Surveillance for ocular hypertension: an evidence synthesis and economic evaluation
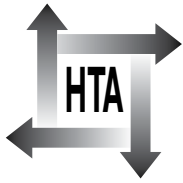
JM Burr, P Botello-Pinzon, Y Takwoingi, R Hernández, M Vazquez-Montes, A Elders, R Asaoka, K Banister, J van der Schoot, C Fraser, A King, H Lemij, R Sanders, S Vernon, A Tuulonen, A Kotecha, P Glasziou, D Garway-Heath, D Crabb, L Vale, A Azuara-Blanco, R Perera, M Ryan, J Deeks and J Cook

How to obtain copies of this and other HTA programme reports

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

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

How to order:

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

Additionally the HTA website allows you to either print out your order or download a blank order form.

Contact details are as follows:

Synergie UK (HTA Department)
Digital House, The Loddon Centre
Wade Road
Basingstoke
Hants RG24 8QW

Email: orders@hta.ac.uk

Tel: 0845 812 4000 – ask for 'HTA Payment Services'
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put 'HTA Order' on the fax header

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

Paying by credit card
You can order using your credit card by phone, fax or post.

Subscriptions

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

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

Please use the form on the HTA website (www.hta.ac.uk/htacd/index.shtml). *HTA on DVD* is currently free of charge worldwide.

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

Surveillance for ocular hypertension: an evidence synthesis and economic evaluation

JM Burr,[1]* P Botello-Pinzon,[1] Y Takwoingi,[2]
R Hernández,[1,3] M Vazquez-Montes,[4] A Elders,[1]
R Asaoka,[5,6] K Banister,[1] J van der Schoot,[7] C Fraser,[1]
A King,[8] H Lemij,[7] R Sanders,[9] S Vernon,[8] A Tuulonen,[10]
A Kotecha,[5,6] P Glasziou,[4] D Garway-Heath,[5] D Crabb,[6]
L Vale,[1,3] A Azuara-Blanco,[1] R Perera,[4] M Ryan,[3]
J Deeks[2] and J Cook[1]

[1]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
[2]Public Health, Epidemiology and Biostatistics Unit, University of Birmingham, Birmingham, UK
[3]Health Economics Research Unit, University of Aberdeen, Aberdeen, UK
[4]Department of Primary Health Care, University of Oxford, Oxford, UK
[5]Glaucoma Research Unit, Moorfields Eye Hospital NHS Foundation Trust, London, UK
[6]Department of Optometry and Visual Science, City University, London, UK
[7]Glaucoma Services, Rotterdam Eye Hospital, Rotterdam, the Netherlands
[8]Queen's Medical Centre, University Hospital, Nottingham, UK
[9]Queen Margaret Hospital, Dunfermline, UK
[10]Eye Centre, Tampere University Hospital, Tampere, Finland

*Corresponding author

This report should be referenced as follows:

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series
Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.
Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

Abstract

Surveillance for ocular hypertension: an evidence synthesis and economic evaluation

JM Burr,[1]* P Botello-Pinzon,[1] Y Takwoingi,[2] R Hernández,[1,3] M Vazquez-Montes,[4] A Elders,[1] R Asaoka,[5,6] K Banister,[1] J van der Schoot,[7] C Fraser,[1] A King,[8] H Lemij,[7] R Sanders,[9] S Vernon,[8] A Tuulonen,[10] A Kotecha,[5,6] P Glasziou,[4] D Garway-Heath,[5] D Crabb,[6] L Vale,[1,3] A Azuara-Blanco,[1] R Perera,[4] M Ryan,[3] J Deeks[2] and J Cook[1]

[1]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
[2]Public Health, Epidemiology and Biostatistics Unit, University of Birmingham, Birmingham, UK
[3]Health Economics Research Unit, University of Aberdeen, Aberdeen, UK
[4]Department of Primary Health Care, University of Oxford, Oxford, UK
[5]Glaucoma Research Unit, Moorfields Eye Hospital NHS Foundation Trust, London, UK
[6]Department of Optometry and Visual Science, City University, London, UK
[7]Glaucoma Services, Rotterdam Eye Hospital, Rotterdam, the Netherlands
[8]Queen's Medical Centre, University Hospital, Nottingham, UK
[9]Queen Margaret Hospital, Dunfermline, UK
[10]Eye Centre, Tampere University Hospital, Tampere, Finland

*Corresponding author

Objectives: To determine effective and efficient monitoring criteria for ocular hypertension [raised intraocular pressure (IOP)] through (i) identification and validation of glaucoma risk prediction models; and (ii) development of models to determine optimal surveillance pathways.

Design: A discrete event simulation economic modelling evaluation. Data from systematic reviews of risk prediction models and agreement between tonometers, secondary analyses of existing datasets (to validate identified risk models and determine optimal monitoring criteria) and public preferences were used to structure and populate the economic model.

Setting: Primary and secondary care.

Participants: Adults with ocular hypertension (IOP > 21 mmHg) and the public (surveillance preferences).

Interventions: We compared five pathways: two based on National Institute for Health and Clinical Excellence (NICE) guidelines with monitoring interval and treatment depending on initial risk stratification, 'NICE intensive' (4-monthly to annual monitoring) and 'NICE conservative' (6-monthly to biennial monitoring); two pathways, differing in location (hospital and community), with monitoring biennially and treatment initiated for a ≥6% 5-year glaucoma risk; and a 'treat all' pathway involving treatment with a prostaglandin analogue if IOP > 21 mmHg and IOP measured annually in the community.

Main outcome measures: Glaucoma cases detected; tonometer agreement; public preferences; costs; willingness to pay and quality-adjusted life-years (QALYs).

Results: The best available glaucoma risk prediction model estimated the 5-year risk based on age and ocular predictors (IOP, central corneal thickness, optic nerve damage

and index of visual field status). Taking the average of two IOP readings, by tonometry, true change was detected at two years. Sizeable measurement variability was noted between tonometers. There was a general public preference for monitoring; good communication and understanding of the process predicted service value. 'Treat all' was the least costly and 'NICE intensive' the most costly pathway. Biennial monitoring reduced the number of cases of glaucoma conversion compared with a 'treat all' pathway and provided more QALYs, but the incremental cost-effectiveness ratio (ICER) was considerably more than £30,000. The 'NICE intensive' pathway also avoided glaucoma conversion, but NICE-based pathways were either dominated (more costly and less effective) by biennial hospital monitoring or had a ICERs >£30,000. Results were not sensitive to the risk threshold for initiating surveillance but were sensitive to the risk threshold for initiating treatment, NHS costs and treatment adherence.

Limitations: Optimal monitoring intervals were based on IOP data. There were insufficient data to determine the optimal frequency of measurement of the visual field or optic nerve head for identification of glaucoma. The economic modelling took a 20-year time horizon which may be insufficient to capture long-term benefits. Sensitivity analyses may not fully capture the uncertainty surrounding parameter estimates.

Conclusions: For confirmed ocular hypertension, findings suggest that there is no clear benefit from intensive monitoring. Consideration of the patient experience is important. A cohort study is recommended to provide data to refine the glaucoma risk prediction model, determine the optimum type and frequency of serial glaucoma tests and estimate costs and patient preferences for monitoring and treatment.

Funding: The National Institute for Health Research Health Technology Assessment Programme.

Contents

List of abbreviations

AGIS	Advanced Glaucoma Intervention Study
BB	beta-blocker
BNF	*British National Formulary*
C/D	cup-to-disc
CCC	concordance correlation coefficient
CCT	central corneal thickness
CDSR	Cochrane Database of Systematic Reviews
CH	corneal hysteresis
CHD	coronary heart disease
CI	confidence interval
CoV	coefficient of variation
CPSD	corrected pattern standard deviation
DARE	Database of Abstracts of Reviews of Effects
DCE	discrete choice experiment
DCT	dynamic contour tonometer
DIGS	Diagnostic Innovations in Glaucoma Study
EGPS	European Glaucoma Prevention Study
EQ-5D	European Quality of Life-5 Dimensions
GAT	Goldmann applanation tonometry
GHT	glaucoma hemifield test
HTA	Health Technology Assessment
HR	hazard ratio
ICC	intracluster correlation
ICER	incremental cost-effectiveness ratio
ID	identifier
IOP	intraocular pressure
IOP_{CC}	corneal-compensated intraocular pressure
IPD	individual patient data
IQR	interquartile range
LASIK	laser-assisted in situ keratomileusis
LoA	limits of agreement
LVCF	last value carried forward
MD	mean deviation
NCT	non-contact tonometer
NICE	National Institute for Health and Clinical Excellence
OAG	open-angle glaucoma
OHT	ocular hypertension
OHTS	Ocular Hypertension Treatment Study
OPA	ocular pulse amplitude
ORA	ocular response analyser
PEX	pseudoexfoliation syndrome
PGA	prostaglandin analogue
PPT	pressure phosphene tonometer
PSD	pattern standard deviation
QALY	quality-adjusted life-year
RC	repeatability coefficient
RCT	randomised controlled trial
RNFL	retinal nerve fibre layer

RT	rebound tonometer
SAP	standard automated perimetry
SD	standard deviation
SOH	surveillance for ocular hypertension
SMD	standardised mean difference
Varmid	variance of the difference between the middle reading and the average of first and last readings
VCD	vertical cup-to-disc
WC	work centre
WEP	work entry point
WTP	willingness to pay

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

Executive summary

Background

Glaucoma is a chronic progressive optic neuropathy leading to impaired vision and blindness if inadequately treated. Open-angle glaucoma is the most common form. A raised intraocular pressure (IOP) is the only modifiable risk factor. Ocular hypertension (OHT) is defined as IOP > 21 mmHg and the absence of clinical signs of glaucoma.

Around 1 million people in the UK have OHT with most identified during a routine 'sight' test; diagnosis is typically confirmed in secondary care. Treatment (daily eye drops) may be indicated to reduce IOP. Surveillance should identify those who would benefit from treatment, and should be affordable and acceptable to patients.

Clinical management establishes that OHT is truly present. Once confirmed, monitoring includes measuring IOP by tonometry and tests to detect glaucoma [visual field by standard automated perimetry (SAP) and evaluation of structural changes in the optic nerve]. Outcomes from all three parameters inform whether or not treatment is necessary. Long-term surveillance requires interpretation of serial tests and, for those requiring treatment, the responsiveness of IOP to treatment. In choosing the monitoring frequency, a strategy that separates 'true' long-term change (signal) from short-term variation and measurement error 'noise' is required.

Guidelines were published by the National Institute for Health and Clinical Excellence (NICE) in 2009, but few data were available to guide best monitoring practice.

Aim

To determine effective and efficient monitoring criteria for OHT.

Objectives

1. To identify and validate the most relevant tool(s) for predicting risk of developing glaucoma.
2. To determine optimal monitoring criteria. (Which tests? How often?).
3. To determine public preferences for a service, taking into account health outcomes and patient experiences.
4. To undertake an economic evaluation of different surveillance pathways, considering costs, clinical outcomes, quality-adjusted life-years (QALYs) and willingness to pay (WTP).
5. To determine risk thresholds for initiating surveillance.
6. To make recommendations for research.

Methods

The study comprised three interlinked substudies, described in *Figure 1*.
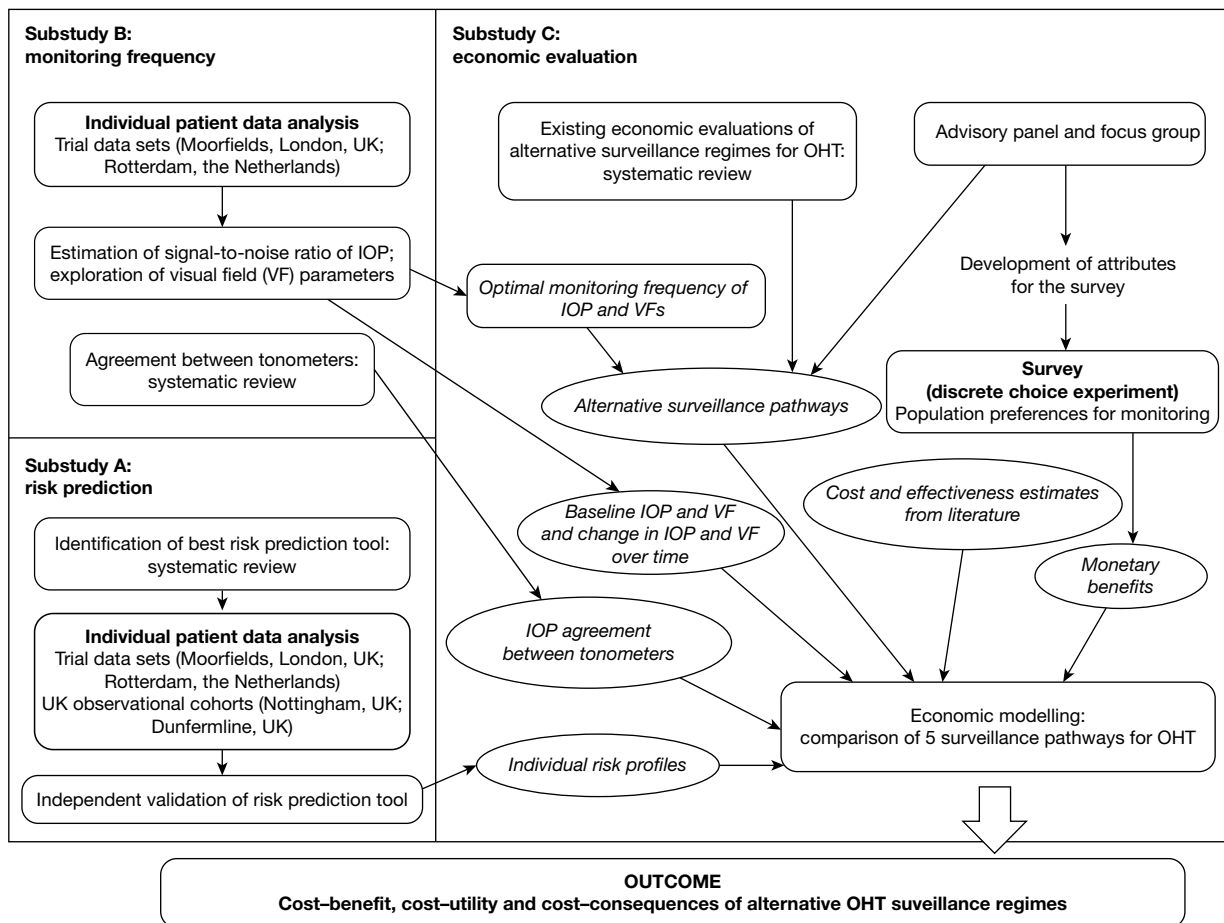
FIGURE 1 Overview of the project.

A: risk prediction models (objectives 1 and 6)

This involved (1) systematic review of prediction models estimating risk of progression to glaucoma and (2) external validation of the most robust using individual patient data from four populations: Moorfields Eye Hospital, London, and Rotterdam Eye Hospital, the Netherlands [randomised controlled trials (RCTs)], and Dunfermline and Nottingham [UK observational cohorts (hospital- and community-based, respectively)]. The 5-year risk was calculated using the prediction equation and the patients' observed or imputed value of the predictors. The discriminatory ability of the model was assessed using Harrell's c-index. Model calibration was assessed using calibration plots and calibration slopes.

B: optimal monitoring criteria (objectives 2 and 6)

This involved (1) evidence synthesis of the measurement agreement between alternative tonometers and the reference standard, Goldmann applanation tonometry (GAT), and (2) statistical modelling of the variability of IOP and visual field indices over time:

1. A systematic search was undertaken. Meta-analyses of the mean differences and the standard deviation of the differences were undertaken for the agreement between each tonometer and GAT. Summary 95% limits of agreement were generated. Data on study characteristics, recordability, acceptability, practicality and reliability were summarised.
2. Secondary analysis of data from the placebo arm of the London trial, conducted in the mid-1990s to evaluate medical treatment for OHT, was used to estimate:
   i. average true long-term change of the whole group

ii. short-term variability: combination of analytic variability and biological fluctuation around a stable average – the 'noise'

iii. long-term variability: variation in long-term true change between individuals – the 'signal'.

The variability was estimated using a direct method and a linear random-effects model. Three imputation methods were employed for missing data. The effect of age and observer was analysed by fitting a separate model for each covariate in the London data set. The models were externally validated using patient-level data from the Rotterdam trial conducted between 1997 and 2008.

C: health economic evaluation (objectives 3–6)

This involved two components:

1. A *discrete choice experiment (DCE)* to investigate the relative importance of attributes of a monitoring service to the public. Attributes and levels were identified by an advisory panel and a focus group. Attributes included a description of health outcomes ('10-year risk of developing glaucoma, severe glaucoma and visual impairment' and 'unwanted effects of treatment'), patient experience ('communication/understanding' and 'location') and a cost attribute (price proxy) to provide a composite monetary measure of utility (WTP). Each DCE question involved a choice between two monitoring programmes (differing in the levels of the attributes) and a no-monitoring alternative. A Bayesian experimental design was used to determine choices, using information from a pilot ($n = 184$). Data were collected using a web-based survey ($n = 814$). The conditional logit model was used to analyse aggregate data and subgroup analyses by age.

2. A *discrete event simulation model* to assess the relative efficiency of monitoring strategies for those with OHT, estimated by cost-effectiveness, cost–utility and cost–benefit analyses (using monetary values generated from the DCE). Pathways were informed by NICE guidelines, by the literature and in consultation with clinical experts, service users and the DCE. The model was populated with parameter estimates informed by components of earlier objectives and the literature. Sensitivity analyses, for the cost–utility analysis, explored the effect of monitoring or treating higher risk only, reducing the unit cost of the prostaglandin analogue by 50% and reducing NHS costs. A scenario analysis was conducted to determine the effects of varying estimates of adherence to medication, IOP measurement precision, accuracy of glaucoma detection and rate of progression to glaucoma.

Results

A: risk prediction models

Three models were identified, derived using data from two large multicentre RCTs, the Ocular Hypertension Treatment Study (OHTS) and the European Glaucoma Prevention Study (EGPS), evaluating ocular hypotensive medication. The OHTS-EGPS means model, the most robust, estimates the 5-year risk based on age, IOP, central corneal thickness (CCT), vertical cup-to-disc (C/D) ratio and pattern standard deviation (PSD); all variables are routinely collected in clinical practice. The model uses the mean values of the right and left eyes of an individual to calculate eye-specific predictors.

The discriminatory ability was good in the four populations tested, with *c*-indexes between 0.69 and 0.83; however, in calibration analyses, the OHTS-EGPS means model generally overestimated the risk of glaucoma, although for the Rotterdam cohort the calibration slope was close to 1 (1.09, 95% confidence interval 0.72 to 1.46), the ideal value when there is complete agreement between predicted and observed risks.

This OHTS-EGPS model was developed using selected trial populations and may not include all important predictors. Both trial and observational validation cohorts were highly selected and none satisfactorily covered the full spectrum of risk. Furthermore, missing data for the predictors was considerable in all cohorts and definitions of glaucoma were not standardised. Despite these limitations, the model is useful in conjunction with clinical assessment.

B: optimal monitoring criteria

Systematic review and meta-analysis of tonometers

A total of 102 comparative studies assessed the agreement of at least one tonometer with GAT (HAAG-STREIT, Koeniz, Switzerland). Comparators were dynamic contour tonometer, non-contact tonometer (NCT) (Canon USA, Inc., Lake Success, NY, USA; Keeler Ltd, Windsor, UK; NIDEK Co. Ltd, Gamagori, Japan; Reichert Ophthalmic Instruments, Buffalo, NY, USA; Topcon Corporation, Tokyo, Japan), Ocular response analyser® (ORA) (Reichert Inc., Depew, NY, USA), Ocuton S® (EPSa Elektronik & Präzisionsbav, Saalfeld, Germany), Perkins® (Kowa HA-2, Kowa, Japan), rebound tonometer, TonoPen® (Mentor O&O Inc., Santa Barbara, CA; Reichert Inc., Depew, NY, USA) and transpalpebral tonometer. Studies were generally poorly reported. The agreement in IOP (95% limits) varied across tonometers, from 0.2 mmHg (–3.8 to 4.3 mmHg) for NCT to 2.7 mmHg (–4.1 to 9.6 mmHg) for Ocuton S. Sizeable inter- and intraobserver variability was observed for all tonometers, including GAT, casting doubt on the validity of GAT as the default standard.

Optimal frequency of monitoring intraocular pressure and tests to detect glaucoma

Statistical modelling was performed on ocular measures from the London placebo group {$n = 153$; mean IOP 24.4 mmHg [standard deviation (SD) 3.5 mmHg], 14 4-monthly visits}. Validation was performed using the Rotterdam placebo data [$n = 132$, mean IOP 25.7 mmHg (SD 2.5 mmHg), 21 visits biannually].

A linear random-effects model, using the last value carried forward to impute missing data, was the best fit to IOP data. The average change in IOP over time for the whole group was < 1 mmHg in 3 years, although a ≥ 5 mmHg change occurred in 25% of participants. For most individuals any true change in underlying IOP ('signal') was smaller than the estimated 'noise'. Observed changes, using a single measure of IOP at each time point, of ≤ 3 mmHg can be explained as 'noise'. Assuming independence of repeated measures, the mean of two baseline IOP readings increased the signal-to-noise ratio such that true change in IOP of 2 mmHg could be detected at 2 years. With three baseline measures averaged, true change could be detectable between 1 and 2 years.

