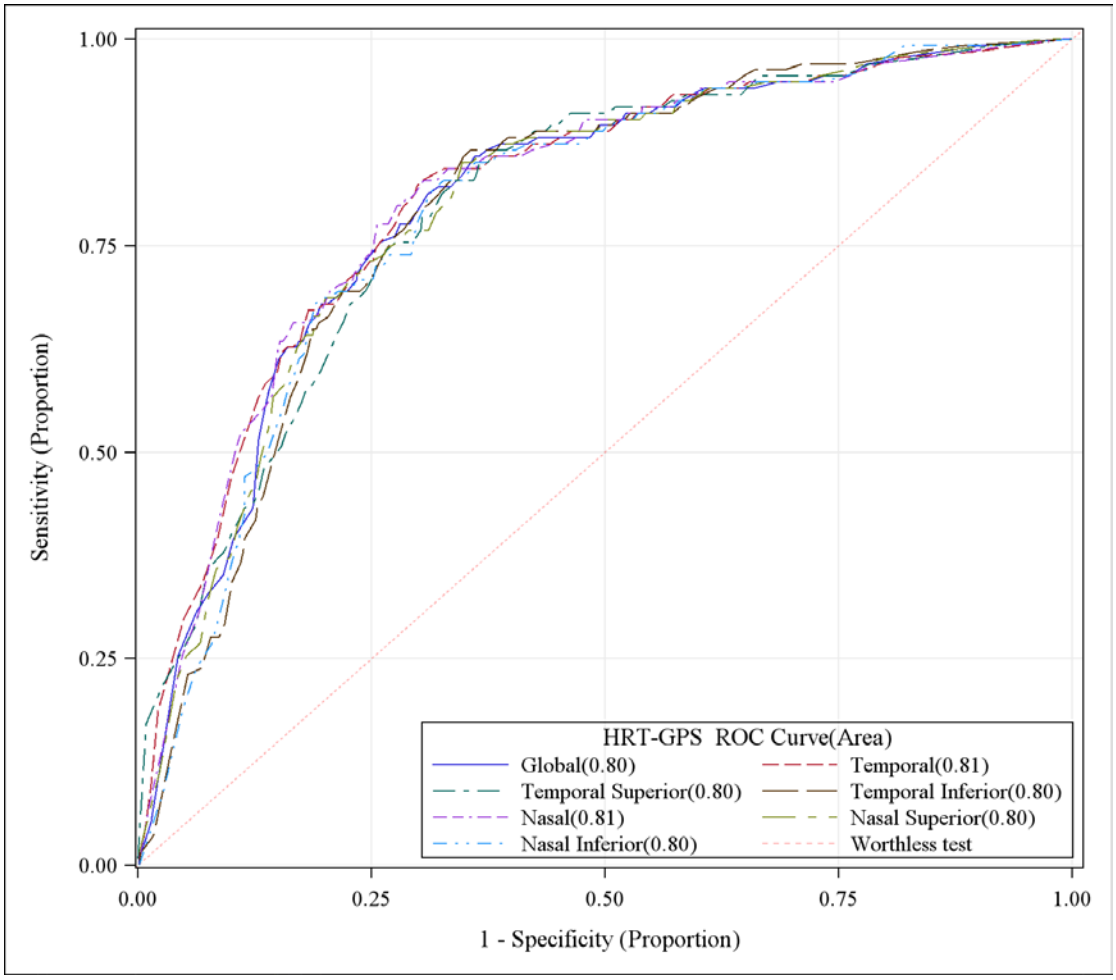


Figure 6.2 ROC curve for HRT-GPS parameters

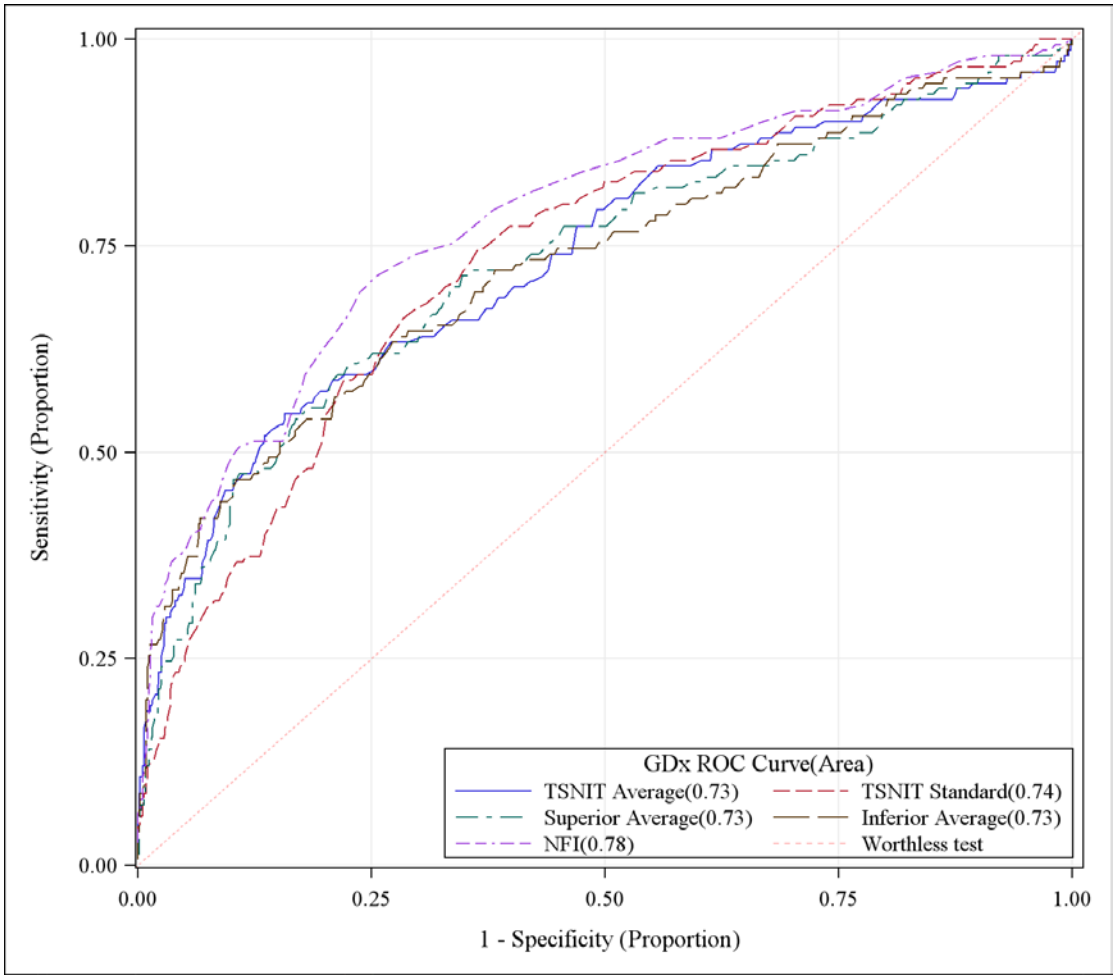


Figure 6.3 ROC curve for GDx parameters