For lower baseline IOP (< 26 mmHg) the model suggested that a true change in IOP would be unlikely within 3 years. The model may have underestimated the small proportion of individuals with a large change in IOP; this was adjusted for in the economic model. Mean deviation (MD) data, a visual field index measured by SAP, were available only in the London data set. MD fluctuated, increasing and decreasing, with minimal signal detected over 4 years.

Because of limited patient data on sequential measures of visual fields, the determination of optimal monitoring frequency was based on IOP variability.

C: health economic evaluation

Discrete choice experiment

There was a general public preference for monitoring of individuals with OHT. Individuals were willing to pay £28 per year for a service, everything else being equal. Coefficients representing

each attribute, other than hospital location, were significant predictors of preferences at the 1% level. Specifically, marginal valuations of the risk of glaucoma and sight loss over 10 years were statistically significant in the expected direction but small. Side effects of treatment reduced the value of any service, with more disbenefits as side effects increased. Good communication with the health professional and understanding of the testing process were important predictors of the value of alternative services. Preferences varied according to age, with those aged > 50 years being less concerned with the risk of sight loss (compared with those aged < 50 years), but more concerned about treatment side effects and the importance of good communication and understanding of the process.

Economic modelling evaluation

Five pathways were compared. Two were based on NICE guidelines with monitoring interval depending on initial risk stratification: 'NICE intensive' – 4-monthly to annual monitoring – and 'NICE conservative' – 6-monthly to biennial monitoring – with treatment according to baseline risk stratification by age, IOP and CCT; two further pathways, differing in location ['surveillance for ocular hypertension (SOH) hospital' and 'SOH primary care'], included monitoring biennially with treatment initiated for a ≥ 6% 5-year glaucoma risk. The pathways included repeated IOP measurements, within 4 months, following treatment initiation or change. A 'treat all' pathway involved treatment if IOP was > 21 mmHg, measurement of IOP annually in community optometry and referral to secondary care if treatment response was inadequate (< 15% IOP reduction).

'Treat all' was the least costly pathway and 'NICE intensive' the most costly pathway. The 'SOH hospital' pathway reduced the number of cases of conversion to glaucoma compared with the 'treat all' pathway and provided more QALYs but the incremental cost per extra QALY (incremental cost-effectiveness ratio; ICER) was considerably more than £30,000. The 'NICE intensive' pathway also avoided conversion to glaucoma, but NICE-based pathways were dominated (more costly and less effective) by the 'SOH hospital' pathway. In the cost–benefit analysis, compared with 'no monitoring' the 'SOH hospital' pathway was the only pathway to show net benefit.

Results were sensitive to the risk threshold for initiating treatment, NHS costs and treatment adherence. If treatment was initiated when the 5-year risk of developing glaucoma was > 10% (e.g. a 60-year-old with an IOP of 27 mmHg, CCT of 560 μm, vertical cup-to-disc (VCD) ratio of 0.4 and a PSD of 1.4 dB has a 10.3% risk in at least one eye), an 'SOH hospital' pathway was less costly and more effective than a 'treat all' pathway. The SOH pathways had ICERs of < £30,000 compared with the 'treat all' pathway when service cost for repeat IOP measurement, in response to treatment change, was < £60. Differences in treatment adherence between the 'treat all' and SOH pathways of approximately 40% or higher led to the SOH pathways having ICERs of ≤ £30,000. NICE-based pathways were more costly and either were dominated or had ICERs well above £30,000 per QALY.

In the cost–utility analysis, surveillance was not compared with a 'no monitoring' alternative as this was not an acceptable option given current NHS policy. A 'treat all' pathway was included based on emerging findings from the literature. The acceptability to users and health-care professionals of a 'treat all' pathway was not explored. The modelling took a 20-year time horizon, which may be insufficient to capture longer-term benefits. Sensitivity analyses conducted may not fully capture the uncertainty surrounding parameter estimates. Although patient views were consulted when developing the DCE, the results were based on public preferences, which may differ from those of patients.

Conclusions

Implications for health care

The best available prediction model (OHTS-EGPS means model) estimates the 5-year risk of glaucoma based on age and the ocular predictors IOP, CCT, VCD ratio and PSD. An IOP measurement algorithm using the average of repeat measurements at one visit reduces noise. Our findings support the clinical importance of establishing a true baseline IOP prior to initiating monitoring or treatment. IOP measurement using the NCT or hand held applanation tonometer appears to give the closest agreement with GAT with > 75% of measurements within 3 mmHg. However, findings suggest that GAT may not be the most appropriate reference standard. The same type of tonometer should be used to compare IOP measurements in an individual.

Our findings, based on a small sample, suggest biennial IOP monitoring for untreated or stable treated OHT. The optimal frequency of clinical testing (perimetry or optic nerve evaluation) to detect glaucoma remains uncertain. The economic evaluation suggests no clear benefit in intensive monitoring to detect glaucoma; any service reconfigurations should consider patient experiences, ensuring adequate time to explain the purpose of monitoring and avoid treatment side effects. If the NHS costs for repeat visits to monitor IOP response to treatment are minimised, biennial hospital-based monitoring appears optimal. The economic model may not have fully captured data uncertainties or the opportunity cost of resource use. The feasibility of community care pathways should be explored.

Recommendations for research

- A prospective cohort study including a representative sample of newly diagnosed OHT to update the risk prediction model and evaluate the optimal interval of serial glaucoma tests; standardisation of a measure of perimetry or optic nerve analysis with consensus on glaucoma conversion criteria; a comparison of alternative tonometers; costs and patient preferences for surveillance and treatment; and an updated economic model.
- Further development of tonometers to meet the needs of patients and the NHS.

Funding

Chapter 1

Background

Description of the health problem (glaucoma)

Glaucoma is a chronic progressive optic neuropathy leading to impaired vision and blindness if left untreated. The underlying pathophysiology is an accelerated loss of retinal ganglion cells compared with the normal ageing process, leading to a characteristic pattern of retinal nerve fibre layer (RNFL) atrophy, optic nerve head changes and loss of peripheral visual function. Open-angle glaucoma (OAG) is the most common form of glaucoma, affecting about 2% of the population aged > 40 years in the UK.[1] In the UK, glaucoma is second to age-related macular degeneration as a main cause of blindness.[2,3]

Open-angle glaucoma is diagnosed primarily by detecting the structural, that is glaucomatous, optic neuropathy (characteristic changes of the optic nerve head, the optic disc) and functional loss based on a corresponding visual field defect. The drainage pathway, the anterior chamber angle, appears open and normal. Early diagnosis is difficult as there is an overlap between normal appearance and glaucomatous changes in terms of both optic disc changes and visual field loss. A repeatable visual field defect consistent with structural damage of the optic nerve determines whether or not OAG is truly present. Having a raised intraocular pressure (IOP) is the most important risk factor for developing glaucoma and is the only one that is treatable. OAG is treated by interventions to lower the IOP. These include a variety of topical ocular hypotensive medications (eye drops), laser trabeculoplasty or glaucoma drainage surgery or a combination of these.

Ocular hypertension as a risk factor for open-angle glaucoma

Risk factors for developing OAG include raised IOP, greater age, non-white ethnicity (those of black ethnicity being a higher risk group), having a first-degree relative with OAG, myopia and diabetes.[4] Ocular hypertension (OHT) is generally defined as an IOP of > 21 mmHg [2 standard deviations (SDs) above the population mean IOP] in the absence of clinical signs of glaucoma (optic nerve damage or visual field loss consistent with glaucoma). The risk of developing OAG, and of worsening of existing disease, increases with increasing IOP,[5–7] and those presenting with advanced glaucoma at diagnosis have higher IOPs.[8,9] The population prevalence estimates for OHT range from 4.5% to 9.4% for those aged > 40 years,[10] with prevalence increasing with age.[10,11]

The effectiveness of treatment for OHT has been evaluated in two large trials of medical treatment [the Ocular Hypertension Treatment Study (OHTS)[12,13] and the European Glaucoma Prevention Study (EGPS)].[14]

Combining the findings from these studies and data from eight smaller trials, a Cochrane systematic review and meta-analysis by Vass and colleagues,[15] evaluating different topical ocular hypotensive medications against placebo or no treatment, showed that lowering the IOP for individuals with OHT was effective, with about a 40% reduction in the incidence of glaucomatous visual field defects with treatment at 5 years [odds ratio 0.62; 95% confidence interval (CI) 0.5 to 0.8].

Baseline factors predicting the development of OAG for those with OHT have been identified. These are described in detail in *Chapter 4*. Because of the risk of developing glaucoma, patients with OHT, either treated or untreated, are monitored in terms of IOP, visual function as measured by a visual field test, usually standard automated perimetry (SAP), and an assessment of the structural damage of the optic nerve and the RNFL.

The uncertainty lies in ascertaining, when OHT exists, how many measures are required to define the baseline IOP, and at what level of IOP should monitoring to detect early glaucoma be recommended. In addition, there is uncertainty whether monitoring should vary according to baseline risk factors for the development of OAG.

Impact of the health problem

Significance for patients

An effective monitoring service for people with OHT has the potential to directly benefit those with OHT in terms of identifying the need for treatment and thus reducing the risk of sight loss due to glaucoma. However, this has to be balanced against the inconvenience of attending for monitoring and any potential harms of monitoring. Although it is important to treat early disease to reduce the costs and disability associated with OAG, only a minority of those with OHT develop OAG, and not all of these are at risk of visual impairment in their lifetime. The challenge is therefore to identify the groups most at risk of developing significant OAG, and to determine effective and efficient monitoring criteria to identify those individuals who are most likely to benefit from treatment.

Significance for the NHS

Around 1 million people in the UK have OHT, assuming an estimate of the prevalence of OHT of 5%[10] applied to the UK population aged > 40 years.[16] Guidelines for the management of OHT and OAG in England, Wales and Northern Ireland were published by the National Institute for Health and Clinical Excellence (NICE) in 2009.[1] The guideline review was based on an economic modelling evaluation supported by evidence from systematic reviews when available. However, it was acknowledged that there was little evidence to guide monitoring possibilities for patients with OHT and glaucoma.

The potential benefits to the NHS of evidence-based guidelines for monitoring people with OHT include:

■ reducing inappropriate referrals and retaining capacity in hospital services for higher-risk patients
■ ensuring that all people who meet the criteria have access to surveillance services
■ managing capacity in the hospital eye service by monitoring people in the community if the latter is an effective and cost-effective pathway
■ increasing patient choice and accessibility so that resources are used appropriately
■ better value for money.[17]

Current service provision

In the UK, the majority of people are identified as having OHT during a 'sight' test, usually to obtain glasses, at a community-based optometrist. People aged ≥ 60 years and those ≥ 40 years with a self-reported family history of glaucoma or considered to be at risk of glaucoma by an

ophthalmologist qualify for a free 'sight test' under the General Ophthalmic Services in England, Wales and Northern Ireland. The UK College of Optometrists guidelines for examining a patient with glaucoma or at risk of glaucoma suggest that an eye examination should normally include an assessment of the optic nerve head and tonometry (measurement of IOP). The guidelines advise that, when IOP is high or borderline, arrangements should be made for the test to be repeated. A central visual field assessment using perimetry with threshold control may also be included at the discretion of the optometrist and practitioners should consider repeating visual field assessment to obtain a meaningful result.[18]

In Scotland, since April 2006, everyone is eligible for a free 'sight' test, which includes a comprehensive eye examination appropriate to need. An eye examination is recommended annually for those with OHT.

There is considerable debate about the role and optimal organisation of a monitoring service for those with OHT and thus at risk of glaucoma. The NICE guidelines[1] aimed to include recommendations on the most appropriate service models for monitoring OHT where evidence of effectiveness is available. OHT is defined in the guidelines as consistently or recurrently elevated IOP > 21 mmHg [measured with slit lamp-mounted Goldmann applanation tonometry (GAT) on more than one occasion] in the absence of optic nerve damage or corresponding visual field defect. The guidelines recommend that optometrists who have been trained to work in the field of glaucoma should refer people with OHT or suspected glaucoma, based on optic nerve damage or repeatable visual field defect, or both, to a consultant ophthalmologist-led service for definitive diagnosis and formulation of a management plan.

The guidelines provide recommendations for monitoring people with OHT and treatment recommendations according to risk (see *Chapter 8* for a more detailed description). Briefly, the recommendations include monitoring IOP by GAT and monitoring of clinical status change [visual fields by automated perimetry (suprathreshold perimetry is acceptable)]. Van Herick's peripheral anterior chamber depth assessment and stereoscopic slit lamp biomicroscopic examination of the optic nerve head are also recommended. Where GAT is not practical, the guidelines suggest Perkins hand-held tonometry as an acceptable alternative.[1]

In Scotland, the NICE guidelines are advisory. The Centre for Change and Innovation for NHS Scotland has developed patient pathways for ophthalmology designed to facilitate optimal patient care. Close collaboration between primary and secondary care is encouraged with the aim of optimising the expertise available in the community, particularly between optometry, general practitioners (GPs) and the hospital eye service. Currently there is no documented patient pathway for the management of OHT or suspect glaucoma.[19] Optometrists in Scotland, under the National Health Service (General Ophthalmic Services) (Scotland) Amendment Regulations 2010,[20] monitor people with OHT (IOP ≥ 21 mmHg as determined by GAT) annually and follow the NICE guidance as to when to refer to the hospital eye service.

Current service cost

The annual additional cost of implementing the NICE guidelines for monitoring people with OHT or suspect glaucoma in England, Wales and Northern Ireland has been estimated as £11M, with potential cost savings of moving hospital-based monitoring to the community of around £7.5M. Future NHS savings, resulting from fewer people losing their vision because of improved treatment and monitoring of those with OHT, are estimated at £0.6M per year for England.[21]

Variation in services

Clinical guidelines for the management of OHT and suspect glaucoma have been developed for other countries; these are consistent with the NICE guidelines in terms of the tests required, but recommendations for the frequency of monitoring according to risk are based on expert consensus opinion.[22–24]

Uncertainties regarding best practice for managing ocular hypertension

Measurement of intraocular pressure

Although the NICE guidelines advocate that IOP should be measured with contact tonometry by ophthalmologists (GAT is the most widely used), if surveillance is to be considered in primary care, tonometers that do not touch the cornea may be more applicable. In addition, a thick or thin cornea can lead to measurement error in tonometry, including GAT.[25–27] New tonometers are available that account for the biomedical properties and thickness of the cornea. In addition, non-invasive self-measurement devices are available and may be highly appropriate and relevant as monitoring devices. The measurement performance of the alternative tonometers has not been systematically compared.

Defining individual risk of developing glaucoma and treatment policy

Although it is important to treat early disease to reduce the costs and disability associated with OAG, only a minority of those with OHT develop OAG and of these not all are at risk of visual impairment in their lifetime. The challenge is therefore to identify the groups most at risk of developing significant OAG, and to determine effective and efficient monitoring criteria to identify those individuals who are most likely to benefit from treatment.

The decision problem to be addressed is whether to treat all with OHT defined as an IOP > 21 mmHg, withhold treatment until a patient shows signs of early glaucoma or to target treatment only to those with OHT at a higher risk of developing glaucoma. Risk prediction algorithms are useful tools for risk assessment for conversion from OHT to OAG. The applicability of published risk prediction models needs to be evaluated in a UK context, and the cost-effectiveness of monitoring according to risk thresholds would inform policy decisions on managing OHT.

Frequency of testing (intraocular pressure measurement, perimetry and optic nerve assessment)

The aims of monitoring people with OHT is (1) initially to establish that OHT is truly present and not 'noise' in the IOP measurement; (2) to decide the long-term variability of the monitoring tests among individuals, commonly referred to as the 'signal' for monitoring purposes, to establish whether or not prophylactic treatment to reduce IOP is advisable; (3) to measure changes in the level of IOP in response to treatment; and (4) to detect those individuals progressing to glaucoma in a timely manner. There is currently insufficient evidence to guide clinicians with respect to the optimal intertest spacing and frequency of testing necessary to identify clinically significant changes in IOP or the optimal frequency and spacing of testing required to detect clinically significant changes in the visual field or optic nerve over a relevant time period, which could be up to 5 years.

Service provision

There is uncertainty regarding the best model for service provision in terms of the best location for a monitoring service, either in primary care or in a specialist-led secondary-care service.

Cost-effectiveness of alternative surveillance strategies for ocular hypertension

Across Europe the annual direct costs associated with managing OHT and glaucoma were estimated in 2003 as €455 per person for OHT, increasing to €595 for early glaucoma and rising to €969 for end-stage disease.[28] These estimates are, however, very imprecise as they are based on retrospectively collected patient data from only 194 patient records collected across four countries, including three centres in the UK. Costs increase with increasing disease stage, with the main cost driver of the total direct health-care cost of glaucoma care being the cost of glaucoma medications. In the USA, a cost–utility analysis, taking a societal perspective and using the OHTS data, found that treatment for individuals with an IOP ≥ 24 mmHg and a ≥ 2% annual risk of developing glaucoma treatment would be cost-effective if society were willing to pay $50,000 for an additional quality-adjusted life-year (QALY).[29]

A recent economic modelling evaluation of monitoring pathways for OHT by van Gestel and colleagues[30,31] suggested that a pathway of treating all with an IOP > 21 mmHg was cost-effective in a Dutch context and concluded that direct treatment of OHT is associated with health gains and cost savings compared with monitoring until conversion to glaucoma.

This study aims to determine the optimum frequency of monitoring individuals identified as being at risk of glaucoma because of OHT and to determine the extent to which this varies according to an individual's risk characteristics. Several methodological approaches were used across three interlinked substudies: expert and consumer opinions; population preferences; systematic reviews; individual patient data (IPD) analysis; and decision-analytic modelling comparing alternative surveillance regimes taking into account the NHS costs and the health and wider benefits, as well as public preferences for the process of care, health and other outcomes.

This study was commissioned prior to the publication of the NICE guidelines. The remit of this study does not include an evaluation of the training required for health-care professionals or the optimal location of testing, although the survey of the general public and the subsequent incorporation of these findings within the economic model evaluate preferences for alternative models of care.

Chapter 2

Aim and objectives

Aim

The overall aim was to determine effective and efficient monitoring criteria for individuals with OHT taking into account the NHS costs and the health and wider benefits, as well as public preferences for the process of care, health and other outcomes.

Objectives

The objectives were to:

1. identify and validate the most relevant prediction tool(s) for predicting the risk of developing OAG in individuals with OHT (see *Chapter 4*)
2. determine optimal monitoring criteria (which tests, frequency of monitoring) (see *Chapters 5* and *6*)
3. determine public preferences for a service, taking into account health outcomes and patient experience factors (see *Chapter 7*)
4. undertake an economic evaluation of different surveillance pathways, considering costs, clinical outcomes, QALYs and willingness to pay (WTP) (see *Chapter 8*)
5. determine risk thresholds for initiating surveillance (see *Chapter 8*)
6. make recommendations for surveillance regimes for OHT and
7. identify future research needs (see *Chapters 4–8*).

The study involves three linked substudies (A, B and C), which are illustrated in *Figure 1*. Priorities for future research are identified from the outputs of each substudy.

A: risk prediction

Chapter 4 describes (1) the systematic identification and critical appraisal of risk prediction models for the progression of OHT to OAG and (2) a comparison of the performance of the models using IPD from four identified cohorts [randomised controlled trials (RCTs) and observational cohorts] of individuals with raised IOP. A schematic diagram showing the outputs of substudy A is given in *Figure 2*.

B: monitoring criteria

To determine effective monitoring criteria to identify those individuals who will most benefit from treatment as well as to identify those who may not need treatment or follow-up we undertook (1) a systematic review to determine the level of agreement between IOP measurement devices (tonometers) (reported in *Chapter 5*) and (2) a statistical modelling evaluation using IPD from identified RCTs to estimate values that would inform the optimal frequency for monitoring according to IOP for individuals with OHT. An exploratory analysis of the variability in visual field indices, namely mean deviation (MD) and pattern standard deviation (PSD), is also reported. A schematic diagram showing the outputs of substudy B is given in *Figure 3*. The findings are reported in *Chapter 6*.
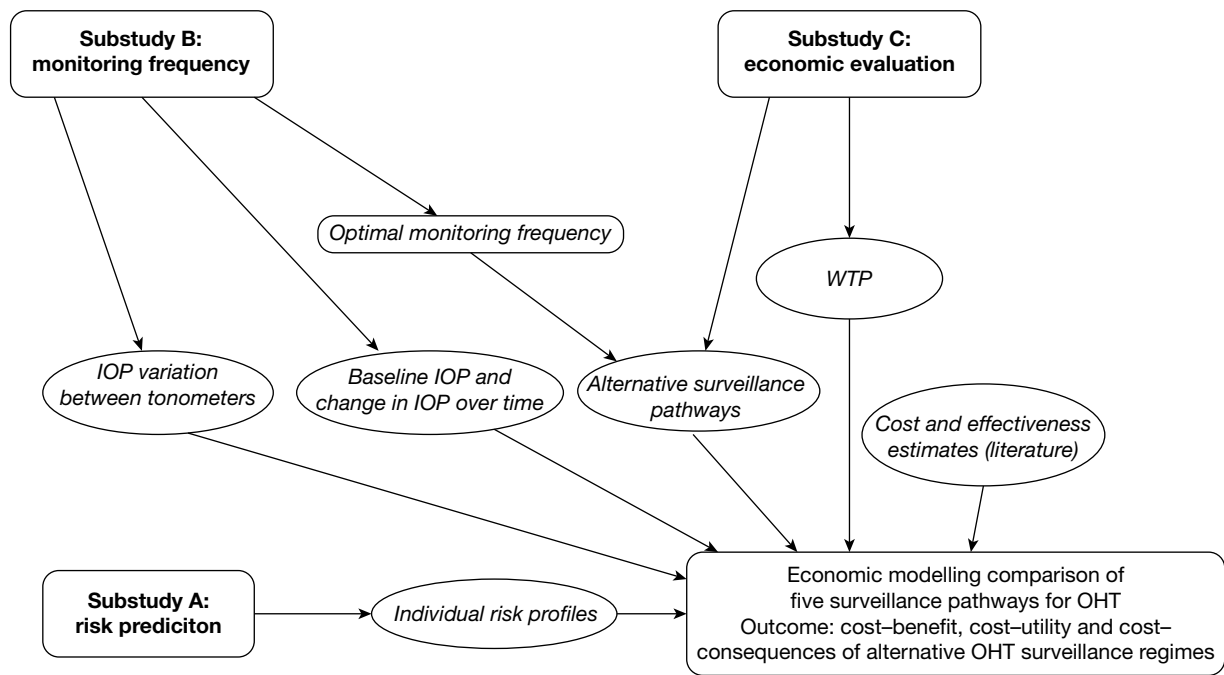
FIGURE 1 Overview of the study process.

FIGURE 2 Schematic diagram showing the outputs of substudy A. VF, visual field.

C: economic modelling evaluation of alternative surveillance regimes

The economic modelling evaluation includes three related components:

1. determining the surveillance pathways for the economic model (see *Chapters 7* and *8*)
2. a systematic review of economic evaluations of surveillance regimes for OHT (see *Chapter 8*)
3. an individual sampling economic model (a discrete event simulation) of the alternative surveillance pathways, using the results of the systematic reviews and primary data sets, to determine the effectiveness, cost-effectiveness, cost–utility and cost–benefit of differing surveillance regimes for people with OHT (see *Chapters 7* and *8*).

A schematic diagram showing the outputs of substudy C is given in *Figure 4*.

Substudy B: monitoring tests and optimal frequency of measurement
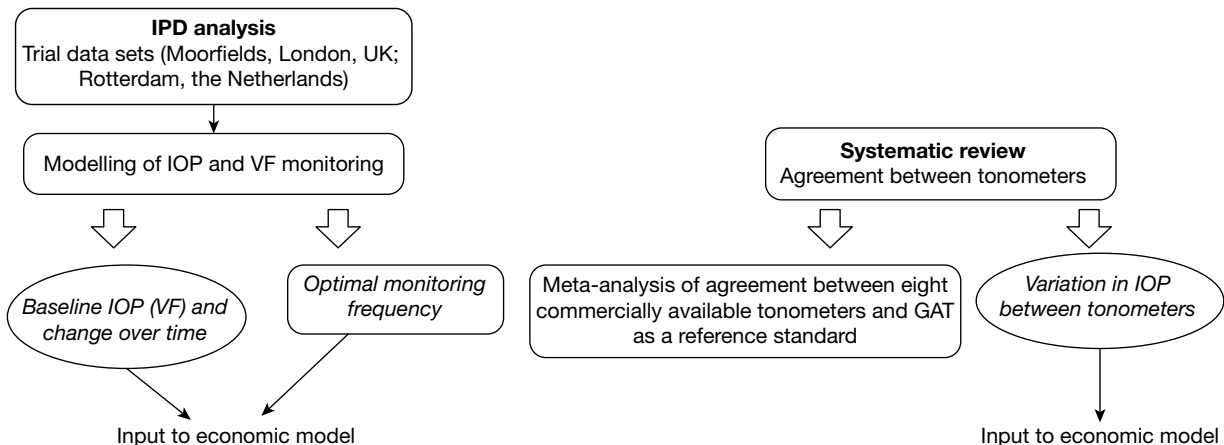
FIGURE 3 Schematic diagram showing the outputs of substudy B.

Substudy C: economic evaluation
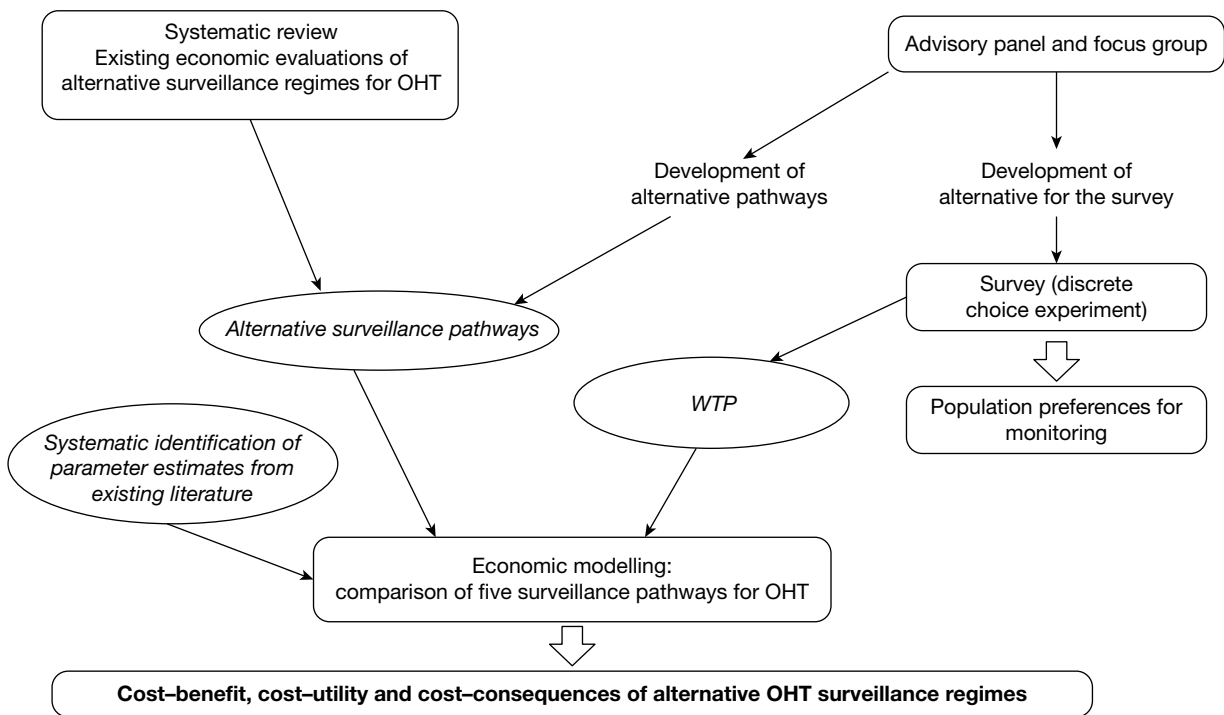
FIGURE 4 Schematic diagram showing the outputs of substudy C. VF, visual field.

Chapter 3

Materials

Overview of searching for the evidence

Electronic searches were conducted to identify reports of published, unpublished and ongoing studies. The search strategies were designed to be highly sensitive, including appropriate subject heading and text word terms that reflect both the clinical content and the type of study required for each component of the research project. This involved:

- identification of studies that report on risk prediction tools (see *Chapter 4*)
- identification of studies that report on the reliability of the selected IOP measurement techniques (see *Chapter 5*)
- identification of economic evaluations of surveillance programmes for OHT (see *Chapter 8*)
- focused searches to identify parameter values for the economic model (see *Chapter 8*).

Full details of databases searched and the search strategies used are given in *Appendix 1*.

Description of individual patient data sets

Before the start of the project we consulted the literature, approached the investigators of the OHT treatment trials and searched for any available existing IPD sets of OHT in primary or secondary care in the UK. Data sets needed to be accessible in terms of electronic storage of the data and agreement from the data holders to collaborate.

The following four data sets met these criteria and provided non-identifiable IPD. The data sets include data from two RCTs (Moorfields Eye Hospital, London, UK, and Rotterdam Eye Hospital, the Netherlands) and data from two observational cohorts – one hospital based (Dunfermline, UK) and one based on a community optometry-led monitoring scheme (Nottingham, UK). The four data sets were used for the validation of the best risk prediction tool (see *Chapter 4*) and the two RCT data sets provided data to determine the optimal frequency for measuring IOP and visual fields to identify true change (see *Chapter 6*).

A summary of the available data in these data sets is provided in *Table 1*. Overviews of the pathways for inclusion of participants in the analyses are provided in *Figures 5–8*. Specific details of each data set are provided in the following text.

Data from randomised controlled trials
Moorfields Eye Hospital
The Moorfields Eye Hospital data set comprises data from a RCT of medical treatment with betaxolol eye drops (FDC International Ltd, Fareham, UK) or placebo in OHT.[32] In the original study 356 patients with OHT were randomised during the period 1992–6.

Eligible participants for the trial included those:

- aged > 35 years with an IOP, by GAT, between 22 mmHg and 35 mmHg and

TABLE 1 Individual patient data sets for analysis

Source	Patient group	Study design	Median length of follow-up (years)
Moorfields Eye Hospital, London, UK	356 participants with OHT randomised; 300 with available data for this study	RCT of treatment with betaxolol 0.5% eye drops or placebo	9.3
Rotterdam Eye Hospital, Rotterdam, the Netherlands	412 participants with OHT randomised; 396 with available data for this study	RCT of treatment with betaxolol 0.25% eye drops, timolol 0.5% eye drops (FDC International Ltd, Fareham, UK) or placebo	8.2
Queen Margaret Hospital, Dunfermline, UK	304 people with OHT 188 with available data for analysis	Registry with hospital-based monitoring	2.7
Queens Medical Centre, Nottingham, UK	185 people with OHT 159 with available data for analysis	Registry with community-based monitoring	4.3

- with no significant ocular or neurological pathology that would affect visual field performance and
- with visual acuity of 6/12 or better and
- with normal visual fields on Humphrey SAP defined as a mean Advanced Glaucoma Intervention Study (AGIS)[33] score on two baseline fields of 0.

There were criteria violations: three participants had an IOP > 35 mmHg on the day of treatment commencement (36, 37 and 40 mmHg) and there were two participants aged < 35 years. Data from these participants were included in the analysis.

Participants were followed up every 4 months with Humphrey SAP, IOP measurement and optic nerve imaging in the study eye for a minimum of 2 years. Treatment status was noted at each visit. Conversion to glaucoma was based on predefined criteria (*Box 1*). Any participants reaching the conversion end point had IOP lowered by a topical ocular hypotensive other than betaxolol.

For the purpose of the secondary analysis of the trial data set for this study, the variables required were not available for 56 of the 356 trial participants and thus 300 participants are included in the analysis. The Moorfields Eye Hospital data set provides valuable RCT data and, additionally, post-trial surveillance with repeated measures of visual field using SAP. A flow diagram to illustrate the selection of patients to be included in the analysis for the Moorfields cohort is provided in *Figure 5*.

BOX 1 Moorfields Eye Hospital trial criteria for conversion to glaucoma

Conversion was defined as the change from an initial AGIS score of 0 to an AGIS score of ≥ 1 on three consecutive reliable visual fields, with at least one of the locations consistently below the threshold for normality. Criteria defining a reliable field were < 25% fixation losses, < 30% false-negative errors and < 30% false-positive errors. The study protocol dictated that, if a patient developed a visual field defect, the test was repeated within 1 month, and if the same defect was then reproduced on a reliable second field then a third test was performed 3–4 months after that. Conversion was confirmed if the field defect was present on the three consecutive tests.

```
                    ┌─────────────────────┐
                    │ 356 patients randomised │
                    └─────────────────────┘
                              │
  ┌─────────────────┐         │
  │ 56 patients excluded │────┤
  │ as unable            │    │
  │ to verify IOP data   │    │
  └─────────────────┘         │
                              │
                    ┌─────────────────────┐
                    │ 300 patients in data set │
                    └─────────────────────┘
                              │
  ┌─────────────────┐         │      ┌─────────────────┐
  │ 1 patient without │───────┼──────│ 1 patient with OAG │
  │ follow-up         │       │      │ at baseline        │
  └─────────────────┘         │      └─────────────────┘
                              │
                    ┌─────────────────────┐
                    │ 298 patients analysed │
                    └─────────────────────┘
                              │
  ┌─────────────────┐         │      ┌─────────────────┐
  │ 254 patients     │────────┴──────│ 44 patients       │
  │ without OAG      │               │ developed OAG     │
  └─────────────────┘               └─────────────────┘
```
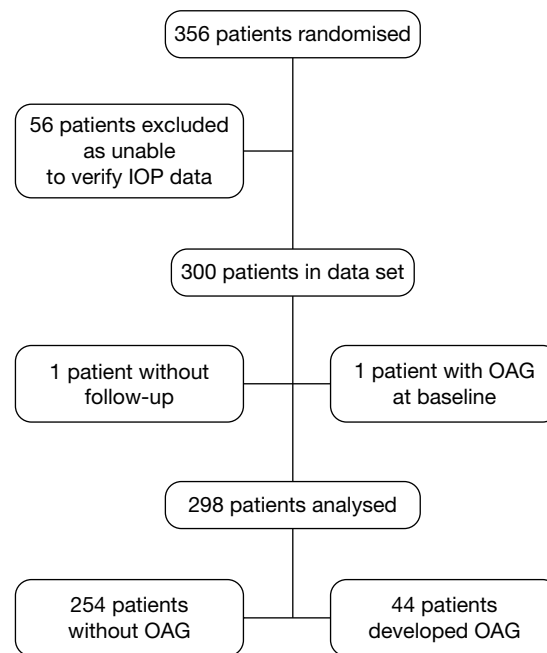
FIGURE 5 Flow diagram of patients from the Moorfields cohort included in the analyses.

Rotterdam Eye Hospital

The Rotterdam Eye Hospital data set comprises data from a placebo-controlled, double-masked, prospective, three-arm RCT (placebo, timolol 0.5% twice a day, betaxolol 0.25% twice a day) that was completed on 31 August 2008. This study randomised 412 participants with OHT between November 1997 and March 2001, including both eyes, with raised IOP ≥ 22 mmHg and ≤ 32 mmHg by GAT, and normal visual fields on Humphrey automated perimetry. Participants were deemed eligible if they were of white ethnic origin and had a best corrected Snellen visual acuity of at least 20/40. Participants with any significant coexisting ocular or systemic disease or any use of ocular hypotensives in the preceding 3 months were excluded. The follow-up entailed 6-monthly IOP measurement, visual field analysis and imaging of the optic disc using automated technology.[34] Conversion to glaucoma was based on predefined criteria (*Box 2*).

For the secondary analysis of the data for the purpose of this study, participants who had used eye drops for at least 6 months or who had not had baseline visual field testing on at least three occasions were not included in the data set. A total of 16 participants were excluded on this basis and thus data on 396 participants were potentially available for analysis from the study start date in November 1997.

On 31 August 2008, the end date for the study, data on 393 participants were available for analysis: two participants did not have follow-up data and were excluded and on reviewing the

BOX 2 Rotterdam Eye Hospital trial criteria for conversion to glaucoma

Conversion to glaucoma was defined as a reproducible defect in the visual field (SAP) of either one individual point below the 0.5% probability level, or two clustered points below the 1% probability level, or three clustered points below the 2% probability level, or four clustered points below the 5% probability level on either the total deviation or the pattern deviation probability plot.

baseline visual field data one participant was deemed to have early glaucomatous visual field defect at baseline and was excluded from the risk prediction analysis (see *Chapter 4*). A flow diagram to illustrate the selection of patients to be included in the analysis for the Rotterdam cohort is provided in *Figure 6*.

Data from observational data sets

Dunfermline

The Dunfermline data set was collated on an electronic patient record of all referrals with confirmed OHT from the year 2000 up until the end of December 2010. A flow diagram to illustrate the selection of patients to be included in the analysis for the Dunfermline cohort is provided in *Figure 7*.
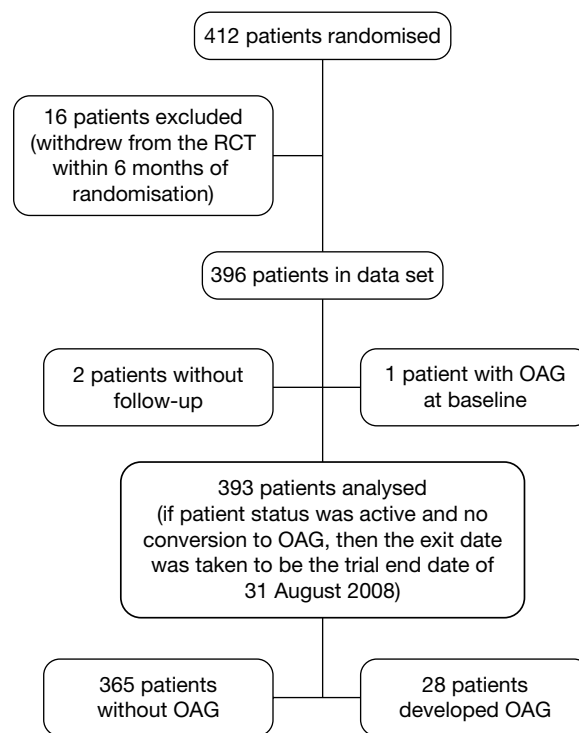
FIGURE 6 Flow diagram of patients from the Rotterdam cohort included in the analyses.
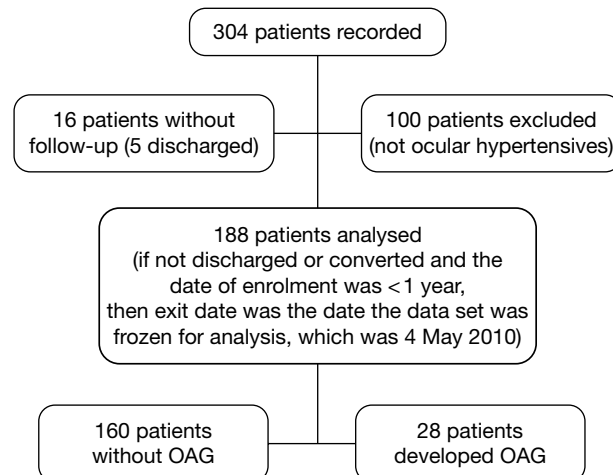
FIGURE 7 Flow diagram of patients from the Dunfermline cohort included in the analyses.

The Dunfermline guidelines for the management of their ocular hypertensive patients recommend that low-risk cases (see *Box 3* for definition) are discharged after two visits (at 1 year); those who develop glaucoma (*Box 4*) and those deemed as being at high risk of developing glaucoma remain under review.

Visual field data using Humphrey SAP were available for both eyes of all patients (except those unable to complete the test).

BOX 3 Dunfermline criteria for discharge from the scheme

Absolute features common to all deemed low risk:

- no field loss
- IOP <23 mmHg at discharge
- stable disc appearances.

Relative features:

1. negative family history of glaucoma
2. not on drugs (such as steroids) that may induce glaucoma.

BOX 4 Dunfermline definition of conversion to glaucoma

Conversion was based on the development of a repeatable visual field defect and/or a significant change in optic disc morphology. A visual field defect was defined as a reproducible defect in SAP of either one individual point below the 0.5% probability level, or two clustered points below the 1% probability level, or three clustered points below the 2% probability level, or four clustered points below the 5% probability level on either the total deviation or the pattern deviation probability plot. At least two sets of fields were required to deem conversion.

Nottingham

The Nottingham data set includes data from a cohort of 185 patients with OHT followed in a community scheme run by locally trained optometrists. The scheme started in 2003 and we included those who had at least 1 year of follow-up at the time of ascertaining the data set for this study in January 2010. Of the 185 patients recorded, 26 did not have follow-up data. Thus, data on 159 participants were analysed. Measurements on both eyes were available. A flow diagram to illustrate the selection of patients to be included in the analysis for the Nottingham cohort is shown in *Figure 8*.

Patients were eligible to enter the scheme if they were being followed up for OHT in hospital-based clinics and had normal visual fields and normal optic discs and an IOP recorded as >21 mmHg on at least one occasion. Patients had to be able to undergo reliable visual field testing and have a visual acuity better than 6/12, and patients with both treated (single drop only) and untreated OHT were eligible. At baseline a visual field test was performed (24-2 SITA Fast) and stereoscopic optic disc imaging was performed as a baseline assessment of optic disc appearance/ morphology. The community optometrists were provided with stereo viewers and stereo photographs of the patients' optic disc appearance at entry for comparison throughout follow-up. In addition, the referral guidelines to the community-based service indicated what cut-off of IOP should prompt referral back to the hospital eye service. Patients were reviewed once a year by visual field tests, GAT and clinical disc examination. Treatment state was recorded at each visit. Patients were referred back to the hospital eye service if the optometrist undertaking the annual assessment judged the optic disc appearance to have changed (comparing it with the baseline

appearance) or noted a visual field defect on visual field testing or if the patient was unable to undergo reliable visual field testing or the IOP had become elevated above the cut-off indicated in the referral template. All patients were referred back to the hospital eye service at 6 years for assessment by a consultant ophthalmologist. Patients were then returned to the community-based service to continue their annual follow-up.