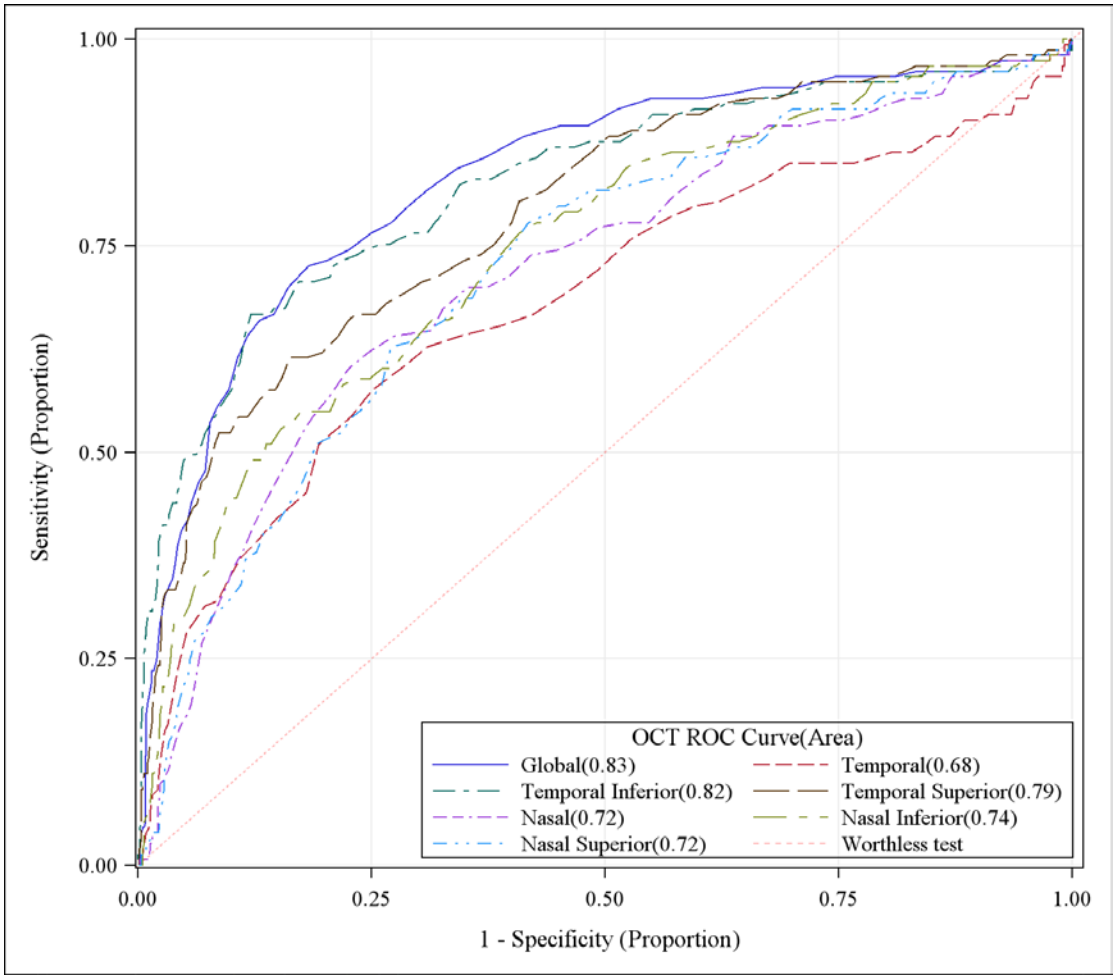


Figure 6.4 ROC curve for OCT parameters

Table 6.1 Area under the ROC curve using imaging test parameters for a diagnosis of glaucoma in the worse eye