Criteria for discharge from the monitoring scheme and the definition for conversion to glaucoma are described in *Boxes 5* and *6*.
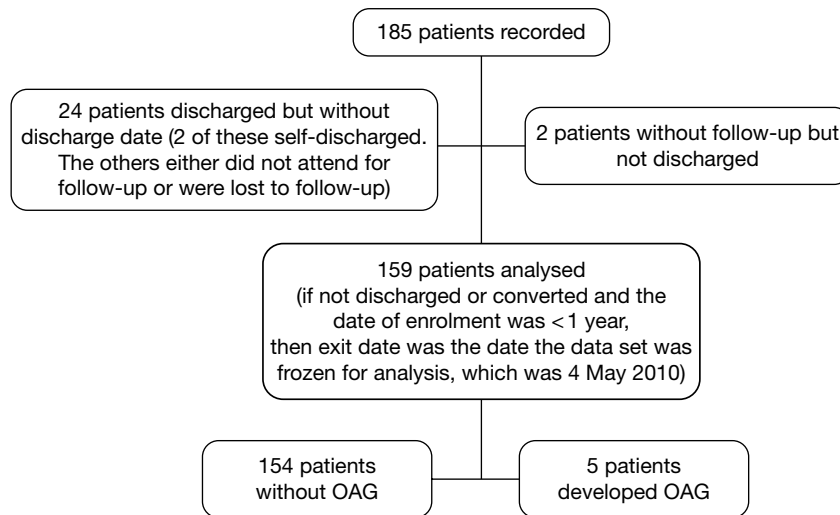
FIGURE 8 Flow diagram of patients from the Nottingham cohort included in the analyses.

BOX 5 Nottingham criteria for discharge

Generally patients were discharged if they had untreated IOP consistently < 21 mmHg and a normal visual field and optic discs.

BOX 6 Nottingham definition of conversion to glaucoma

Conversion was based on the development of a repeatable visual field defect and/or a significant change in optic disc morphology. This was initially detected by the optometrist on their annual review and confirmed with a repeat visual field on return to the hospital eye service for consultant assessment.

Chapter 4

Risk prediction tools for development of open-angle glaucoma

Introduction

The aetiology of OAG is not well understood, and elevated IOP, a well-known risk factor for the development of OAG, cannot be prevented. However, IOP can be lowered by medication, laser or surgery. Treatment decisions such as initiation and extent of therapy rely on many factors including the patient's risk of developing OAG and life expectancy. Risk factors for conversion from OHT to OAG have been investigated in several longitudinal population-based[35–39] and randomised[40–44] studies. Although both increased age and elevated IOP have been consistently shown to be important factors,[45] patients with IOP in the normal range (10–21 mmHg) can develop glaucoma. Consequently, multifactorial risk prediction models have been developed in an attempt to quantify the risk of developing the disease.

Validated risk calculators have become a useful tool in risk assessment for coronary heart disease (CHD)[46,47] and parallels have been drawn between CHD and glaucoma; both are chronic diseases with known modifiable risk factors.[48,49] Ideally, internal and external validation of a prediction model should be performed. For internal validation of performance, only patients from the model's derivation set are used by applying bootstrapping, cross-validation or split-sampling techniques.[50] However, prediction models often do not generalise beyond the population used in model derivation, and so it is widely accepted that a prediction model should not be applied in clinical practice before it has been validated in at least one other population and preferably by different investigators.[51,52] Studies with long-term follow-up of OHT patients with varying characteristics can be used to assess transferability and generalisability of an OAG prediction model.

Aims

1. To identify prediction models for development of OAG that include IOP as a predictor.
2. To critically appraise the construction and validation of the models.
3. To compare the performance of the models in four existing data sets.

Methods

Search strategy

Sensitive electronic searches were conducted to identify reports on the development and validation of risk prediction models for patients with OHT. Databases were searched from 1987 until January 2011 with no language restriction. Conference proceedings were not included. The following bibliographic databases were searched: MEDLINE, MEDLINE-In Process & Other Non-Indexed Citations, EMBASE, Science Citation Index and BIOSIS. In addition, full-text searches of key journals from 2004 onwards were undertaken for relevant published and in-press publications, including *American Journal of Ophthalmology*, *Archives of Ophthalmology*, *British*

Journal of Ophthalmology, *Eye*, *Graefe's Archive for Clinical and Experimental Ophthalmology*, *Investigative Ophthalmology and Visual Science*, *Journal of Glaucoma* and *Ophthalmology*.

Additional searches were undertaken in the Health Technology Assessment (HTA) database and the Database of Abstracts of Reviews of Effects (DARE) for relevant evidence synthesis reports. An internet search using Copernic Agent was also undertaken and included key professional organisations.

Full details of the search strategies used are provided in *Appendix 1*.

Four authors independently reviewed all titles and abstracts to identify studies that should be read in full to determine their eligibility for inclusion. The reference lists of all included studies were also scanned for additional reports. Disagreements were discussed and resolved by consensus.

Study selection

Prospective studies and studies in which patients were retrospectively identified but prospectively followed up were included if:

- only patients with OHT were recruited or they could be identified from the rest of the study cohort and models were fitted separately for them
- they were conducted post 1987, when reliable computerised perimetry became the standard of care
- a prediction equation for the development of OAG could be obtained
- the reported model included at least two variables, one of which was IOP
- the performance of the model was reported in any data set (derivation or validation) of longitudinal follow-up of a cohort initially free of OAG irrespective of the length of follow-up.

Population

Adults with OHT (defined as elevated IOP but no evidence of glaucomatous optic nerve damage or visual field loss) aged ≥ 18 years.[53]

Quality assessment

The quality of included studies was assessed using a checklist that included assessment of the definition of OAG, the method of measurement of candidate predictors and how continuous predictors were used in the models. Categorising a continuous predictor can lead to loss of information, but sometimes it may be necessary to categorise in order to assess the variation of risk across the range of values of a predictor. Such categorisation should be prespecified and not data driven. The definition of OAG was acceptable if the diagnosis was made using visual field testing with automated perimetry and optic nerve head examination. The acceptable method of visual field measurement was automated white-on-white perimetry. GAT, the clinical reference standard for IOP measurement, was the only acceptable method of IOP measurement. We also considered the number of conversions per variable in the multivariate analysis because estimates may not be accurate or precise if there are few conversions per predictor. We assessed this criterion based on the rule of thumb that logistic and Cox models should be used with a minimum of 10 outcome events per predictor variable.[54,55]

Data extraction

We extracted the following data from each eligible study:

- demographic and clinical characteristics of the derivation and/or validation population such as age, gender, race and eye-specific measures – IOP, central corneal thickness (CCT), PSD and vertical cup-to-disc (VCD) ratio
- features of the model development process such as method of analysis and variable selection
- coefficients or hazard ratios (HRs) of predictors and prediction equations
- measures of model performance (*c*-index, calibration chi-squared text and calibration slope).

Statistical analysis

Data synthesis

Meta-analysis of each predictor was not performed because the effects of other predictors in a multivariate model are not accounted for and so this would be misleading. Each model was reviewed and summarised separately.

Model validation

Baseline data from the four cohorts described in *Chapter 3* were used to validate each included prediction model. However, for the Rotterdam cohort, with the exception of VCD ratio measurements, which were taken 2–3 years after randomisation, we used ocular measurements taken at the first follow-up visit (usually 1 month after randomisation) because of learning curve errors in the visual fields taken at baseline. Because the vertical VCD ratio is used in the assessment of OAG, if the time interval between measurement of the vertical VCD ratio and conversion to OAG was < 12 months, the measurements were not used and were considered missing. Sensitivity analysis was performed to assess the effect of using a time interval of < 24 months. The four data sets were analysed individually to enable assessment of the transferability of models between different geographical locations and settings.

The missing data mechanism was investigated in each cohort. To make the best use of the available data and to avoid bias, missing values were imputed using multivariate imputation by chained equations.[56] By replacing missing values with plausible values based on the distribution of the observed data, 10 imputed data sets were created based on the predictors and outcome variables (Nelson–Aalen estimator for time to glaucoma and the censoring indicator). To avoid excluding a completely missing predictor from the risk estimation, we imputed the average value from a similar data set. Visual inspection of the distribution of imputed values was used to assess comparability with the observed data. For comparison with analyses based on imputed data, complete case analyses were also performed.

Ideally, model validation should be performed using an untreated population but this approach would have more than halved the sample size of some cohorts. Thus, we performed Cox regression analysis with and without stratification by treatment on each imputed data set. Estimates of the coefficient of each predictor were averaged across the 10 data sets to obtain a single estimate. Standard errors were computed based on Rubin's rules,[57] which take into account the between- and within-imputation components of variation in the parameter estimates.

The 5-year risk of developing glaucoma for each patient was calculated using each included prediction equation and the patient's observed or imputed value of the predictors. The predicted 5-year risk of glaucoma was grouped into three categories, low (< 6%), intermediate (6–13%) and high (> 13%), as defined by the OHTS.[13] Five-year Kaplan–Meier failure curves for each cohort were plotted according to this risk classification and log-rank tests performed to assess equality of the survivor function across groups (note that each plot and the associated log-rank test depicts a single imputed data set as an example).

The predictive ability of a model is frequently assessed in terms of discrimination and calibration.[58] The ability of the model to discriminate between patients who did or did not

develop OAG was assessed using Harrell's *c*-index. This measure is similar to the area under the receiver characteristic operating curve and is more appropriate for survival analysis.[59] A *c*-index of 1.0 indicates perfect discrimination (e.g. predicted risks for those with OAG are all greater than for those without OAG, with perfect separation of both groups) whereas a *c*-index of 0.5 indicates random discrimination. Approximate 95% CIs were computed for the *c*-index by applying Rubin's rules to jackknife standard errors obtained for the *c*-indexes of the 10 imputed data sets. Where the upper limit of the 95% CI exceeded 1.0 the value was truncated at 1.0.

Calibration refers to the agreement between predicted and observed risk. Model overfitting (where predictions in the model are more certain than justified because of the play of chance in the derivation data set) can be identified by assessing model calibration in independent validation data sets. Patients were divided into quintiles according to their predicted risk. Within each quintile, the average predicted risk was compared with the corresponding Kaplan–Meier estimate of the observed risk. Calibration plots were used to illustrate the fit of the models; for a perfectly calibrated model, all points will lie on the 45° line. Additionally, we used the slope of the prognostic index (linear predictor), known as the calibration slope, to quantify calibration. The slope of the prognostic index is the regression coefficient β in a Cox regression model with the prognostic index as the only covariate. The calibration slope should ideally be 1 when predicted risks agree completely with observed risks,[60] which will be the case if the effects of predictors in the validation data set are on average similar to those in the development data set (note that, although the tests and summary statistics are based on appropriate analyses of all 10 imputed data sets, the plot depicts a single imputed data set and thus may not directly depict the final results). Because pretreatment values of the baseline predictors were known in the trial-based cohorts, we produced calibration plots for treated and untreated patients separately to assess the extent to which calibration of the model was affected by treatment. Calibration and Kaplan–Meier plots were based on one of the imputed data sets for each cohort. All analyses were performed using Stata 11.1 (StataCorp LP, College Station, TX, USA).

Results

Systematic review

Study selection
The titles and abstracts of 565 articles were screened to identify eligible studies. Full-text papers of 54 of these were retrieved for more detailed evaluation of eligibility (*Figure 9*). Forty-nine papers were excluded and the reasons are provided in *Appendix 2*. Of the five included papers, four were based on the results of two RCTs, the OHTS[61] and the EGPS.[14,41] These provided three models for which prediction equations were available (full and reduced OHTS models[42,62,63] and the pooled OHTS-EGPS means model[64,65]). The fifth paper reported the independent validation study of the OHTS model in the Diagnostic Innovations in Glaucoma Study (DIGS) cohort.[63]

Assessment of study quality
Both the OHTS and EGPS were large prospective studies that included patients with OHT aged ≥ 30 years who had no evidence of glaucomatous damage at baseline. The OHTS randomised 1636 individuals, with an IOP between 24 mmHg and 32 mmHg in one eye and between 21 mmHg and 32 mmHg in the other eye, to treatment or observation. The EGPS randomised 1081 individuals with IOP ≥ 22 mmHg in at least one eye to treatment or placebo. The inclusion and exclusion criteria used in the DIGS cohort were very similar to those used by the OHTS. For a patient to be included, his or her baseline IOP had to be ≥ 24 mmHg in one eye and ≥ 21 mmHg in the other eye, on at least two occasions.
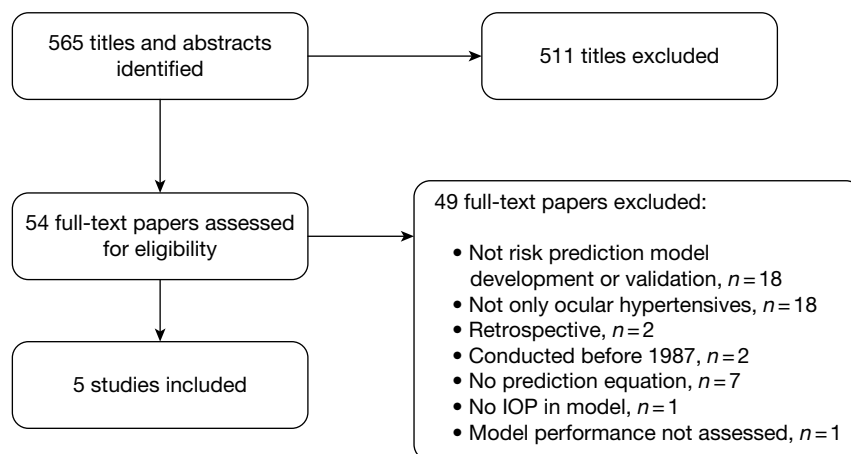
FIGURE 9 Flow chart of the study selection process.

In the OHTS, all visual fields were assessed using full-threshold white-on-white Humphrey program 30-2 perimetry, whereas in the EGPS this was used for 79.6% of the participants and Octopus 32-2 visual fields for the rest. Visual field assessment was based on three consecutive abnormal and reliable tests with defect in the same location and index. In the DIGS, visual field examination was performed with SAP. At baseline, SAP testing was performed using the program 24-2 full-threshold strategy, but during follow-up, SAP testing was performed using either the full-threshold or the Swedish Interactive Threshold Algorithm strategy.[63] Using different visual field programs based on different algorithms is not ideal, but in a clinical context of determining whether or not glaucoma has developed it is acceptable. In the OHTS, optic discs were assessed using two consecutive sets of photographs judged to have a clinically significant change, whereas the EGPS used a set of photographs judged to have changed by at least two of three masked readers. In both studies, detection of OAG was by a masked end-point committee and was based on reproducible visual field abnormalities or optic disc deterioration. In the DIGS, structural damage to the optic disc at baseline was based on assessment of simultaneous stereoscopic optic disc photographs by two masked readers; conversion from OHT to OAG was based on the development of a reproducible visual field defect or glaucomatous change in the appearance of the optic disc in at least one eye. The original protocols of the OHTS and EGPS did not include CCT and measurements were taken later, 2–3 years after randomisation of the last patient enrolled in the studies. All patients in the DIGS had CCT measurements taken during follow-up.

Cox regression analysis was used to obtain all of the prediction models and all continuous predictors were modelled without categorisation. The OHTS identified candidate variables during the planning phase of the study. For the pooled OHTS-EGPS model, variables in univariate models in the OHTS or the EGPS with $p < 0.10$ were the candidate variables. Statistically significant variables in the pooled univariate analyses were included in the pooled multivariate analyses. Measures of model performance in the development set were not reported for the OHTS models.

Assessment of a patient's risk requires the baseline hazard/survivor function in addition to the prognostic index. In the validation of the OHTS models, the baseline survivor function was approximated by the average survival probability at 5 years obtained from the Kaplan–Meier estimates because the function was not published by the OHTS.[63] The OHTS reduced model was published in a brief report,[62] and the same population and analyses described for the full model applies. The methodological quality of the two studies[42,64] that fully described the development of the OHTS full model and the OHTS-EGPS means model, respectively, and the validation

study in the DIGS cohort[63] is summarised in *Table 2*. Unlike the OHTS models, the OHTS-EGPS model has not been independently validated. The model derived from the observation group of the OHTS was validated in the EGPS placebo group, but the pooled OHTS-EGPS model derived from both populations was not externally validated.

Prediction models

In univariate analyses of the OHTS or pooled OHTS and EGPS data, statistically significant predictors for development of OAG were age, IOP, CCT, VCD ratio, horizontal cup-to-disc (C/D) ratio, PSD, history of heart disease, gender, race and diabetes mellitus. In both OHTS and EGPS, history of diabetes and heart disease were self-reported and not clinically verified. *Table 3* summarises the models including the reduced OHTS model and validation of the full and reduced OHTS models in the DIGS cohort. The factors common to all models in multivariate analyses were IOP, age and CCT.

Ocular Hypertension Treatment Study prediction model

The OHTS multivariate Cox regression analysis used the means of both eyes for eye-specific variables and identified six factors that were significantly associated with the risk of conversion from OHT to OAG: age, IOP, CCT, VCD ratio, PSD and diabetes mellitus.[42] Horizontal C/D ratio was also significantly associated with glaucoma development in multivariate analysis but, because of its high correlation with the VCD ratio, only the VCD ratio was included in the final multivariate model. Also, the VCD ratio had slightly better predictive ability. Analyses were performed using only the untreated group as well as the combined treated and untreated groups. Similar results were obtained and only the model based on the entire sample was reported.

Changes in PSD and VCD ratio may be indicative of early OAG damage and their inclusion in the model may result in an overestimation of risk. Thus, to avoid inclusion of variables used in the definition of the disease, another model, the reduced OHTS model, was developed.[62] HRs from the multivariate Cox models with and without PSD and VCD ratio were found to be similar and in the reduced model all factors remained statistically significant.

Using 252 eyes from 126 untreated OHT patients retrospectively selected from DIGS, a prospective longitudinal study, Medeiros and colleagues[63] validated the OHTS prediction model.

TABLE 2 Summary of methodological quality

Quality item	OHTS[42]	OHTS-EGPS[64]	DIGS[63]
1. Was the sample of patients representative (i.e. consecutively or randomly sampled)?	Yes	Yes	Yes
2. Were groups of interest included?	Yes	Yes	Yes
3. Was the definition of OAG appropriate?	Yes	Yes	Yes
4. Was the method of measurement of visual field parameters acceptable?	Yes	Yes	Yes
5. Was the method of measurement of IOP acceptable?	Yes	Yes	Yes
6. Was the duration of follow-up adequate (at least 5 years)?	Yes	Yes	Yes
7. Was the number of conversions adequate for the number of predictors in the model?	Yes	Yes	NA
8. Did the model include IOP as well as other known risk factors?	Yes	Yes	NA
9. Was categorisation of continuous variables prespecified and rationalised to avoid data-dependent analysis?	NA	NA	NA
10. Was the model validated in a data set different to the derivation data set (a, b or c)?	Yes	No	NA
(a) A different sample, source and investigator (this provides best evidence)	Yes		
(b) A different sample from a different source			
(c) A different sample but from the same source as the data used to derive it			

NA, not applicable.

TABLE 3 Multivariate HRs and performance measures for the risk prediction models

Characteristic	Full OHTS model (development set)	Full OHTS model (DIGS validation set)	Reduced OHTS model (development set)	Reduced OHTS model (DIGS validation set)	OHTS-EGPS model (development set)
n/N^a	125/1618	31/126	125/1618	31/126	154/1123
Baseline predictor, HR (95% CI)					
Age (decade)	1.25 (1.04 to 1.49)	1.49 (1.06 to 2.11)	1.29 (1.09 to 1.53)	1.56 (1.13 to 2.16)	1.26 (1.06 to 1.50)
IOP (mmHg)	1.11 (1.03 to 1.17)	1.17 (1.05 to 1.30)	1.10 (1.04 to 1.17)	1.16 (1.04 to 1.29)	1.09 (1.03 to 1.17)
CCT (per 40 μm thinner)	1.82 (1.51 to 2.19)	1.92 (1.25 to 2.96)	1.92 (1.60 to 2.30)	2.10 (1.39 to 3.18)	2.04 (1.70 to 2.45)
History of diabetes mellitus	0.35 (0.15 to 0.78)	1.13 (0.40 to 3.18)	0.38 (0.17 to 0.86)	1.28 (0.47 to 3.46)	Excluded
VCD ratio (per 0.1 larger)	1.32 (1.20 to 1.45)	1.10 (0.85 to 1.44)	Excluded	Excluded	1.19 (1.09 to 1.31)
PSD (per 0.2 dB greater)	1.25 (1.06 to 1.48)	1.15 (0.95 to 1.39)	Excluded	Excluded	1.13 (1.04 to 1.24)
Performance measure					
c-index	NR	0.68	NR	0.73	0.74 (0.70 to 0.78)
Calibration χ^2	NR	NE	NR	NE	7.05
Calibration slope	NR	0.826 (SE 0.195)	NR	1.086 (SE 0.255)	NE

NE, not estimated; NR, not reported; SE, standard error of the calibration slope.
a n is the number of OAG cases and N is the total number of patients. It was not always apparent whether these were the numbers analysed in multivariate Cox regression analysis though there was no report of imputation for missing values. CCT measurements were only available for 119 cases and 1279 non-cases in OHTS. The numbers shown for the OHTS-EGPS model were the numbers analysed.