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Global area	0.78	0.73	0.82
	Temporal area	0.72	0.67	0.76
	Temporal superior area	0.78	0.74	0.83
	Temporal inferior area	0.79	0.74	0.83
	Nasal	0.70	0.65	0.75
	Nasal superior area	0.75	0.71	0.80
	Nasal inferior area	0.73	0.69	0.78
HRT-GPS	Global probability	0.80	0.77	0.84
	Temporal probability	0.81	0.77	0.85
	Temporal superior probability	0.80	0.76	0.84
	Temporal inferior probability	0.80	0.76	0.83
	Nasal probability	0.81	0.77	0.85
	Nasal superior probability	0.80	0.76	0.84
	Nasal inferior probability	0.79	0.76	0.83
GDx	NFI	0.78	0.74	0.83
	TSNIT average	0.73	0.69	0.78
	TSNIT Standard deviation	0.74	0.69	0.78
	Superior average	0.73	0.68	0.78
	Inferior average	0.73	0.68	0.78
OCT	Global thickness	0.83	0.79	0.87
	Temporal thickness	0.68	0.63	0.73
	Temporal superior thickness	0.79	0.75	0.83

Test	Parameter	Value	95% CI	
			Lower	Upper
	Temporal inferior thickness	0.82	0.78	0.86
	Nasal thickness	0.72	0.68	0.77
	Nasal superior thickness	0.72	0.68	0.77
	Nasal inferior thickness	0.74	0.70	0.79

Appendix 7 Additional triage analysis to inform the health economic model

Overview

An additional set of two statistical analyses (*Triage sensitivity analyses 9 and 10*) were carried out to specifically inform the economic modelling for GATE. These were set up to mirror the model structure in terms of population (i.e. with the simplification of ignoring the presence of non-glaucoma related co-morbidities). The first additional analysis used a reference standard definition of disease of glaucoma, glaucoma suspects, OHT and PAC; the second analysis used diagnosis of glaucoma alone as the reference standard (*see Table A.1*). The test was a composite as previously described in Chapter 2 and 5 of the respective imaging test result, IOP and visual acuity measurements (referred to throughout this appendix by the name of imaging test used within the composite test e.g. either HRT-MRA, HRT-GPS, GDx or OCT). Where a classification was not provided by the imaging test, the patient was defined as a "For referral". For the first analysis, borderline imaging results were also classified as "For referral" Whereas for the other analysis they were classified "Not for referral". Triage sensitivity analyses 9 and 10, respectively represent the analyses used to populate the diagnostic accuracy results of the base-case and the sensitivity analysis scenarios (*see Chapter 6 for further details*). Sub-group sensitivity and specificity values were calculated for each diagnosis separately (e.g. glaucoma, "at risk of glaucoma" and neither groups) breaking down the respective performance of the triage test to provide estimates for the economic model.

Table A.1 Additional analyses carried out to inform the health economic model

Analysis	Reference standard definition of disease	Test 'For referral' definition	Handling of 'no result' categories	Figure no.	Table no.
Triage sensitivity analysis 9	Glaucoma, OHT, PAC and glaucoma suspects.	Imaging (outside normal limits or borderline) or IOP>21mmHg or VA6/12 or poorer	A-D For referral E excluded	A.1	A.2
Triage sensitivity analysis 10	Glaucoma	Imaging (outside normal limits)	A-D For referral E excluded	A.2	A.3
No result categories *A: test performed and imaging report produced but quality is lower than manufacturer quality cut-off* *B: test performed and imaging report produced but no overall classification generated by machine* *C: test performed but there was a clear imaging artefact on the report* *D: test attempted but no imaging could be acquired from the patient's eyes – no report generated* *E: missing imaging output (due to study related or data collection issues)*					

Diagnostic performance of the triage tests

The diagnostic accuracy results of the two analyses are given in the following two sections.

Triage sensitivity analysis 9

The flow of study participants according to triage sensitivity analysis 10 is shown in Figure A.1 with respective numbers of referral, not for referral and no results cases by triage test. The diagnostic performance for the four tests is given in Table A.2. Results showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (HRT-GPS was only very slightly lower) but also the second lowest specificity (HRT-GPS had the lowest), GDx had the lowest sensitivity but the highest specificity, and OCT provided intermediate results. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 4.29 for GDx to 16.83 for HRT-MRA.

From this analysis, the sensitivity for participants with glaucoma was calculated as 99, 99, 88 and 97% for HRT-MRA, HRT-GPS, GDx and OCT, respectively; similarly the sensitivity for participants at risk of glaucoma was calculated as 97, 97, 77 and 87% respectively and the specificity for participants classified as normal (not glaucoma or ‘at risk of glaucoma’) was 30, 28, 51 and 35% respectively for HRT-MRA, HRT-GPS, GDx and OCT.

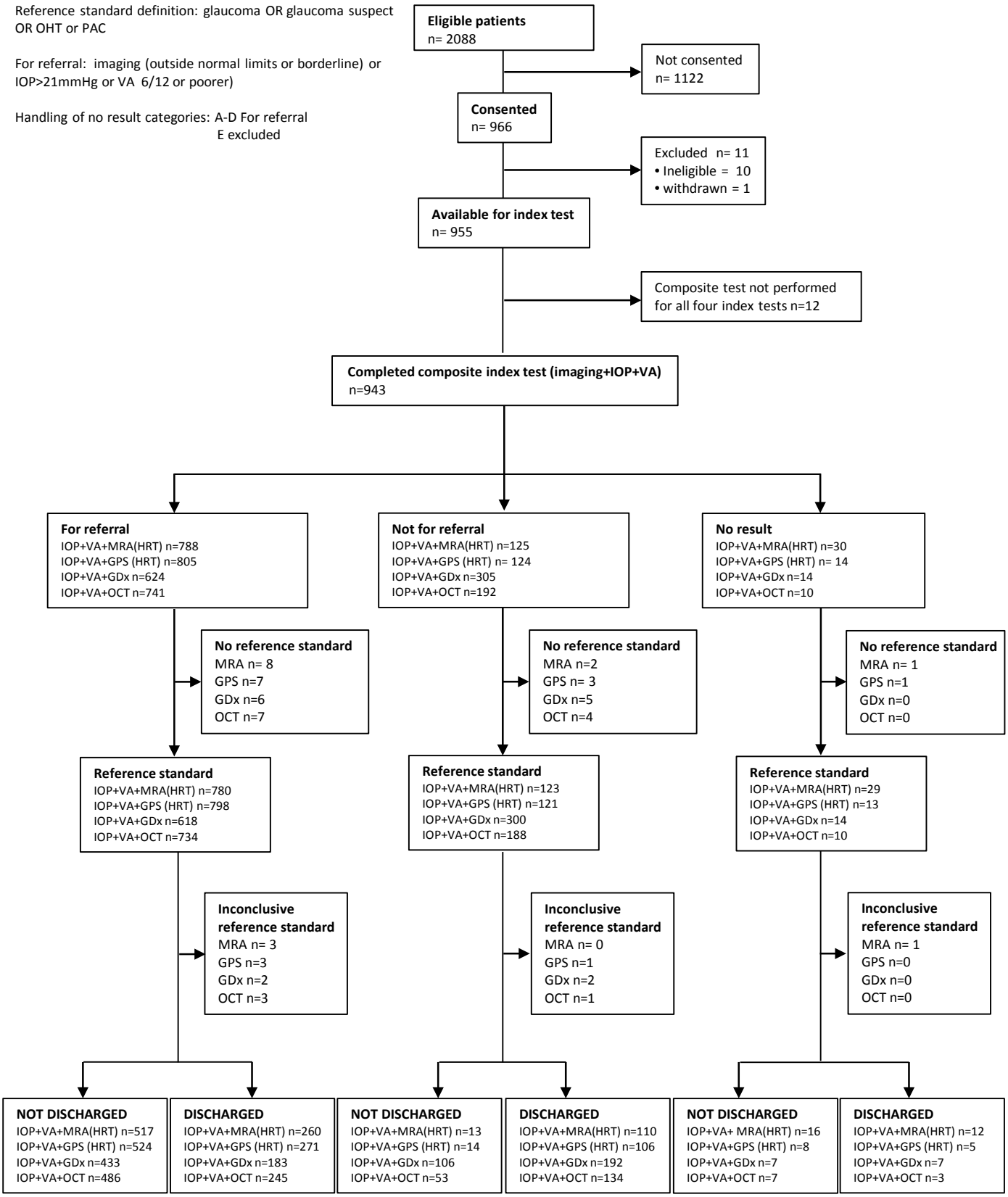


Figure A.1 Flow diagram Triage sensitivity analysis: 9

Table A.2 Triage Sensitivity analysis 9

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	97.5	95.8	98.7
	Specificity - %	29.7	25.1	34.7
	Positive Likelihood Ratio	1.39	1.30	1.49
	Negative Likelihood Ratio	0.08	0.05	0.14
	DOR	16.83	9.29	30.47
HRT-GPS	Sensitivity - %	97.4	95.7	98.6
	Specificity - %	28.1	23.6	32.9
	Positive Likelihood Ratio	1.35	1.27	1.45
	Negative Likelihood Ratio	0.09	0.05	0.16
	DOR	14.64	8.23	26.05
GDx	Sensitivity - %	80.3	76.7	83.6
	Specificity - %	51.2	46.0	56.4
	Positive Likelihood Ratio	1.65	1.47	1.84
	Negative Likelihood Ratio	0.38	0.32	0.47
	DOR	4.29	3.2	5.75
OCT	Sensitivity - %	90.2	87.3	92.5
	Specificity - %	35.4	30.5	40.4
	Positive Likelihood Ratio	1.39	1.29	1.51
	Negative Likelihood Ratio	0.38	0.21	0.37
	DOR	5.02	3.52	7.14

Triage sensitivity analysis 10

The flow of study participants according to triage sensitivity analysis 10 is shown in Figure A.2 with respective numbers of referral, not for referral and no results cases by triage test. The diagnostic performance for the four tests is given in Table A.3. Results generally showed a trade-off between detection of patients who need to be referred and discharging those who do not need to be referred: HRT-MRA had the highest sensitivity (HRT-GPS was only very slightly lower) but also the second lowest specificity (HRT-GPS had the lowest), GDx had the lowest sensitivity but the highest specificity, and OCT provided intermediate results. Likelihood ratios (and 95% CI) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 triage tests (CIs did not contain 1.0). DORs ranged from 5.11 for GDx to 12.83 for HRT-MRA.