Fifteen (12%) of the 126 patients received treatment during follow-up, and for these patients only the pretreatment period was evaluated. To derive the OHTS prediction equation, the average survival probability at 5 years was approximated using the Kaplan–Meier estimates because the baseline survivorship function was not included in the published OHTS results. For the full OHTS model (containing all six predictor variables), the risk estimate for glaucoma development in 5 years was calculated as:

$$\text{Risk estimate} = 1 - 0.906\exp(\text{PI}) \qquad\qquad [\text{Model 1}]$$

where 0.906 is the average survival probability at 5 years and

$$\text{PI} = \sum_{i=1}^{k} \beta_i \left(X_i - \overline{x}_i \right) = \ln(1.25)\times(\text{AGEDECADE} - 5.54) + \ln(1.11)\times(\text{IOP} - 24.9)$$
$$+ \ln(1.82)\times(\text{T\_CCT} + 14.3625) + \ln(1.25)\times(\text{T\_PSD} - 9.5)$$
$$+ \ln(1.32)\times(\text{T\_VCD} - 3.9) + \ln(0.35)\times(\text{DIABETES} - 0.121) \quad [\text{Equation 1}]$$

where β_i is the regression coefficient in the Cox proportional hazards model for the ith predictor and is the log HR, X_i is an individual's value for one of the k predictors and $\overline{x}_i$ to $\overline{x}_k$ are the mean values from the OHTS cohort for the k predictors. Four of the five predictors were transformed as follows:

$$\text{AGEDECADE} = \text{AGE}/10 \qquad\qquad [\text{Equation 2}]$$

$$\text{T\_CCT} = -1\times\text{CCT}/40 \qquad\qquad [\text{Equation 3}]$$

$$\text{T\_VCD} = \text{VCD}/0.1 \qquad\qquad [\text{Equation 4}]$$

$$\text{T\_PSD} = \text{PSD}/0.2 \qquad\qquad [\text{Equation 5}]$$

The 'DIABETES' variable takes the value 1 if the patient has a history of diabetes mellitus and 0 otherwise. VCD is the vertical C/D ratio. A risk scoring system based on this model was also developed.

For the reduced model, excluding the VCD ratio and PSD, the risk estimate was calculated as:

$$\text{Risk estimate} = 1 - 0.906\exp(\text{PI}) \qquad \text{[Model 2]}$$

where

$$\text{PI} = \sum_{i=1}^{k} \beta_i \left(X_i - \overline{x}_i \right) = \ln(1.29) \times (\text{AGEDECADE} - 5.54) + \ln(1.10) \times (\text{IOP} - 24.9)$$
$$+ \ln(1.92) \times (\text{T\_CCT} + 14.3625) + \ln(0.38)$$
$$\times (\text{DIABETES} - 0.121) \qquad \text{[Equation 6]}$$

The full model containing the VCD ratio and PSD performed similarly to the reduced model that excluded these variables. In updated univariate and multivariate analyses of the OHTS model, the protective effect of a history of diabetes mellitus was no longer statistically significant.[66] The authors suggest that the difference from the 2002 article was a reflection of more complete ascertainment of diabetes mellitus. Also, individuals with diabetic retinopathy were excluded from the OHTS and so the patients in the study may not be representative of patients with diabetes.

Ocular Hypertension Treatment Study-European Glaucoma Prevention Study prediction model

A prediction model was developed from the OHTS observation group and validated in the EGPS placebo group.[64] Unlike the OHTS, the EGPS included patients with pigment dispersion syndrome or pseudoexfoliation syndrome (PEX), conditions that may lead to secondary OAG. However, the 19 patients with either condition were excluded from the analyses. In both the OHTS and the EGPS, interactions were not detected between predictors.[64] The final model was derived from both groups combined. The OHTS-EGPS model estimates the 5-year risk of developing glaucoma based on age, IOP, CCT, VCD ratio and PSD. This model uses the means of the right and left eyes of each participant to calculate eye-specific predictors. Similar results were obtained when the pooled analysis was performed with and without the EGPS participants with only one eye eligible for the study and so they were all included in the analyses.

The OHTS and EGPS groups further assessed whether the accuracy and clinical application of the OHTS-EGPS model based on the means of the right and left eyes, the 'means model', are equal, superior or inferior to those of prediction models that include eye-specific information.[65] These models include the 'means plus asymmetry' model, which includes age and the means of the right and left eyes as well as the absolute difference between eyes for eye-specific variables, and the 'worse' eye model, which includes age and values from the eye at higher risk for developing OAG. The 'worse' eye was selected using the OHTS-EGPS means model, which was developed from the same sample as the 'worse' eye model, thus introducing statistical circularity and increasing the predictive accuracy of the 'worse' eye model. The prediction model that uses the means of both eyes for eye-specific variables was reported as the simplest to use and the most robust to measurement variability and error.[65] Using the same transformation as in the OHTS model for age, CCT, PSD and VCD, the risk estimate for glaucoma development in 5 years was calculated as:

$$\text{Risk estimate} = 1 - 0.91831\exp(\text{PI}) \qquad \text{[Model 3]}$$

where

$$\mathrm{PI} = \sum_{i=1}^{k} \beta_i \left(X_i - \overline{x}_i \right) = 0.23260 \times (\mathrm{AGEDECADE} - 5.64301) + 0.09025 \times (\mathrm{IOP} - 24.1386)$$
$$+ 0.71503 \times (\mathrm{T\_CCT} + 14.3349) + 0.12376 \times (\mathrm{T\_PSD} - 9.76001)$$
$$+ 0.17689 \times (\mathrm{T\_VCD} - 3.60828) \qquad \text{[Equation 7]}$$

A calculator to estimate the 5-year risk of developing OAG, based on the pooled OHTS–EGPS means model, is available online at http://ohts.wustl.edu/risk/calculator.html. The calculator truncates predictor values at the bounds of the values recorded for each predictor in the OHTS-EGPS cohort (see *Table 8* for the lower and upper limit values of each predictor). To simplify risk assessment, a points-based system was also developed by the OHTS-EGPS group.

Model validation

Because the inclusion of a history of diabetes mellitus as a predictor in the OHTS models (model 1 and model 2) was no longer valid and the models have been superseded by the OHTS-EGPS model (model 3), only model 3 was validated using the four cohorts. The OHTS-EGPS prediction model used the average values of eye-specific predictors and so when values were available for both eyes we adopted the same approach in our validation. Only patients with an exit date, that is a conversion date, discharge date, date of death or at least one follow-up date, were included in the analyses. For those considered as active in the Rotterdam cohort but without a known follow-up date, the trial end date of 31 August 2008 was taken to be their exit date. Patients in the Dunfermline cohort who were not due for their first follow-up were assigned the date that the data sets were frozen for analyses (4 May 2010). Analyses were based on 393, 298, 188 and 159 patients for the Rotterdam, Moorfields, Dunfermline and Nottingham cohorts, respectively, as described in *Chapter 3*.

Missing predictors

All cohorts provided complete data on age. IOP was missing for a few patients, especially for those recently enrolled in the Dunfermline cohort. The frequency of missing data for each predictor in each cohort is shown in *Table 4*. The proportion of missing values for CCT was high in all cohorts (between 22.6% and 100%). CCT was sporadically collected for both the Moorfields and Dunfermline cohorts. However, in the Dunfermline cohort, CCT was not recorded in the electronic patient record system. To retrieve this information would have required manual retrieval of all case notes, an immense task with possibly very little yield. Therefore, we imputed the average value, 556 µm, from the Nottingham cohort to allow for calculation of each patient's risk.

For the Rotterdam cohort, VCD ratio was measured from November 2000 onwards only and CCT from September 2003 onwards. Patients who dropped out or were excluded prior to these dates did not have either a VCD ratio (32.8%) or a CCT measurement (22.6%). For two patients, the time interval between measurement of VCD ratio and conversion to OAG was < 1 year and so their measurements were regarded as missing. Of the 393 patients in the cohort, there were 64 patients (16%) with at least two missing predictors and in all 64 patients both CCT and VCD ratio were missing. For the Moorfields cohort, there were 87 patients (29.2%) with multiple missing predictors, of whom both CCT and PSD were missing in 74. In more than half of the Moorfields cohort, PSD was missing because corrected PSD (CPSD) was recorded more frequently, with this being missing in only 24 (8.1%) patients. In contrast to the Rotterdam and Moorfields cohorts, the Dunfermline and Nottingham cohorts included very few patients with multiple missing predictors – 14 (7.4%) and eight (5.0%), respectively.

TABLE 4 Frequency of missing data for the five predictors in each cohort

Characteristic	Rotterdam, *n* (%)	Moorfields, *n* (%)	Dunfermline, *n* (%)	Nottingham, *n* (%)
n[a]	393	298	188	159
Complete cases	238 (60.6)	78 (26.2)	135 (71.8)	112 (70.4)
Age	0	0	0	0
IOP	2 (0.5)	0	12 (6.4)	2 (1.3)
CCT	89 (22.6)	143 (48.0)	188 (100)	37 (23.3)
VCD ratio	129 (32.8)	13 (4.4)	17 (9.0)	10 (6.3)
PSD	1 (0.3)	155 (52.0)	46 (24.5)	10 (6.3)

a   *n* is the number of completed cases from imputed data and is the number of patients analysed.

Validation sample

The median follow-up times of our observational cohorts (2.7 years for Dunfermline and 4.3 years for Nottingham) were much shorter than those of our trial-based cohorts (8.2 years for Rotterdam and 9.3 years for Moorfields). The median follow-up time was 4.8 years and 6.6 years for the EGPS and the OHTS, respectively. The Kaplan–Meier estimate of the 5-year cumulative probability of developing OAG was 9.3% in the OHTS observation group and 16.8% in the EGPS placebo group. These values were 4.0% (95% CI 2.3% to 6.9%) for Rotterdam, 11.3% (95% CI 7.8% to 16.4%) for Moorfields, 23.7% (95% CI 14.1% to 39.9%) for Dunfermline and 5.1% (95% CI 2.0% to 12.7%) for Nottingham. The baseline characteristics of the four cohorts were generally similar to those of the OHTS-EGPS cohort, although patients in our observational cohorts were on average older than those in the randomised cohorts (*Table 5*). On average, patients in the Moorfields cohort had a lower CCT than those in the OHTS-EGPS, Rotterdam or Nottingham cohorts. Average PSD in the group that converted to OAG and the group that did not were lower in the Rotterdam cohort than in the corresponding groups in other cohorts.

In univariate analysis, a history of cardiovascular disease was associated with OAG in the OHTS, the EGPS and the pooled OHTS-EGPS. This information was not recorded for the Rotterdam and Moorfields cohorts. Only a few of the Dunfermline cohort had a history of heart disease (10/188, 5.3%), and 5.0% of those who did not convert to OAG had heart disease, compared with 7.1% of those who converted. In the OHTS-EGPS, 7.8% of those who did not convert to OAG had heart disease, compared with 11.5% of those who converted. The proportions of patients for whom predictor values were outside the bounds of the OHTS-EGPS cohort is shown in *Table 6*. In 22.1% (66/298) of the Moorfields cohort, IOP was < 20 mmHg, and 12.0% (17/142) of the Dunfermline cohort had a PSD value > 3.

Model performance

Examination of the distribution of imputed values showed similarity with observed values and the imputed data were used to compute predicted risks. Using one of the 10 imputed data sets, an example of the distribution of predicted risks for those who developed and those who did not develop OAG is shown in *Figure 10*. Stratified (treated vs untreated) and unstratified analyses gave similar results, and so only the unstratified analysis is reported. *Figures 11–14* show the Kaplan–Meier curves for the risk of conversion from OHT to glaucoma by risk group from the same imputed data sets used for *Figure 10*. Based on analysis of all 10 imputed data sets for the trial-based cohorts, there was a statistically significant difference in survival across risk groups. Multivariate HRs (and 95% CIs) for the risk factors and the performance measures of the model applied to each cohort using imputed data and complete cases are shown in *Tables 7* and *8*, respectively. For the Rotterdam and Dunfermline cohorts, there was little or no difference in model performance using multiple imputed data instead of complete case analyses. The converse

TABLE 5 Baseline clinical and demographic characteristics of the OHTS–EGPS derivation cohort and the four validation cohorts

Characteristic	OHTS–EGPS control group		Rotterdam		Moorfields		Dunfermline		Nottingham	
	No OAG	OAG	No OAG	OAG	No OAG	OAG	No OAG	OAG	No OAG	OAG
Patients[a]	1147	165	365	28	254	44	160	28	154	5
Age (years)	56.1 (9.9)	59.3 (9.5)	56.0 (11.0)	60.4 (10.2)	59.3 (10.2)	60.5 (10.5)	62.9 (11.8)	62.2 (9.2)	60.9 (10.1)	64.0 (9.4)
IOP (mmHg)	24.0 (2.4)	24.9 (2.7)	25.3 (2.3)	27.5 (2.5)	22.4 (3.8)	22.8 (3.3)	24.4 (3.6)	24.6 (3.6)	22.3 (3.3)	21.1 (3.4)
CCT (μm)	577.0 (36.0)	550.7 (36.3)	574.9 (35.6)	553.0 (27.9)	558.8 (31.4)	539.2 (26.8)	NA	NA	572.6 (43.7)	538.1 (4.5)
VCD ratio (by contour)	0.35 (0.2)	0.43 (0.2)	0.37 (0.16)	0.55 (0.22)	0.48 (0.15)	0.50 (0.14)	0.42 (0.18)	0.50 (0.17)	0.41 (0.15)	0.48 (0.19)
PSD (dB)	1.94 (0.4)	2.01 (0.4)	1.72 (0.36)	1.89 (0.31)	2.09 (1.00)	2.26 (0.69)	2.01 (0.82)	2.16 (0.77)	1.72 (1.22)	2.11 (0.91)
CPSD (dB)	1.14 (0.5)	1.11 (0.5)	NA[b]	NA	1.56 (1.16)[b]	1.44 (1.00)	NA	NA	NA	NA
Male	504 (43.9)	80 (48.5)	173 (47.4)	14 (50.0)	149 (58.7)	25 (56.8)	89 (55.6)	16 (57.1)	76 (49.4)	0
Race[c]										
Caucasian	938 (82.4)	121 (73.8)	365 (100)	28 (100)	210 (82.7)	36 (81.8)	160 (100)	28 (100)	0	0
A-A	168 (14.8)	33 (20.1)	0	0	13 (5.1)	6 (13.6)	0	0	0	0
Asian	0	0	0	0	4 (1.6)	0	0	0	0	0
Other	32 (2.8)	10 (6.1)	0	0	2 (0.8)	2 (4.5)	0	0	0	0
Unknown	0	0	0	0	25 (9.8)	0	0	0	154 (100)	5 (100)
Diabetes (yes)	120 (10.5)	5 (3.0)	NA	NA	NA	NA	13 (8.1)	4 (14.3)	12 (7.8)	0
Heart disease (yes)	90 (7.8)	19 (11.5)	NA	NA	NA	NA	8 (5.0)	2 (7.1)	12 (7.8)	1 (20.0)
Treated (yes)[d]	0	0	244 (66.8)	17 (60.7)	126 (49.6)	20 (45.5)	3 (1.9)	0	46 (30.1)[e]	2 (40.0)

A-A, African ancestry; NA, not available.

Values are expressed as mean (SD) for continuous variables and as number (percentage) for categorical variables.

a　This is the number of patients available in each cohort. Owing to missing data (except for age), only a subset was used to compute the summary statistics. (see the OHTS–EGPS report[64] and *Table 3*).

b　CPSD was measured in 232 patients who did not develop OAG and in all 44 patients who developed OAG.

c　For OHTS–EGPS, we grouped the Native American, Hispanic and other categories as Other; black, non-Hispanic as A-A; and white non-Hispanic as Caucasian. In the Moorfields cohort, African Caribbean was classified as A-A. Two patients in the Other category in the Moorfields cohort who did not develop OAG were Lebanese and Sri Lankan; the two in this category in this cohort who did develop OAG were Greek Cypriot and Egyptian.

d　Randomised to or prescribed an IOP-lowering treatment.

e　Based on 153 patients.

TABLE 6 Percentage of values in each cohort outside the range of predictor values in the OHTS-EGPS cohort

	Rotterdam		Moorfields		Dunfermline		Nottingham	
	N	*n* (%)	*N*	*n* (%)	*N*	*n* (%)	*N*	*n* (%)
Age (years)	393		298		188		159	
<30		2 (0.5)		2 (0.7)		0		0
>80		2 (0.5)		4 (1.3)		10 (5.3)		3 (1.9)
IOP (mmHg)	391		298		176		157	
<20		3 (0.8)		66 (22.1)		11 (6.3)		33 (21.0)
>32		3 (0.8)		3 (1.0)		5 (2.8)		0
CCT (µm)	304		155		0		122	
<475		0		1 (0.6)		NA		2 (1.6)
>658		2 (0.7)		0		NA		2 (1.6)
VCD ratio (by contour)	264		285		171		149	
<0		0		0		0		0
>0.8		2 (0.8)		3 (1.1)		1 (0.6)		0
PSD (dB)	392		143		142		149	
<0.5		0		1 (0.7)		0		0
>3		3 (0.8)		10 (7.0)		17 (12.0)		6 (4.0)

NA, not available.
N is the number of non-missing values for the predictor; *n* is the number of values less than the lower bound value of the predictor or greater than the upper bound value of the predictor in the OHTS-EGPS cohort.

was the case for the Moorfields and Nottingham cohorts. Data were complete for only 78 patients (26%) in the Moorfields cohort.

The *c*-indexes were between 0.69 and 0.83; the Rotterdam cohort had the best discriminatory ability (0.83, 95% CI 0.75 to 0.91), whereas the Moorfields cohort yielded the lowest *c*-index (0.69, 95% CI 0.59 to 0.78). *Figures 15–22* illustrate calibration of the model in the four cohorts based on the same imputed data sets as in *Figures 10–14*. For the Rotterdam and Moorfields cohorts, the calibration plots are shown separately for the treated, untreated and combined data to highlight differences in calibration that may exist as a result of the effect of IOP-lowering treatment. The effect of treatment on calibration is not clear from *Figures 15*, *16*, *18* and *19*, although calibration appears to be better in the treated groups. However, these plots should be viewed cautiously because they were based on small study sizes and few conversions to OAG, and they depict only a single imputation, which may be atypical. The calibration slopes presented in *Table 7* were computed based on the 10 imputed data sets. There was little variation in the estimated calibration slope between imputations for all data sets except the Moorfields cohort, and this may be owing to sensitivity to missing data. For the Rotterdam cohort, the calibration slope was close to 1 (1.09, 95% CI 0.72 to 1.46), the ideal value when there is complete agreement between predicted and observed risks, although the example calibration plot presented in *Figure 17*, based on a single imputation, shows an atypical relationship. The calibration slopes were 0.59 (95% CI 0.19 to 0.99), 0.21 (95% CI –0.33 to 0.74) and 0.24 (95% CI –0.17 to 0.65) for the Moorfields, Dunfermline and Nottingham cohorts, respectively. Values of the calibration slope that are < 1 represent overfitting, suggesting that, on average, the effects of the predictors were smaller in these cohorts than in the OHTS-EGPS cohort.

Sensitivity analysis

Vertical cup-to-disc ratio was missing in > 30% of the Rotterdam cohort because the measurements were not taken at baseline. The time interval between VCD ratio measurement

TABLE 7 Multivariate HRs and performance measures for all cohorts based on multiple imputed data

Characteristic	OHTS-EGPS (development set)	Rotterdam (validation set)	Moorfields (validation set)	Dunfermline (validation set)	Nottingham (validation set)
n/N[a]	154/1123	28/393	44/298	28/188	5/159
Baseline predictor, HR (95% CI)					
Age (decade)	1.26 (1.06 to 1.50)	1.44 (0.95 to 2.18)	1.20 (0.86 to 1.68)	0.72 (0.48 to 1.08)	1.23 (0.48 to 3.15)
IOP (mmHg)	1.09 (1.03 to 1.17)	1.15 (1.04 to 1.27)	1.07 (0.99 to 1.17)	1.01 (0.92 to 1.11)	0.94 (0.71 to 1.24)
CCT (per 40 μm thinner)	2.04 (1.70 to 2.45)	1.31 (0.79 to 2.17)	2.76 (1.44 to 5.26)	Excluded	1.62 (0.46 to 5.65)
VCD ratio (per 0.1 larger)	1.19 (1.09 to 1.31)	1.52 (1.19 to 1.94)	1.15 (0.90 to 1.47)	1.36 (1.07 to 1.72)	1.34 (0.76 to 2.37)
PSD (per 0.2 dB greater)	1.13 (1.04 to 1.24)	1.15 (0.95 to 1.40)	1.05 (0.96 to 1.15)	1.06 (0.96 to 1.16)	1.02 (0.94 to 1.11)
Performance measure					
c-index	0.74 (0.70 to 0.78)	0.83 (0.75 to 0.91)	0.69 (0.59 to 0.78)	0.72 (0.63 to 0.82)	0.70 (0.39 to 1.00)
Calibration slope	NE	1.09 (0.72 to 1.45)	0.59 (0.19 to 0.99)	0.21 (−0.33 to 0.74)	0.24 (−0.17 to 0.65)
Calibration χ^2	7.32	NE	NE	NE	NE

NE, not estimated.
a　n is the number of OAG cases and N is the total number of patients.