From this analysis, the sensitivity for participants with glaucoma was 93, 89, 49 and 83% for HRT-MRA, HRT-GPS, GDx and OCT respectively; the sensitivity for participants at risk of glaucoma was calculated as 61, 59, 17 and 36% respectively and the specificity for participants in the normal health state (without glaucoma or at risk of glaucoma) was calculated as 60, 61, 85 and 72% respectively.

Reference standard definition: glaucoma

For referral: imaging (outside normal limits or borderline)

Handling of no result categories: A-D For referral
E excluded

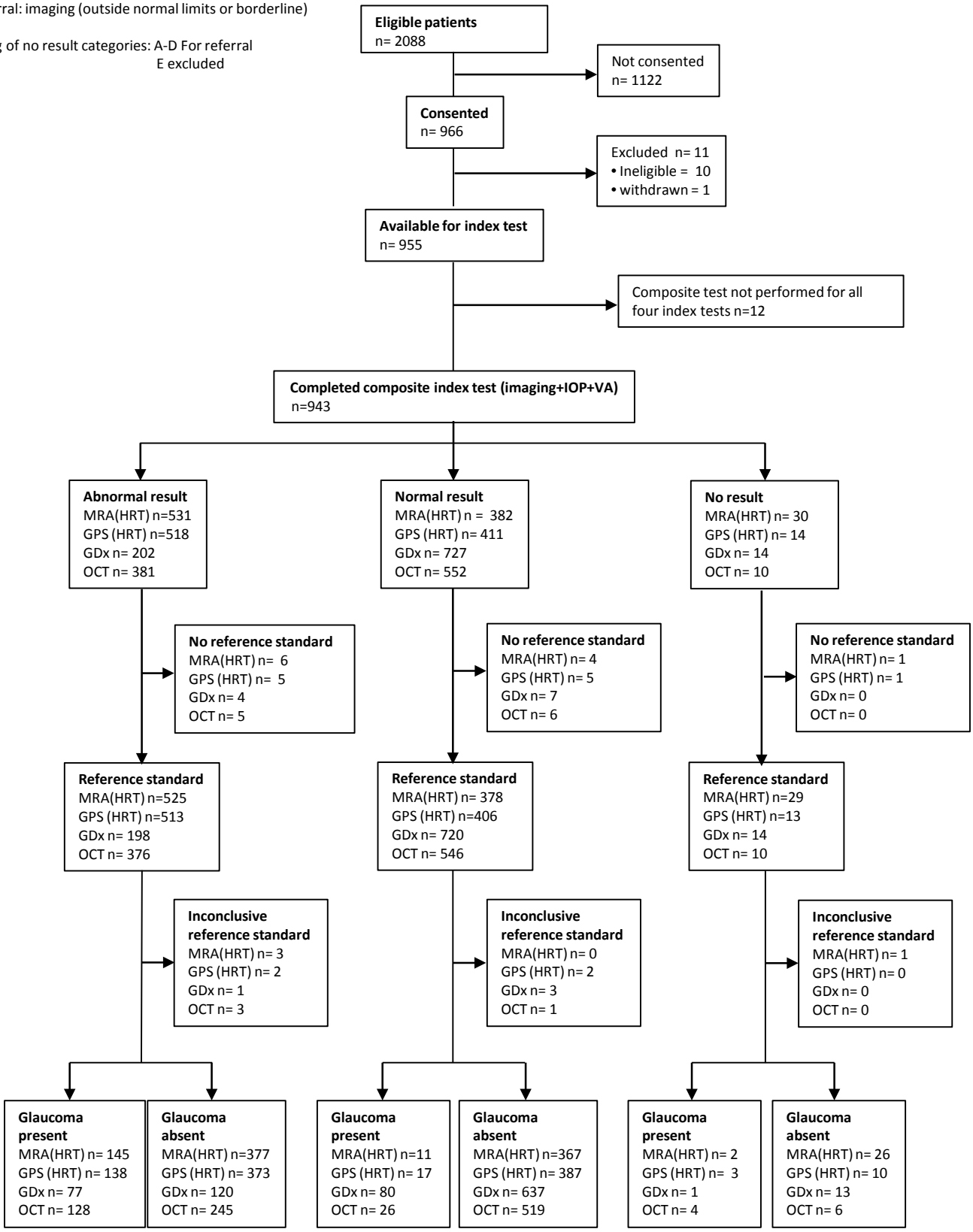


Figure A.2 Flow diagram Triage analysis sensitivity 10

Table A.3 Triage sensitivity analysis 10

Test	Parameter	Value	95% CI	
			Lower	Upper
HRT-MRA	Sensitivity - %	92.9	87.7	96.4
	Specificity - %	49.3	45.7	53.0
	Positive Likelihood Ratio	1.83	1.69	21.99
	Negative Likelihood Ratio	0.14	0.08	0.25
	DOR	12.83	6.84	24.08
HRT-GPS	Sensitivity - %	89.0	83.0	93.5
	Specificity - %	50.9	47.3	54.5
	Positive Likelihood Ratio	1.81	1.66	1.99
	Negative Likelihood Ratio	0.22	0.14	0.3
	DOR	8.42	4.99	14.88
GDx	Sensitivity - %	49.0	41.0	57.1
	Specificity - %	84.1	81.3	86.7
	Positive Likelihood Ratio	3.09	2.46	3.89
	Negative Likelihood Ratio	0.61	0.52	0.71
	DOR	5.11	3.53	7.39
OCT	Sensitivity - %	83.1	76.2	88.7
	Specificity - %	67.9	64.5	71.2
	Positive Likelihood Ratio	2.59	2.29	2.94
	Negative Likelihood Ratio	0.25	0.17	0.35
	DOR	10.43	6.66	16.33

Appendix 8 Interim life table

Age x	qx^a	*Age x*	qx^a
0	0.00501	51	0.00388
1	0.00034	52	0.00424
2	0.00019	53	0.00464
3	0.00014	54	0.00502
4	0.00011	55	0.00572
5	0.00012	56	0.00625
6	0.00011	57	0.00675
7	0.00009	58	0.00731
8	0.00011	59	0.00796
9	0.0001	60	0.00876
10	0.0001	61	0.00933
11	0.0001	62	0.01019
12	0.00011	63	0.0115
13	0.00012	64	0.01272
14	0.00014	65	0.0138
15	0.00022	66	0.01543
16	0.0003	67	0.01677
17	0.00047	68	0.01894
18	0.00056	69	0.02078