TABLE 8 Multivariate HRs and performance measures for all cohorts based on complete cases

Characteristic	OHTS-EGPS (development set)	Rotterdam (validation set)	Moorfields (validation set)	Dunfermline (validation set)	Nottingham (validation set)
n/N[a]	154/1123	16/238	16/78	26/135	4/112
Baseline predictors, HR (95% CI)					
Age (decade)	1.26 (1.06 to 1.50)	1.56 (0.89 to 2.73)	1.03 (0.97 to 1.10)	0.73 (0.48 to 1.13)	1.46 (0.48 to 4.38)
IOP (mmHg)	1.09 (1.03 to 1.17)	1.17 (1.04 to 1.33)	1.06 (0.92 to 1.23)	1.03 (0.92 to 1.14)	0.84 (0.59 to 1.21)
CCT (per 40 μm thinner)	2.04 (1.70 to 2.45)	0.99 (0.56 to 1.77)	3.61 (1.25 to 10.4)	Excluded	1.55 (0.53 to 4.56)
VCD ratio (per 0.1 larger)	1.19 (1.09 to 1.31)	1.51 (1.14 to 2.00)	1.15 (0.82 to 1.63)	1.36 (1.07 to 1.73)	1.13 (0.48 to 2.63)
PSD (per 0.2 dB greater)	1.13 (1.04 to 1.24)	1.28 (0.96 to 1.71)	1.24 (1.08 to 1.42)	1.05 (0.95 to 1.16)	1.25 (0.97 to 1.60)
Performance measure					
c-index	0.74 (0.70 to 0.78)	0.83 (0.72 to 0.94)	0.75 (0.63 to 0.88)	0.71 (0.61 to 0.82)	0.79 (0.55 to 1.00)
Calibration slope	NE	1.04 (0.57 to 1.50)	1.36 (0.65 to 2.06)	0.23 (−0.31 to 0.77)	0.92 (0.04 to 1.81)
Calibration χ^2	7.32	NE	NE	NE	NE

NE, not estimated.
a　n is the number of OAG cases and N is the total number of patients.
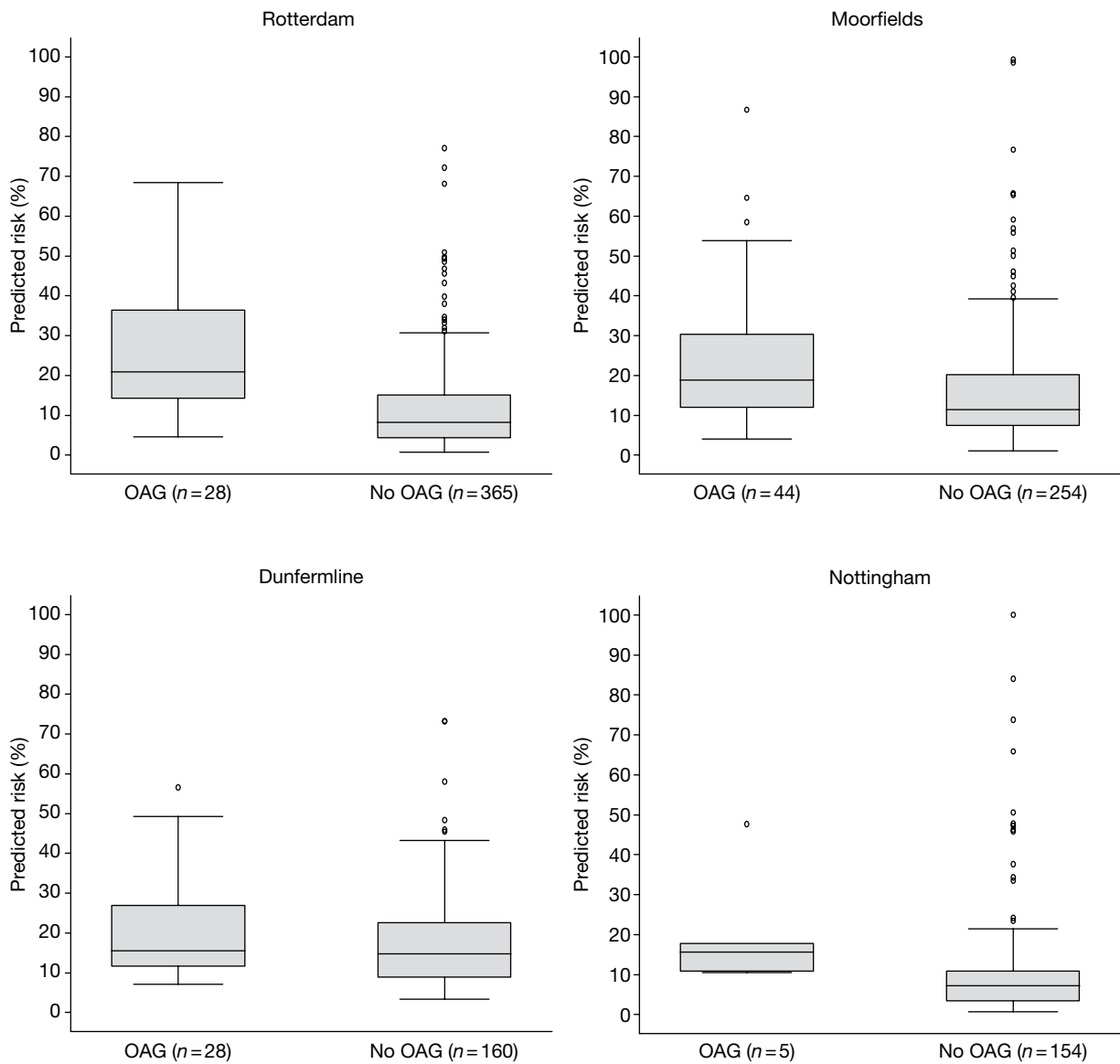
FIGURE 10 Illustrative distribution of predicted 5-year risk in those who developed OAG and those who did not develop OAG during the follow-up period in each of the cohorts for the first imputed data set.
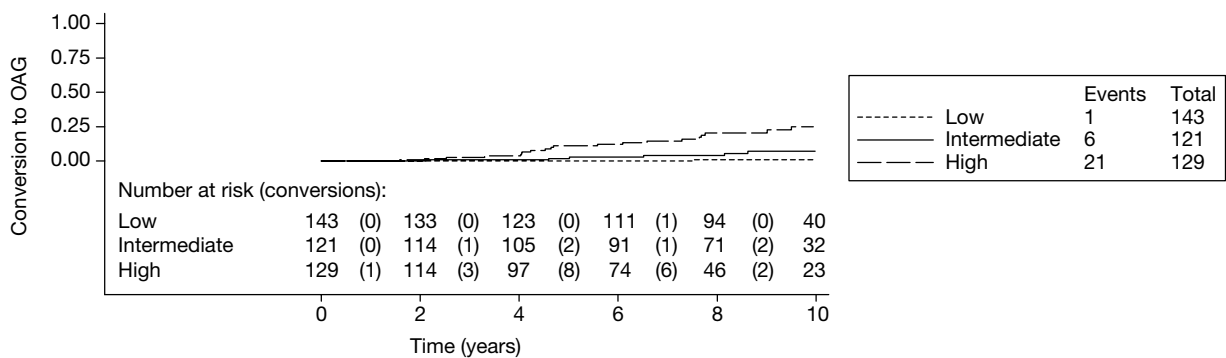
FIGURE 11 Rotterdam cohort: illustrative Kaplan–Meier curve for the risk of conversion from OHT to glaucoma by risk category for the first imputed data set. Log-rank test: $p < 0.001$. The number at risk in each risk category – low (<6%), intermediate (6–13%) and high (>13%) – is shown at 2-yearly intervals. The numbers who converted to OAG during each time interval are shown in parentheses.
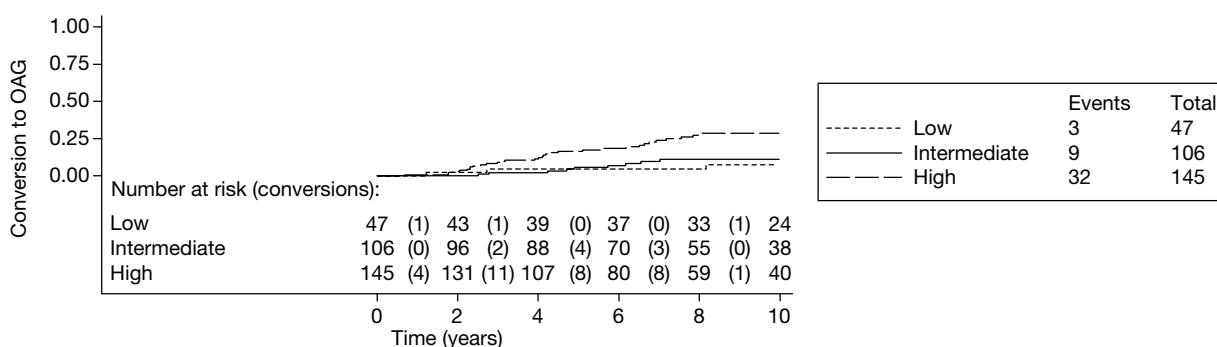
FIGURE 12 Moorfields cohort: illustrative Kaplan–Meier curve for the risk of conversion from OHT to glaucoma by risk category for the first imputed data set. Log-rank test: $p = 0.001$. The number at risk in each risk category – low (< 6%), intermediate (6–13%) and high (> 13%) – is shown at 2-yearly intervals. The numbers who converted to OAG during each time interval are shown in parentheses.
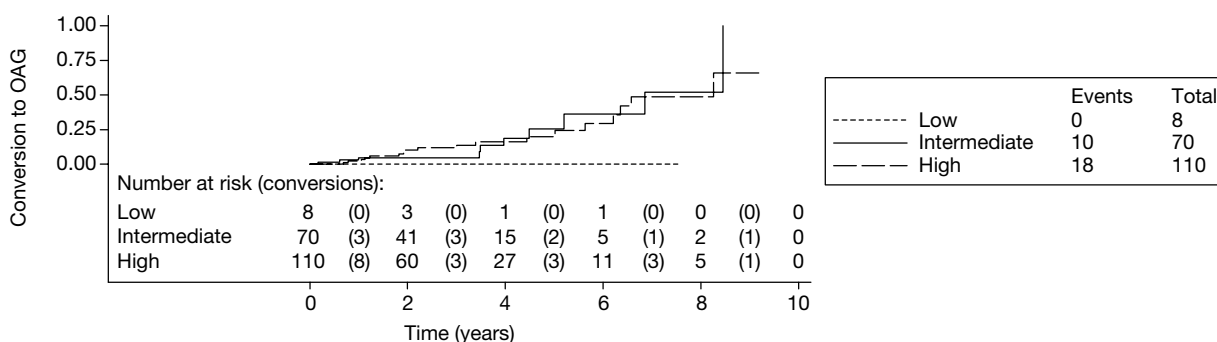
FIGURE 13 Dunfermline cohort: illustrative Kaplan–Meier curve for the risk of conversion from OHT to glaucoma by risk category for the first imputed data set. Log-rank test: $p = 0.6$. The number at risk in each risk category – low (< 6%), intermediate (6–13%) and high (> 13%) – is shown at 2-yearly intervals. The numbers who converted to OAG during each time interval are shown in parentheses.
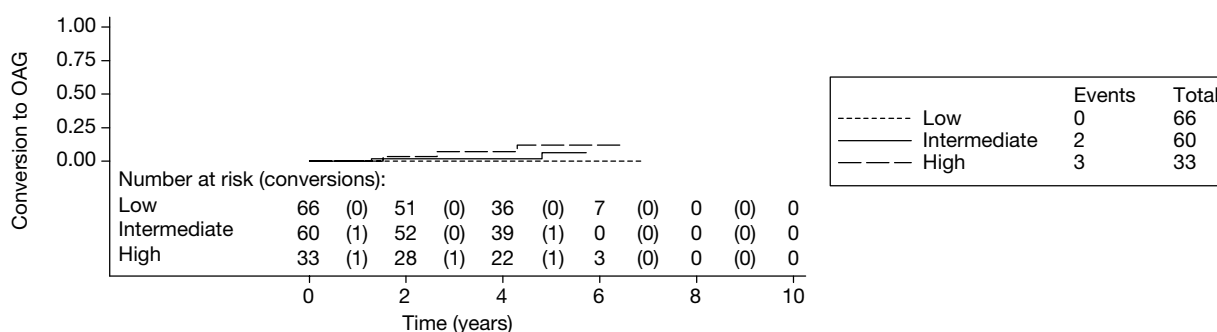
FIGURE 14 Nottingham cohort: illustrative Kaplan–Meier curve for the risk of conversion from OHT to glaucoma by risk category for the first imputed data set. Log-rank test: $p = 0.08$. The number at risk in each risk category – low (< 6%), intermediate (6–13%) and high (> 13%) – is shown at 2-yearly intervals. The numbers who converted to OAG during each time interval are shown in parentheses.
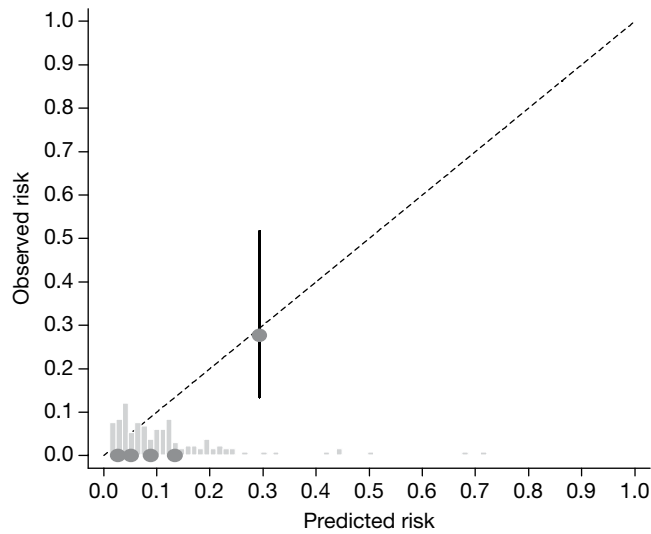
FIGURE 15 Example calibration plot for the Rotterdam cohort for the first imputed data set: untreated group. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
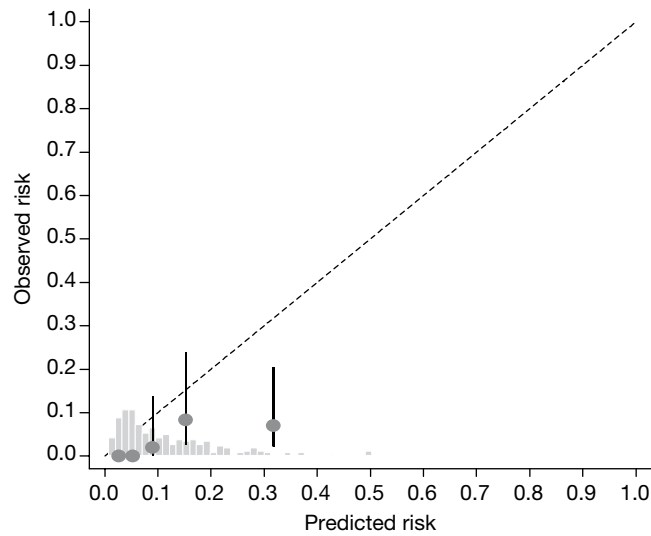
FIGURE 16 Example calibration plot for the Rotterdam cohort for the first imputed data set: treated group. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
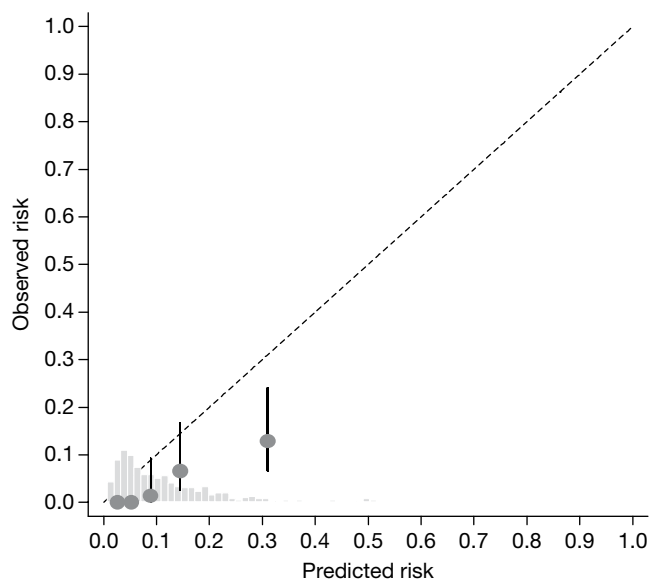
FIGURE 17 Example calibration plot for the Rotterdam cohort for the first imputed data set: all participants. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
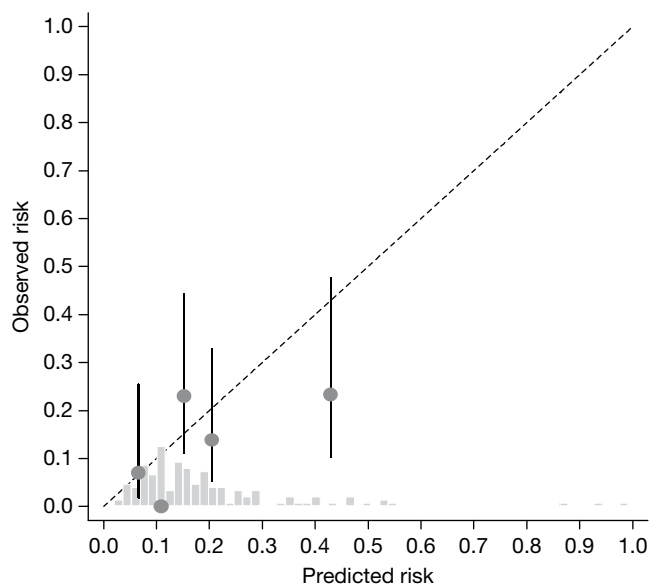
FIGURE 18 Example calibration plot for the Moorfields cohort for the first imputed data set: untreated group. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
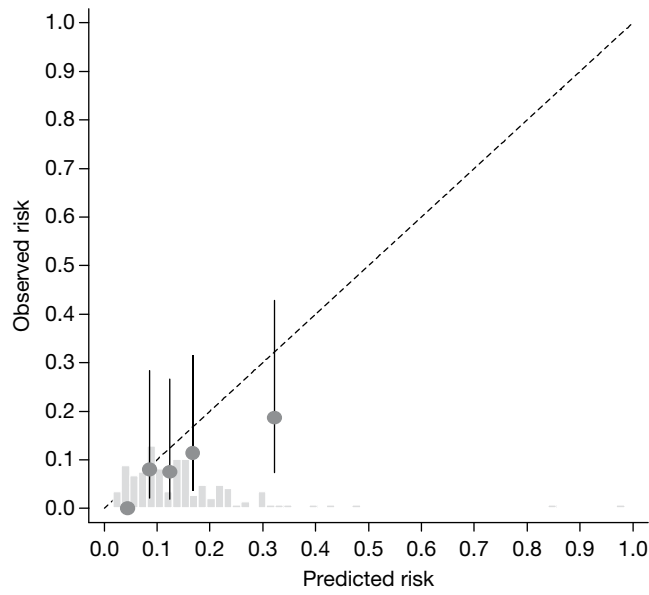
FIGURE 19 Example calibration plot for the Moorfields cohort for the first imputed data set: treated group. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
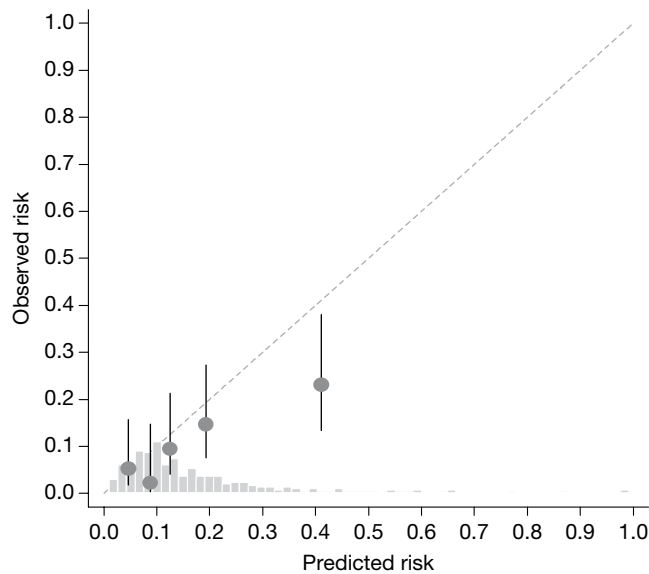
FIGURE 20 Example calibration plot for the Moorfields cohort for the first imputed data set: all participants. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
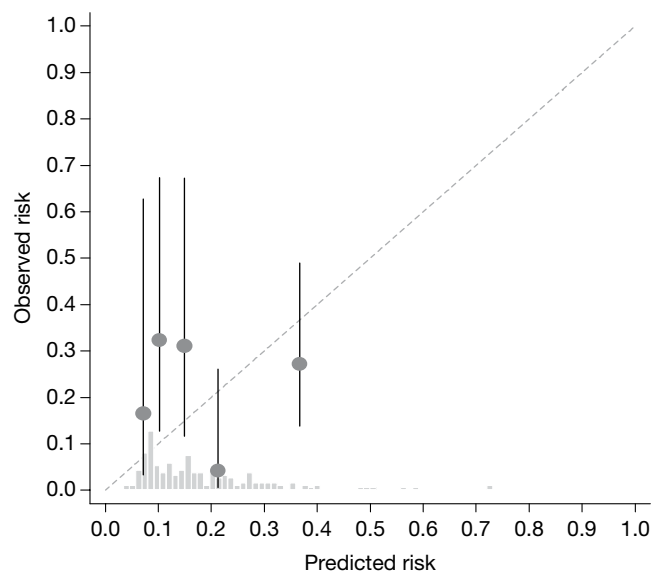
FIGURE 21 Example calibration plot for the Dunfermline cohort for the first imputed data set: all participants. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.
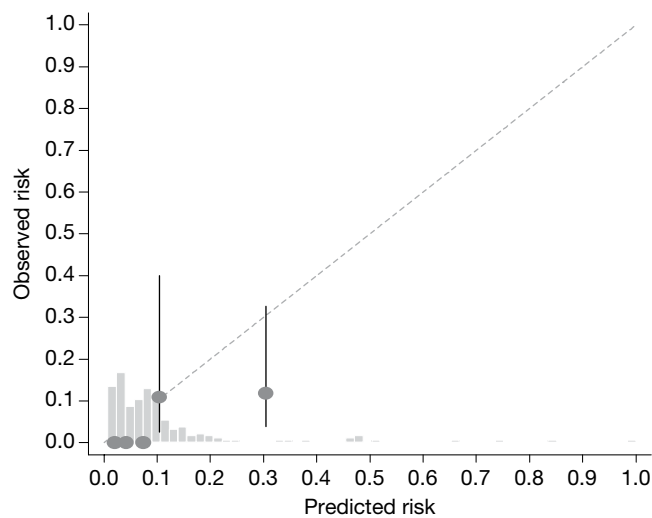
FIGURE 22 Example calibration plot for the Nottingham cohort for the first imputed data set: all participants. For each quintile, the average observed risk is plotted with its 95% CI against the predicted 5-year risk of developing OAG. The dotted straight line through the origin represents perfect calibration. The histogram above the horizontal axis represents the distribution of predicted risks.

and conversion to OAG was < 24 months for five patients in the Rotterdam cohort. Sensitivity analysis using imputed values for these patients did not affect performance of the model (c-index: 0.83, 95% CI 0.75 to 0.91; calibration slope: 1.06, 95% CI 0.64 to 1.47).

Discussion

Systematic review

Several studies have identified potential risk factors for OAG, but there are few risk prediction tools that predict an individual's likelihood of developing the disease over a specified period of time (e.g. 5 years). We found only three models with published prediction equations. These models were derived using either the OHTS or the combined OHTS and EGPS populations. The OHTS-EGPS collaboration produced a model, the 'means' model, which was the most robust, and it is the model referred to as the OHTS-EGPS model. Unlike the OHTS model, this OHTS-EGPS model has not been validated in a different population by independent investigators. The OHTS and OHTS-EGPS models demonstrated good discriminatory ability (c-indexes between 0.68 and 0.74),[63,64] similar to those of the Bach model for lung cancer (0.69, 95% CI 0.66 to 0.72),[67] the EuroScore model for identifying patients with prolonged intensive care unit stay (0.71, 95% CI 0.70 to 0.72),[68] the Framingham model for CHD (between 0.69 and 0.83 in six different cohorts)[69] and better than that of the Gail model for invasive breast cancer (0.58, 95% CI 0.56 to 0.60).[70] All of the variables included in the models are or can be routinely collected in clinical practice. A simple calculator based on the OHTS-EGPS model is available online and can be freely downloaded, enabling quick and easy estimation of the 5-year risk of a patient with OHT developing OAG in at least one eye. A points-based system is also provided.

Intraocular pressure and central corneal thickness

Intraocular pressure and age are important and well-established risk factors, but CCT was a major predictor for the development of OAG over 5 years. For the OHTS-EGPS model, there was an increase in risk per 40-μm decrease in corneal thickness (HR 2.04, 95% CI 1.70 to 2.45). The log-HR standardised according to the SD of the measurements of each predictor in the OHTS-EGPS cohort provides a measure of the contribution of each predictor in the model; CCT was the most important predictor. It is unclear to what extent lower CCT is responsible for the increased risk of OAG and previous studies have shown that CCT tends to decrease with increasing age.[71,72] It is also well known that corneal thickness influences the measurement of IOP because GAT has an inherent measurement error because of CCT variations; IOP is overestimated in thicker corneas and underestimated in thinner ones.[73-75] Nevertheless, there is no consensus on the clinical significance of the effect of CCT on IOP measurements.[76] No correlation was found between CCT and IOP (or any of the other predictors) in either the OHTS or the EGPS. This may be owing to exclusion of patients with normal or very low IOPs and also those with very high IOPs.[77] In the Early Manifest Glaucoma Trial, CCT was identified as a significant predictive factor for glaucoma progression in patients with higher baseline IOP but not in those with lower baseline IOP.[78] Racial differences in CCT have been reported in several population studies, with individuals of African ancestry having thinner corneas, on average, than Caucasians, Hispanics or Asians.[79-85] It has been suggested that black race may not be an independent risk factor because black patients tend to have higher IOP, thinner corneas and higher C/D ratios than other patients with OHT and are therefore generally at a higher risk than white patients.[86] In the OHTS and OHTS-EGPS, when either VCD ratio or CCT was included in the multivariate model, race was no longer statistically significant.[42,64] CCT has also been shown to be a highly heritable trait.[87-89] Because thinner CCT has also been found to predict progression of visual loss in patients with OAG, it is plausible that a biological link exists between aspects of the cornea that regulate its thickness and the physical and structural properties of tissues involved in glaucoma pathogenesis.[76]

Choice of one eye, two eyes or worse eye

Glaucoma can affect either eye or both eyes but it is often asymmetrical, occurring first in one eye only. The prediction models use the mean of the measurements taken from both eyes for all of the eye-specific predictors. Analysis based on a single eye per individual or on the average of both eyes is convenient but may be suboptimal because of intereye differences. Using a single eye could introduce bias if, for example, data on both eyes are available in some individuals, with one eye randomly selected for analysis, but in other individuals are available for only one eye and so are included by default.[90] With respect to IOP, 21% (105/500) of the EGPS participants randomised to the placebo group and included in the OHTS-EGPS analyses had only one eligible eye. Only the data from the included eye were used.[64] Nevertheless, unpublished data from the EGPS showed that conversion to OAG was almost the same in the group of patients with both eyes included as in those with only one eye included in the trial.[91] When both eyes are available, an alternative approach is to use the 'worse' eye, the eye at higher risk of developing OAG; however, this can be problematic when a single variable cannot be used to identify the 'worse' eye. Identification of the worse eye in the OHTS-EGPS 'worse' eye model was based on IOP, CCT, VCD ratio and PSD. Selection of the 'worse' eye was challenging because the eyes of many participants had nearly identical risk estimates at baseline.[65] Using the means of right and left eyes enables the inclusion of relevant data from both eyes of each patient, but potentially at the expense of predictive accuracy. Nevertheless, the performance of the OHTS-EGPS prediction model based on the means of right and left eyes for eye-specific variables was compatible with prediction models that included additional information on baseline differences between eyes.[65]

Limitations of the prediction models

The models are limited because they do not include other major risk factors such as race and family history of glaucoma. Race was not associated with the risk of developing OAG in univariate or multivariate analyses of the OHTS or the pooled OHTS-EGPS data. All of the 439 patients from the EGPS included in the pooled data were Caucasians, and the lack of association may be owing to limited power in the OHTS and the combined data set. In the OHTS, data on family history of glaucoma (parent or sibling) were collected by patient recall, without verification, and so may be subject to recall bias. This information was not collected in the EGPS. A positive family history of glaucoma was not significantly associated with an increased risk of OAG in the OHTS analyses. Conversely, the Barbados Eye Studies found more than a twofold increase in risk (relative risk 2.4, 95% CI 1.3 to 4.6) after adjustment for age, gender, IOP, and IOP and blood pressure treatment.[38] Other ocular conditions, such as PEX[36,37,41,92] and optic disc haemorrhage,[78,91,93] have been found to be predictive for the development of OAG. The OHTS excluded individuals with PEX and pigment dispersion whereas the EGPS included only 19 individuals with these conditions and they were excluded from the pooled analyses.

Another limitation of the models is the use of only baseline measurements of predictors – except for CCT, which was measured later during follow-up – without accounting for fluctuations and changes over time in any of the predictors, especially IOP. IOP fluctuation has been suggested as an independent risk factor for glaucomatous optic nerve damage, but reports are conflicting.[94] In the EGPS, the area under the curve of IOP, a 'time-dependent' evaluation of IOP, taking into account the IOP behaviour for each 6-month period, was associated with the development of OAG yet IOP fluctuation during follow-up was not.[91] Medeiros and colleagues[95] found that long-term IOP fluctuations were not a significant risk factor for glaucoma development in untreated patients with OHT observed over time, but higher mean IOP levels were consistently associated with increased likelihood of developing glaucoma. Similarly, in the Malmö Ocular Hypertension Study, IOP fluctuations were not an independent risk factor for the incidence of glaucomatous visual field loss in patients with OHT, and higher IOP fluctuation was found in eyes with higher IOP levels.[96]

Model validation

We assessed the external validity of the OHTS-EGPS model by investigating its discriminatory ability and calibration in four different populations. In the Rotterdam, Moorfields and Nottingham cohorts, the model reliably stratified patients into the three risk categories (low, intermediate and high), but the model performed best in the Rotterdam cohort. The ability of the model to distinguish between patients with OHT who developed OAG and those who did not was good in all four cohorts, with *c*-indexes ranging between 0.69 and 0.83. A *c*-index of 0.83 indicates that, in approximately 83% of the cases, the model allocated a higher predicted probability to an individual who converted than to one who did not.

In the model calibration analyses, the OHTS-EGPS model generally overestimated the risk of OAG in all cohorts. In the Moorfields and Dunfermline cohorts, there were fewer patients in the low-risk group than in the intermediate- and high-risk groups. In contrast, patients in the Rotterdam cohort were almost equally distributed among the three risk groups. Because of differences in case mix, a prediction model is likely to be more appropriate in a population similar to the one used to derive the model.[97] For instance, a risk model derived from a more heterogeneous case mix provided better calibration in risk assessment across the spectrum of acute coronary syndromes than another model derived using a clinical trial population, although both models showed similar discriminatory ability.[98]

Predictor effects

The HRs obtained when the OHTS-EGPS model was fitted to our cohorts were similar to those obtained in the development cohort, although they were not always statistically significant. Our cohorts were much smaller than the OHTS-EGPS cohort. CCT was the only statistically significant predictor in the Moorfields cohort (HR 2.76, 95% CI 1.44 to 5.26), but was not significant in the Rotterdam cohort (HR 1.31, 95% CI 0.79 to 2.17). This may be because average CCT was lower in the Moorfields cohort than in the other cohorts. Seventeen patients (12%) in the Dunfermline cohort had an average PSD value > 3, the upper limit allowed in the OHTS-EGPS model, and the HR was 1.06 (95% CI 0.96 to 1.16) per 0.2-dB increase in PSD. A large PSD could be a sign of early optic nerve damage and unconfirmed glaucoma but can also reflect an unreliable field, for example because of incorrect patient positioning when undertaking the perimetry. In the observational cohorts patients are likely to have had only one baseline visual field examination, unlike the trial cohorts for which data from at least two reliable baseline visual fields would have been required.

Generalisability

The trial-based cohorts were a highly selected population. The OHTS and EGPS excluded pregnant or nursing women, those with best corrected visual acuity worse than 20/40 in either eye, a history of previous intraocular surgery, life-threatening or debilitating disease, secondary causes of elevated IOP, systemic or ocular conditions capable of causing visual field loss or optic disc abnormalities, patients with diabetes who had evidence of diabetic retinopathy and those on medication. In addition, in the OHTS, patients with pigment dispersion or exfoliation syndrome were excluded. Most of these criteria were applied in the selection of the Rotterdam and Moorfields cohorts (see *Chapter 3*). The OHTS and DIGS used similar patient selection criteria and the calibration slope (1.086) of the OHTS model in the DIGS cohort was similar to that of the OHTS-EGPS model in the Rotterdam cohort. In both cases, the values were not statistically different from 1, the ideal value. Selection criteria were more flexible in the two observational cohorts, yet both cohorts may not be representative of the general OHT population. Low-risk cases in the Dunfermline hospital-based registry were discharged after two visits (at 1 year), and in the Nottingham community-based registry they were discharged if they had untreated IOP consistently < 21 mmHg as well as normal visual fields and optic discs. Patients in the Nottingham scheme suspected of conversion or with IOP above recommended predefined guidelines (see

Chapter 3) were referred back to secondary care. If the secondary-care clinician confirmed conversion, the patient was treated and retained in the secondary-care system. However, if no conversion was judged to have occurred, the patient was returned to the community scheme for annual review. The Dunfermline and the Nottingham data sets represent routinely collected data in clinical practice. A risk prediction tool is potentially most valuable in this context and used as an aid to define criteria for monitoring and treatment guidelines for the management of OHT. Because the discriminatory ability of the model was generally good, the model can be adjusted or recalibrated to correct for model optimism; calibration errors can be eliminated without affecting the discriminatory ability of the model. However, given the limitations of our data, we did not attempt to recalibrate the model to correct for the overestimation of risk.

Variable and outcome definition

Differences in the methods used for data collection or in the definitions of the predictors or the outcome variable can affect model performance.[97] Variability of IOP measurements is a well-known phenomenon, and in both the OHTS and the EGPS between two and nine measurements of IOP were taken in each eye to obtain a stable estimate of baseline IOP. The variability of IOP measurements in the same eye between consecutive visits has been shown to be moderate and is greater than the variability of IOP measurements between right and left eyes at the same visit.[99] IOP of $\geq 22\,\mathrm{mmHg}$ and $< 35\,\mathrm{mmHg}$ on two or more occasions measured within a 2-week period was the entry criterion for the Moorfields cohort.[32] Yet 22% of the Moorfields cohort had an IOP of $< 20\,\mathrm{mmHg}$ at baseline, a likely reflection of the regression to the mean. One IOP measurement was recorded in each eye, which is typically the case in clinical practice.

In the OHTS and the EGPS, CCT was measured using ultrasonic pachymeters and an average of five measurements per eye was carried out; this was also the case in the Nottingham scheme. Intereye difference in CCT tends to be small but, to ensure data quality and reproducibility, in the OHTS, repeat measurements were made when the intereye difference was $\geq 40\,\mu\mathrm{m}$.[71] In clinical practice, only one CCT measurement, based on an average of five readings at one visit, is undertaken.

Pattern standard deviation is an objective index of severity of visual field loss. However, the evaluation of the VCD ratio (especially in clinical practice) is subjective and is expected to have a large intra- and interobserver variability.

For risk estimation, the online calculator requires three IOP measurements per eye taken using GAT, three CCT measurements per eye using an ultrasound pachymeter and two PSD measurements per eye using any of the following: Humphrey full threshold 30-2 or 24-2; SITA standard 30-2 or 24-2; or loss variance from Octopus 32-2. Only one measurement of each of these variables was recorded per eye in our cohorts, except for the visual field parameter (PSD), for which at least two baseline visual fields were required.

Events and conversion criteria

With the exception of Dunfermline, there were few incident cases of OAG and the rate of conversion varied between cohorts. Substantial variation in the incidence of OAG was also observed between clinics in the OHTS (3.7–42.9%) and the EGPS (0–25%). The variation was attributed to differences in characteristics of the participants enrolled at the various clinics as well as the small number of participants in some clinics.[64] The rate of conversion from OHT to OAG may also depend on the conversion criteria used. Glaucoma is diagnosed on the basis of a characteristic pattern of damage to the optic nerve with a corresponding and reproducible visual field defect. The conversion criteria used in the observational cohorts were not as rigid and well defined as those used in the trial-based cohorts; conversion was generally based on the development of a visual field defect. This is likely to be a better reflection of clinical practice

than the criteria and processes adopted in the trials, but is unlikely to be consistent across the validation data sets as there is no 'gold' standard case definition of early glaucoma. The OHTS and EGPS also had a mixed population comprising newly diagnosed and existing patients with OHT attending clinics. Including the latter group could have introduced bias because those with OHT who converted to OAG before recruitment started were excluded from the study, and those who did not convert may be less likely to convert than the whole population with OHT.

Limitations of the model validation

One of the limitations of this study was the inclusion of treated patients in all cohorts, although the only predictor affected by treatment is IOP. This may be important for those who were untreated at the time of recruitment into the study, but is less important for those who were already on treatment when entered into the study cohort, as may be the case with the observational (Dunfermline and Nottingham) cohorts. Although the OHTS-EGPS model was developed to estimate the risk for a patient left untreated for 5 years, the HRs were similar to those of the OHTS model based on treated and untreated patients. This OHTS model was validated in an untreated population and similar HRs were obtained for the predictors.[63]

A second limitation was the number of patients and OAG cases in some cohorts. In the UK, patients considered to have a moderate or high risk are often kept in secondary care whereas those deemed to be very low risk are discharged (i.e. no follow-up). This may explain the higher incidence of OAG in the Dunfermline cohort (28/188, 14.9%) than in the other cohorts, and the very low incidence in the Nottingham cohort (5/159, 3.1%) compared with the DIGS cohort (31/126, 24.6%),[63] although it exceeds that of the Visual Impairment Project (66/2415, 2.7%).[37] Some patients in the Nottingham cohort received treatment for OHT prior to entry into the scheme, and this could account for the low incidence of OAG, as well as the strict entry criterion (see *Chapter 3*, *Data from observational data sets*). Thus, the Dunfermline (secondary care) and Nottingham (community care) cohorts may be unrepresentative of the wider OHT population.

A third limitation was the large proportion of missing values for some predictors in some cohorts. CCT was missing in >20% of patients in each cohort (100% in the Dunfermline cohort), and was more likely to be missing in those who were enrolled earlier or dropped out/discharged early in the study. The OHTS was the first study to prospectively document that a lower CCT measurement predicts the development of OAG and recommended its measurement in the clinical evaluation of patients with OHT.[42] CCT measurements were not included originally in the protocols of OHTS and EGPS, and 184 patients without CCT measurements were excluded from their analyses. The Moorfields study commenced in 1992, at a time when CCT was not measured routinely, and CCT data were collected during follow-up. It is possible that only those at higher risk of developing glaucoma had a CCT measurement.

A fourth limitation was the enrolment of only Caucasians in the Rotterdam and Dunfermline cohorts, and the fact that race was not documented in the Nottingham cohort (it is likely that there were few non-Caucasians because black patients were considered to be at higher risk of OAG and so were unlikely to be referred to the scheme). Although race was not a predictor in the OHTS-EGPS model, the OHTS population was racially diverse, and approximately 25% of patients enrolled in the study were African Americans. Because there is evidence to suggest that blacks and whites differ with respect to several risk factors for OAG,[100] the poor calibration in some cohorts may be partly because of racial differences.

A risk prediction model cannot replace clinical decision-making but is a tool that can facilitate more objective assessment and management of patients. It has been shown that ophthalmologists underestimated the risk of glaucoma in patients with OHT compared with the risk obtained using a risk calculator.[101] Furthermore, the use of a risk calculator changed treatment recommendations made by glaucoma specialists.[102]

Conclusions

To guide therapeutic intervention and for the prevention of OAG there is a need for accurate and reliable risk assessment for conversion from OHT to OAG. The applicability of a multivariate risk prediction model beyond the setting in which it was developed is essential to its adoption and acceptability in clinical practice. We identified and validated a risk prediction model that can be used for objective risk assessment in the management of patients with OHT (the OHTS-EGPS model). Despite the limitations of our cohorts, the OHTS-EGPS model generally discriminated well in estimating the 5-year risk of conversion to OAG but was not well calibrated in all populations. It could provide more accurate risk assessment in conjunction with clinical assessment, but its usefulness in different settings and populations other than trial populations remains uncertain. Validation in a population-based setting is important because of its relevance to daily clinical practice. Therefore, further validation, especially in non-Caucasians, and recalibration in relevant patient subgroups is needed. Prediction models tend to evolve over time as the role and impact of existing risk factors become more apparent, or new candidate predictors are identified. The model has not been revised since its original publication and future research may seek to update the model to reflect such new knowledge and potentially improve predictive ability.