19	0.00057	70	0.02235
20	0.00063	71	0.02474
21	0.00063	72	0.02746
22	0.00061	73	0.03039
23	0.00066	74	0.03323
24	0.00065	75	0.03727
25	0.00066	76	0.04157
26	0.00074	77	0.04606
27	0.00074	78	0.05137
28	0.00081	79	0.05736
29	0.00084	80	0.06507
30	0.00092	81	0.07265
31	0.00092	82	0.08049
32	0.00097	83	0.08871
33	0.00103	84	0.10089
34	0.00116	85	0.11282
35	0.00126	86	0.12553
36	0.00125	87	0.13789
37	0.00132	88	0.15117

Age x	***qx*[a]**	***Age x***	***qx*[a]**
38	0.00149	89	0.15729
39	0.00155	90	0.16935
40	0.00168	91	0.18235
41	0.00179	92	0.20721
42	0.00185	93	0.23379
43	0.00199	94	0.25755
44	0.00222	95	0.284
45	0.00241	96	0.30686
46	0.00255	97	0.33477
47	0.00272	98	0.35701
48	0.00294	99	0.37245
49	0.00317	100	0.41787
50	0.00353		

[a] *qx*, is the mortality rate between age x and (x +1), that is the probability that a person aged x exact will die before reaching age (x +1).

Appendix 9 Cost effectiveness supplementary tables

Table 1 Incremental cost effectiveness for the base case with different NHS reference costs applied to the triage strategies

NHS reference cost (£)	Intervention	Cost (£)	QALYs	ICER
10	GDX	2841	19.7701	
	OCT	2967	19.7746	27812
	HRT MRA	3001	19.7771	13807
	HRT GPS	3011	19.7771	Dominated[a]
	Current Practice	3084	19.7780	98231
13	GDX	2856	19.7701	
	OCT	2982	19.7746	27784
	HRT MRA	3016	19.7771	13780
	HRT GPS	3026	19.7771	Dominated[a]
	Current Practice	3084	19.7780	80605
16	GDX	2872	19.7701	
	OCT	2996	19.7746	27757
	HRT MRA	3031	19.7771	13754
	HRT GPS	3040	19.7771	Dominated[a]
	Current Practice	3084	19.7780	62979
19	GDX	2887	19.7701	
	OCT	3011	19.7746	27729
	HRT MRA	3046	19.7771	13727
	HRT GPS	3055	19.7771	Dominated[a]
	Current Practice	3084	19.7780	45353
22	GDX	2902	19.7701	
	OCT	3026	19.7746	27701
	HRT MRA	3060	19.7771	13700
	HRT GPS	3070	19.7771	Dominated[a]
	Current Practice	3084	19.7780	27727
25	GDX	2917	19.7701	
	OCT	3041	19.7746	27673

	HRT MRA	3075	19.7771	13673
	Current Practice	3084	19.7780	10101
	HRT GPS	3085	19.7771	Dominated[a]
28	GDX	2932	19.7701	
	OCT	3056	19.7746	27646
	Current Practice	3084	19.7780	8313
	HRT MRA	3090	19.7771	Dominated[a]
	HRT GPS	3100	19.7771	Dominated[a]
31	GDX	2947	19.7701	
	OCT	3071	19.7746	27618
	Current Practice	3084	19.7780	3853
	HRT MRA	3105	19.7771	Dominated[a]
	HRT GPS	3115	19.7771	Dominated[a]
34	GDX	2962	19.7701	
	Current Practice	3084	19.7780	15579
	OCT	3086	19.7746	Dominated[a]
	HRT MRA	3120	19.7771	Dominated[a]
	HRT GPS	3129	19.7771	Dominated[a]
37	GDX	2977	19.7701	
	Current Practice	3084	19.7780	13663
	OCT	3101	19.7746	Dominated[a]
	HRT MRA	3135	19.7771	Dominated[a]
	HRT GPS	3144	19.7771	Dominated[a]
40	GDX	2992	19.7701	
	Current Practice	3084	19.7780	11747
	OCT	3116	19.7746	Dominated[a]
	HRT MRA	3149	19.7771	Dominated[a]
	HRT GPS	3159	19.7771	Dominated[a]
43	GDX	3007	19.7701	
	Current Practice	3084	19.7780	9831
	OCT	3130	19.7746	Dominated[a]
	HRT MRA	3164	19.7771	Dominated[a]
	HRT GPS	3174	19.7771	Dominated[a]

46	GDX	3022	19.7701	
	Current Practice	3084	19.7780	7315
	OCT	3145	19.7746	Dominated[a]
	HRT MRA	3179	19.7771	Dominated[a]
	HRT GPS	3189	19.7771	Dominated[a]
49	GDX	3037	19.7701	
	Current Practice	3084	19.7780	5999
	OCT	3160	19.7746	Dominated[a]
	HRT MRA	3194	19.7771	Dominated[a]
	HRT GPS	3204	19.7771	Dominated[a]
52	GDX	3052	19.7701	
	Current Practice	3084	19.7780	4083
	OCT	3175	19.7746	Dominated[a]
	HRT MRA	3209	19.7771	Dominated[a]
	HRT GPS	3218	19.7771	Dominated[a]
55	GDX	3067	19.7701	
	Current Practice	3084	19.7780	2168
	OCT	3190	19.7746	Dominated[a]
	HRT MRA	3224	19.7771	Dominated[a]
	HRT GPS	3233	19.7771	Dominated[a]
58	GDX	3082	19.7701	
	Current Practice	3084	19.7780	252
	OCT	3205	19.7746	Dominated[a]
	HRT MRA	3238	19.7771	Dominated[a]
	HRT GPS	3248	19.7771	Dominated[a]
61	Current Practice	3084	19.7780	
	GDX	3097	19.7701	Dominated[a]
	OCT	3220	19.7746	Dominated[a]
	HRT MRA	3253	19.7771	Dominated[a]
	HRT GPS	3263	19.7771	Dominated[a]
64	Current Practice	3084	19.7780	
	GDX	3112	19.7701	Dominated[a]
	OCT	3235	19.7746	Dominated[a]

	HRT MRA	3268	19.7771	Dominated[a]
	HRT GPS	3278	19.7771	Dominated[a]
67	Current Practice	3084	19.7780	
	GDX	3127	19.7701	Dominated[a]
	OCT	3250	19.7746	Dominated[a]
	HRT MRA	3283	19.7771	Dominated[a]
	HRT GPS	3292	19.7771	Dominated[a]
70	Current Practice	3084	19.7780	
	GDX	3142	19.7701	Dominated[a]
	OCT	3265	19.7746	Dominated[a]
	HRT MRA	3298	19.7771	Dominated[a]
	HRT GPS	3307	19.7771	Dominated[a]
73	Current Practice	3084	19.7780	
	GDX	3157	19.7701	Dominated[a]
	OCT	3279	19.7746	Dominated[a]
	HRT MRA	3313	19.7771	Dominated[a]
	HRT GPS	3322	19.7771	Dominated[a]
76	Current Practice	3084	19.7780	
	GDX	3172	19.7701	Dominated[a]
	OCT	3294	19.7746	Dominated[a]
	HRT MRA	3327	19.7771	Dominated[a]
	HRT GPS	3337	19.7771	Dominated[a]
79	Current Practice	3084	19.7780	
	GDX	3187	19.7701	Dominated[a]
	OCT	3309	19.7746	Dominated[a]
	HRT MRA	3342	19.7771	Dominated[a]
	HRT GPS	3352	19.7771	Dominated[a]
82	Current Practice	3084	19.7780	
	GDX	3202	19.7701	Dominated[a]
	OCT	3324	19.7746	Dominated[a]
	HRT MRA	3357	19.7771	Dominated[a]
	HRT GPS	3367	19.7771	Dominated[a]
85	Current Practice	3084	19.7780	

	GDX	3217	19.7701	Dominated[a]
	OCT	3339	19.7746	Dominated[a]
	HRT MRA	3372	19.7771	Dominated[a]
	HRT GPS	3381	19.7771	Dominated[a]

Table 2 Incremental cost effectiveness ratios of increasing costs of triage strategies and not treating patients diagnosed as "at risk"

Increasing cost of triage strategy (£)	**Intervention**	**Cost(£)**	**QALYs**	**ICER**
+10	GDX	2719	19.7393	
	OCT	2840	19.7410	68260
	HRT MRA	2869	19.7414	83488
	HRT GPS	2879	19.7414	Dominated[a]
	Current Practice	2954	19.7415	488759
+ 13	GDX	2733	19.7393	
	OCT	2853	19.7410	68229
	HRT MRA	2883	19.7414	83457
	HRT GPS	2893	19.7414	Dominated[a]
	Current Practice	2954	19.7415	409713
+16	GDX	2747	19.7393	
	OCT	2867	19.7410	68198
	HRT MRA	2897	19.7414	83426
	HRT GPS	2906	19.7414	Dominated[a]
	Current Practice	2954	19.7415	330667
+19	GDX	2761	19.7393	
	OCT	2881	19.7410	68167
	HRT MRA	2910	19.7414	83396
	HRT GPS	2920	19.7414	Dominated[a]
	Current Practice	2954	19.7415	251620
+22	GDX	2775	19.7393	
	OCT	2895	19.7410	68137

Increasing cost of triage strategy (£)	Intervention	Cost(£)	QALYs	ICER
	HRT MRA	2924	19.7414	83365
	HRT GPS	2934	19.7414	Dominated[a]
	Current Practice	2954	19.7415	172574
+25	GDX	2788	19.7393	
	OCT	2908	19.7410	68106
	HRT MRA	2938	19.7414	83335
	HRT GPS	2947	19.7414	Dominated[a]
	Current Practice	2954	19.7415	93527
+28	GDX	2802	19.7393	
	OCT	2922	19.7410	68075
	HRT MRA	2952	19.7414	83304
	Current Practice	2954	19.7415	14481
	HRT GPS	2961	19.7414	Dominated[a]
+31	GDX	2816	19.7393	
	OCT	2936	19.7410	68044
	Current Practice	2954	19.7415	34813
	HRT MRA	2965	19.7414	Dominated[a]
	HRT GPS	2975	19.7414	Dominated[a]
+34	GDX	2830	19.7393	
	OCT	2949	19.7410	68014
	Current Practice	2954	19.7415	8882
	HRT MRA	2979	19.7414	Dominated[a]
	HRT GPS	2989	19.7414	Dominated[a]
+37	GDX	2843	19.7393	
	Current Practice	2954	19.7415	48341
	OCT	2963	19.7410	Dominated[a]
	HRT MRA	2993	19.7414	Dominated[a]
	HRT GPS	3002	19.7414	Dominated[a]
+40	GDX	2857	19.7393	0
	Current Practice	2954	19.7415	42328
	OCT	2977	19.7410	Dominated[a]

Increasing cost of triage strategy (£)	Intervention	Cost(£)	QALYs	ICER
	HRT MRA	3006	19.7414	Dominated[a]
	HRT GPS	3016	19.7414	Dominated[a]
+43	GDX	2871	19.7393	
	Current Practice	2954	19.7415	36314
	OCT	2991	19.7410	Dominated[a]
	HRT MRA	3020	19.7414	Dominated[a]
	HRT GPS	3030	19.7414	Dominated[a]
+46	GDX	2885	19.7393	
	Current Practice	2954	19.7415	30300
	OCT	3004	19.7410	Dominated[a]
	HRT MRA	3034	19.7414	Dominated[a]
	HRT GPS	3043	19.7414	Dominated[a]
+49	GDX	2898	19.7393	
	Current Practice	2954	19.7415	24287
	OCT	3018	19.7410	Dominated[a]
	HRT MRA	3048	19.7414	Dominated[a]
	HRT GPS	3057	19.7414	Dominated[a]
+52	GDX	2912	19.7393	
	Current Practice	2954	19.7415	18273
	OCT	3032	19.7410	Dominated[a]
	HRT MRA	3061	19.7414	Dominated[a]
	HRT GPS	3071	19.7414	Dominated[a]
+55	GDX	2926	19.7393	
	Current Practice	2954	19.7415	12260
	OCT	3045	19.7410	Dominated[a]
	HRT MRA	3075	19.7414	Dominated[a]
	HRT GPS	3084	19.7414	Dominated[a]
+58	GDX	2940	19.7393	
	Current Practice	2954	19.7415	6246
	OCT	3059	19.7410	Dominated[a]
	HRT MRA	3089	19.7414	Dominated[a]

Increasing cost of triage strategy (£)	Intervention	Cost(£)	QALYs	ICER
	HRT GPS	3098	19.7414	Dominated[a]
+61	GDX	2954	19.7393	
	Current Practice	2954	19.7415	233
	OCT	3073	19.7410	Dominated[a]
	HRT MRA	3102	19.7414	Dominated[a]
	HRT GPS	3112	19.7414	Dominated[a]
+64	Current Practice	2954	19.7415	
	GDX	2967	19.7393	Dominated[a]
	OCT	3087	19.7410	Dominated[a]
	HRT MRA	3116	19.7414	Dominated[a]
	HRT GPS	3126	19.7414	Dominated[a]
+67	Current Practice	2954	19.7415	
	GDX	2981	19.7393	Dominated[a]
	OCT	3100	19.7410	Dominated[a]
	HRT MRA	3130	19.7414	Dominated[a]
	HRT GPS	3139	19.7414	Dominated[a]