Chapter 5

Agreement and reliability of candidate tonometers for measuring intraocular pressure

Introduction

Background

Raised IOP is the most important risk factor for glaucoma and is the only one that is treatable. The instrument used to measure IOP is called a tonometer. The desirable attributes for a tonometer for use in a monitoring programme are accuracy, precision, acceptability to patients and ease of use. GAT, a contact tonometer, is currently the tonometer most widely used by ophthalmologists and is accepted as the current clinical standard. However, GAT has several limitations (see below for further information), and because of the skills required for its interpretation it is not ideal for the monitoring setting. In this setting, tonometers that do not touch the cornea (non-contact) and the use of which does not require extensive training would be both preferable and more practical. In recent years, a variety of new tonometers for estimating IOP have emerged with potential advantages, including being easy to deliver, non-contact, automated and self-administered, and compensating for, or not being influenced by, corneal thickness and other properties of the cornea.[25,26]

Description of technologies

Tonometers can be categorised as contact and non-contact tonometers depending on whether or not they involve direct corneal contact. In most cases, measurement of IOP using a tonometer should be carried out by a health-care professional. However, as described below, some tonometers [Ocuton S® (EPSa Elektronik & Präzisionsbav, Saalfeld, Germany) and transpalpebral tonometers] have been specifically designed for self-measurement.[27]

Contact tonometers

This term refers to tonometers that have direct contact with the cornea.

Goldmann applanation tonometer

The GAT, a contact tonometer, is currently the instrument most commonly used by ophthalmologists to estimate the IOP. It is a slit lamp-mounted device. However, there are some limitations associated with GAT, including the influence of corneal thickness and corneal biomechanical properties, the potential for transmitting infections or causing corneal abrasion and its relative difficulty of use. The GAT calibration assumes that the cornea has a central thickness of between 530 and 560 μm and IOP is likely to be an under- or overestimated when using GAT in people with thinner or thicker corneas, respectively.[103] The operating manual recommends three measurements, but this is often not undertaken in clinical practice.

Dynamic contour tonometer

The dynamic contour tonometer (DCT) is a slit lamp-mounted, contact, digital, non-applanation tonometer and is operated in a fashion similar to GAT. It is commercially available as the PASCAL® DCT (SMT Swiss Microtechnology, Port, Switzerland). Corneal anaesthesia is required

and there is no need for fluorescein. As corneal thickness is an important factor influencing the IOP measurement, the PASCAL DCT minimises the effect of the architecture of the cornea by using a built-in 'sensor tip' with solid-state 'pressure sensor', which matches corneal curvature. The concave contour surface approximates the corneal shape when the pressures on both sides of it are equal, with minimum distortion, and directs all forces acting within the cornea to the pressure sensor surface, providing an IOP measure. The pressure in the eye is detected 100 times per second and presented as the diastolic IOP (mmHg) in the LCD (liquid crystal display) screen together with the ocular pulse amplitude (OPA) and a quality score, 'Q'. OPA is the difference between the systolic and diastolic IOP. The quality score of the data and results can be interpreted as follows: 'Q1' is optimum; 'Q2' and 'Q3' are acceptable; and 'Q4' and 'Q5' are poor and indicate that the measurement should be repeated. The PASCAL DCT self-calibrates at the beginning of every measurement although a performance test can also be performed.[104] A measurement takes about 5 seconds to achieve. Good cooperation is necessary to maintain a steady eye and head position, which are required during the measurement.

Ocuton S

The Ocuton S is a self-measurement tonometer that calculates and displays the IOP value automatically through direct contact of the measuring prism with the cornea. The use of topical anaesthetic is required.[105]

The Perkins applanation tonometer

The Perkins applanation tonometer is a hand-held device but otherwise uses the same principles as GAT and requires topical anaesthesia and fluorescein instillation. All hand-held applanation tonometers were included under this heading.

Rebound tonometer

The rebound tonometer (RT) is a simple portable device, commercially available as the Icare® tonometer (Tiolat, Helsinki, Finland). Although it is a contact tonometer, topical anaesthetic drops are not required and the tonometer has a disposable tip to minimise the risk of cross-infection. The device processes the rebound movement of a rod probe resulting from its interaction with the eye; there is a shorter duration of impact as the IOP increases. The rebound is influenced by corneal thickness, and for this reason measurement of IOP by this tonometer is prone to measurement error as a result of corneal properties.[106,107] Like GAT, it is calibrated for a typical CCT value. Six measurements are recommended to provide accurate results, and the average of six IOP measurements is displayed on the LCD.[108] Regular calibration is required.

TonoPen

The TonoPen® (Mentor O&O Inc., Santa Barbara, CA, USA; Reichert Inc., Depew, NY, USA) is a hand-held, self-contained portable tonometer that determines IOP by making contact with the cornea (central contact is recommended) through a probe tip, causing applanation/indentation of a small area. The tip contains a transducer that measures the applied force on the cornea. Topical anaesthetic eye drops are used. After four valid readings are obtained, the averaged measurement will appear on the LCD screen. Up to 10 measurements can be performed according to the manufacturer recommendations. It is recommended that calibration is performed daily before instrument use, when indicated by the LCD screen or whenever batteries are replaced.

Non-contact tonometers

This term refers to tonometers that do not have direct contact with the cornea.

Non-contact tonometer

The air-puff tonometer uses a rapid air pulse to applanate (flatten) the cornea, thus working on the same basic principle as the Goldmann tonometer. The force of the air stream increases

linearly over milliseconds, progressively flattening a known area of the cornea. The moment of applanation is determined by an optical sensor and the air pulse is then interrupted. The advantages of the non-contact tonometer (NCT) include speed, the fact that there is no need for topical anaesthesia, and thus a low risk of corneal abrasion (hence its use requires minimal training), and, because there is no direct contact with the eye, any infection issues are avoided. IOP measurement by NCT is affected by the corneal thickness. There are several models available in the market; calibration and the number of recommended measurements may vary accordingly. Some models seek to correct the measurement for the CCT. Some patients find the air puff uncomfortable.

Ocular response analyser

The Ocular response analyser® (ORA) (Reichert Inc., Depew, NY, USA) utilises air-puff technology and an electro-optical system to record two applanation measurements of IOP, one while the cornea is moving inward and the other as the cornea returns. Because of its viscoelastic property, the cornea resists the dynamic air puff, delaying the inward and outward applanation process and resulting in two different pressure values; the average of these two IOP values provides a repeatable, Goldmann-correlated IOP measurement (IOP_G). The difference between these two IOP readings is corneal hysteresis (CH), a new measurement related to corneal tissue properties that is a result of viscous damping in the corneal tissue.[109] The CH measurement provides a basis for two additional new parameters: corneal-compensated intraocular pressure (IOP_{CC}) and corneal resistance factor.[110,111] The IOP_{CC} is an IOP measurement that is less affected by the corneal properties. Four good-quality readings per eye are recommended.[112]

Transpalpebral tonometer

This type of tonometry includes devices that measure IOP through the eyelid avoiding direct corneal contact. Topical anaesthesia is not required. The Diaton® tonometer (BiCOM Inc., Long Beach, NY, USA; previously commercialised as TGDc-01, Ryazan State Instrument-Making Enterprise, Ryazan, Russia) is a hand-held, pen-like portable device applying this principle. The pressure phosphene tonometer (PPT; Proview® Eye Pressure Monitor, Bausch & Lomb Inc., Rochester, NY, USA) has been developed as a self-measurement tonometer. The PPT is a spring compression device calibrated in mmHg that consists of a probe with a flat applicator of the same diameter (3.06 mm) as the area applanated by the GAT. This instrument delivers a phosphene spot when pressure is applied through the closed eyelid in the superior nasal portion of the eye, a self-perceptible visual phenomenon. The threshold pressure for creating a phosphene spot is the estimated IOP.[113]

Aim and objectives

Aim
- To compare the agreement, recordability, practicality, acceptability and reliability of the tonometers used in clinical practice using GAT as the reference tonometer.

Primary objectives
- To compare the agreement of IOP readings of one or more tonometers in adults with the readings of GAT as the reference tonometer.
- To explore the factors affecting the agreement between tonometers including CCT, IOP level, previous corneal refractory surgery, type of examiner and use of disposable tonometer heads.

Secondary objectives
- To report the recordability (proportion of measurements that are recordable) of the alternative tonometers.

- To report the practicality of the alternative tonometers from the 'doer' (examiner) point of view and their acceptability from the 'user' perspective.
- To compare the reliability when reported of the comparator tonometers with that of GAT, including intra/interobserver reliability.

Methods

Inclusion and exclusion criteria

Types of study

Direct comparative studies that assessed the agreement of one or more tonometers with the reference standard tonometer (GAT) in the same group of people were included. Non-English-language studies and conference abstracts were excluded.

Types of participant

Adults aged > 16 years, including those with a diagnosis of OHT or glaucoma, representative of the general population, were included. When the age range was not reported, confirmation from the authors was sought. If no response was received and a mean age and SD or median age and interquartile range (IQR) were provided, a formula was applied (mean $- 3SD \geq 16$ or median $- IQR/1.35 \times 3 \geq 16$) to assess inclusion. This was carried out to prevent exclusion purely on the failure to report the age range of the participants when it was very unlikely that any of the participants were under 16 based on data on the age distribution. Participants with corneal abnormalities were excluded (corneal pathology, including keratoconus, bullous keratopathy or post-corneal grafts).

Types of technology
Reference tonometer
The reference tonometer was the GAT.

Comparator tonometers

All tonometers that could be conceivably used in a monitoring context were eligible for inclusion. Studies that evaluated the agreement of manometry[114] (invasive procedure) were not eligible. Tonometers that were primarily used as a research device (such as ocular blood flow[115] – recommended for research only) or were unsuitable/unavailable in a clinical setting were not eligible. If a study compared both an eligible and a non-eligible tonometer with GAT the study was included.

Types of examiner

Tonometry performed by any type of examiner including optometrists, ophthalmologists, nurses, technicians and patients was included.

Primary outcome

The primary outcome was the agreement [mean difference and limits of agreement (LoA)] between a tonometer and the reference standard.

Secondary outcomes

- Interobserver reliability for two observations taken by different observers with the same tonometer.
- Intraobserver reliability for two observations taken by the same observer with the same tonometer.
- Practicality for 'doers' using the technologies.
- Acceptability of the tonometers to users and providers.
- Proportion of participants with a recordable IOP (recordability).

Search strategy

Sensitive electronic searches were conducted to identify reports of published and ongoing studies on the reliability and agreement of tonometers. Databases were searched from 1987 until February 2010 and searches were restricted to articles published in English. Conference proceedings were not included. Studies prior to 1987 were not considered because of technology changes. The following bibliographic databases were searched: MEDLINE, MEDLINE-In Process & Other Non-Indexed Citations, EMBASE, Science Citation Index, BIOSIS and the Cochrane Central Register of Controlled Trials. The websites of key journals were screened for additional relevant or in-press publications, including *American Journal of Ophthalmology*, *Archives of Ophthalmology*, *British Journal of Ophthalmology*, *Eye*, *Graefe's Archive for Clinical and Experimental Ophthalmology*, *Investigative Ophthalmology and Visual Science*, *Journal of Glaucoma* and *Ophthalmology*.

Additional searches were undertaken in current research registers, including ClinicalTrials.gov, Current Controlled Trials and World Health Organization International Clinical Trials Registry Platform, and in the HTA database, DARE and the Cochrane Database of Systematic Reviews (CDSR) for relevant evidence synthesis reports. An internet search using Copernic Agent was also undertaken and included key professional organisations and manufacturers of tonometers.

Full details of the search strategies used are provided in *Appendix 1*.

Data extraction strategy

Two reviewers (AA, AA-B, KMc, ABP or JB) independently screened the titles and abstracts (if available) of all reports identified by the electronic searches. Full-text copies of all studies deemed to be potentially relevant were obtained and independently assessed for inclusion by two reviewers (AA, AA-B, KMc, ABP or JB) using a screening tool developed for this review. Authors were contacted by email (if provided in the manuscript) when only age range data were missing. Any disagreements were resolved by consensus or arbitration by a third party (AA-B, JB).

A data extraction form was developed and piloted. Two reviewers (AA, ABP, JC or AE) independently extracted data on study design, participant characteristics, type of tonometer used and outcome data. We conducted a 20% check of all extracted data (ABP). When outcome data were provided per eye (right/left) and overall (average measurement per participant) for each comparison, right eye data were used. If studies compared different versions of the same technology in the same study only data on the most recent tonometer version were included. IOP_{CC} data were used for ORA. When an individual study provided data from two different groups (e.g. normal/glaucoma), the study with the bigger sample size was included in the analysis. If the sample size was the same in both groups, data from the group with the higher IOP measured by GAT were included. When measurements were performed during different sessions (days apart), data from the first session were used. If the measurements were taken and results reported at different time points during the same day, a measurement time likely to occur in practice (e.g. 10AM) was used. When measurements taken before and after surgery were available, preoperative results were included in the main analysis.

When raw outcome data were provided, mean values and SDs per tonometer were calculated. When outcome data were not provided, and if authors' details (email address) were available, they were asked for the mean difference or the SD of the difference between GAT and the comparator tonometer. When mean differences were not reported, they were calculated from the reported data (e.g. GAT and comparator means) by reading the values from a published Bland–Altman plot. When a difference of opinion existed, a third party was consulted (JC).

Quality assessment strategy

Two reviewers (AA, AA-B, ABP or JB) independently assessed the quality of all included studies using a modified checklist adapted from Whiting and colleagues[116] and Craig and colleagues.[117] Each item was graded as 'yes', 'no', or 'unclear'. The quality assessment of diagnostic accuracy studies (QUADAS) tool[116] is a recently developed quality assessment tool for use in systematic reviews of diagnostic accuracy studies; however, for this review it was adapted to make it applicable to assessing the quality of reliability and agreement studies. We conducted a 20% check of quality assessment (ABP, JC). Discrepancies were resolved by discussion or arbitration. We classified studies as low quality if they did not meet one or more of the quality criteria.

Data analysis

General approach

The primary outcome, agreement, was assessed by calculating summary LoA.[118] Secondary outcomes were tabulated with no quantitative analysis conducted. The 95% LoA interval was calculated for each candidate tonometer from pooled estimates of the mean difference (systematic difference) between a tonometer and the reference standard and of the corresponding variability of agreement (random error). Pooled estimates of mean difference and random error were calculated using the DerSimonian and Laird random-effects method.[119] Measures of agreement variability were based on reported (or derived) SDs of within-participant differences. Imputation of within-participant correlation coefficients to allow calculation of the SD of differences was employed, if required, when a correlation estimate was available from other studies of the same type of tonometer. The imputed value was the arithmetic mean of available correlation estimates when more than one study estimate was available. Sensitivity analyses included a fixed-effect analysis and/or imputation of correlations using the minimum correlation coefficient reported from the studies comparing the same tonometer. Additionally, an approximate 95% prediction interval was calculated for both parameters using the estimated τ-value ($\pm 1.96\tau$) from the random-effects analysis to quantify the impact of between-study heterogeneity on the systematic difference and the random error. The prediction interval provides a range of values that could plausibly be observed if a new study was undertaken, based on the observed between-study heterogeneity. Finally, the proportion of studies within 2 and 3 mmHg of GAT was estimated from the pooled difference and SD using the cumulative distribution function for the standard normal deviates for values within ±2/3 mmHg.

Additional sensitivity analyses

Further sensitivity analyses looked at the impact of excluding studies that used suboptimal methods according to our quality assessment tool (i.e. studies in which at least one of the requirements is clearly not met), excluding studies that reported data clustered within persons (i.e. studies in which some or all of the participants contributed data on more than one eye but in which data for one eye only was unavailable) and using the standardised mean difference (SMD) metric, that is, mean difference divided by pooled SD, to address variation in the number of measurements (systematic difference only). An additional analysis was conducted to correct for the underestimated variation in studies with repeated measurements by using reported estimates of within-participant variation to adjust the results of such studies to reflect the variation if only a single measurement had been taken.[120]

Clinical factor analyses

Heterogeneity between the study estimates in the meta-analyses was explored by visual inspection of forest plots and calculation of I^2 statistics. Possible reasons for heterogeneity were explored through prespecified clinical factor analyses. Where possible, studies were categorised according to CCT, previous corneal refractory surgery, type of examiner and IOP level, with corresponding meta-analyses being conducted. For some studies, data relating to a subset of the

main cohort were used for the factor analyses if sufficient data were given for the subset but not the full cohort. For the IOP analyses, studies in which the estimated proportion of the sample with OHT (i.e. > 21 mmHg) was > 33% were compared with studies in which this proportion was estimated to be < 33%. The proportion was estimated using the mean and SD for GAT, assuming a normal distribution (studies in which this was not reported were excluded from the subgroup analysis). A similar approach was taken for the CCT analysis with one of the two subgroups being defined as studies in which 33% of patients in the sample had CCT of < 555 µm (i.e. patients with the highest risk of glaucoma conversion). It was not possible to carry out analyses on the use of disposable tonometer heads as none of the included studies reported such data.

A clinical factor analysis was conducted to compare studies with no previous refractive surgery in their sample with studies that had postoperative subsamples. Another analysis compared studies in which the comparator examiner(s) were known to be solely ophthalmologists with studies in which the comparator examiner(s) were all known to be non-ophthalmologists. For self-tonometers, studies in which patients were the sole examiner type were included as a third category in this subgroup analysis. Because of the observed level of heterogeneity, a further clinical factor analysis investigated the impact of manufacturers in studies in which multiple manufacturers produced the same type of tonometer. Formal comparison between factor subgroups was not conducted because of the high level of heterogeneity in the main analyses. When individual studies reported on the impact of the clinical factors considered in the analyses the results of the individual studies were summarised in narrative form without any statistical analysis.

Data abstraction

When the mean difference was not reported, it was derived from the reported mean GAT and mean comparator IOP values. Authors were contacted for clarification if a discrepancy existed between reported mean difference and mean GAT/comparator scores (other than owing to rounding). The study was excluded from the meta-analysis if it was not possible to obtain a satisfactory value for the mean difference. Where IPD were published in the included study report, the IPD were used to calculate all necessary statistics.

The reported mean difference and SD of the differences were 'validated' by assessing the consistency between various reported statistics, where available (i.e. reported SD, 95% LoAs, 95% CIs, paired t-tests and Pearson correlation coefficients). In cases in which LoAs were not explicitly reported, the limits were derived if an appropriate Bland–Altman[121] plot had been presented. However, LoAs were considered valid only if the mid-point corresponded with a validated measure of the mean difference. If there were any discrepancies between any of the reported statistics relating to the SD, then the value that was most prevalent was used. If a value for the SD was explicitly reported, then it was assumed that the reported value was correct and this was used in the meta-analysis. If this process still failed to be conclusive, then the authors were contacted for clarification. However, if no satisfactory SD was reported (or satisfactory statistics from which it could be derived), then mean correlation imputation (previously described) was used.

Intraobserver repeatability coefficients (RCs) were used as the measure of within-subject variance to adjust the meta-analysis for repeated measurements with the same tonometer, as proposed by Bland and Altman.[120] Reported within-subject SDs were converted into RC, where appropriate. If a measure of within-subject variation was not reported, it was imputed using the highest reported value (i.e. the most conservative) from other studies within the same tonometer comparison. Adjustments for repeatability were made for GAT and also for the comparator tonometer (if there was at least one study for the tonometer) reported repeatability statistics.

Data were validated and prepared using Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) and SPSS version 18 (IBM Corporation, NY, USA). Meta-analyses were carried out using the *metan* command in Stata version 11.

Results

This section is broken down into three main parts: an overview of included studies, a summary of agreement across tonometers and, finally, results per candidate tonometer including quality assessment of included studies, the meta-analysis of agreement and results for the other outcomes (recordability, acceptability, practicality and reliability).

Overview

The study selection process is summarised in *Figure 23*. In total, 642 reports were identified from the electronic searches as being possibly relevant. The full text of 189 reports was obtained for assessment: 143 from the electronic searches and an additional 46 from reference lists of the selected studies. Finally, 102 reports met the inclusion criteria, including six RCTs.[122–127]

Characteristics of included and excluded studies

Details of the characteristics of the included studies are provided in *Appendix 3*. A total of 102 studies reporting 130 comparisons involving 11,582 participants (15,525 eyes) were included. The earliest studies took place in 1988[128] and the latest in 2010.[129,130] Fifteen studies took place in the USA,[111,124,127,131–142] nine in Italy,[143–151] nine in the UK,[110,128,129,152–157] seven in Germany,[158–164] six in each of Australia,[165–170] China (excluding Taiwan)[171–176] and Japan,[107,130,177–180] five in each of Belgium[122,181–184] and Switzerland,[104,185–188] four in Spain,[189–192] three in each of Saudi Arabia,[193–195] India,[196–198] Portugal[106,199,200] and Greece[201–203] and two in Taiwan,[125,204] Israel,[205,206] the Netherlands,[207,208] Sweden[209,210] and Turkey.[211,212] One study took place in each of Austria,[126] Brazil,[213] France,[123] Ireland,[214] New Zealand,[215] Norway[216] and Denmark.[217] One additional study was reported as a multicentre study, taking place in Italy and Spain.[218]

After contacting authors to clarify the age of study participants, eight authors from nine studies[135,140,141,148,189,190,192,201,212] confirmed eligibility for inclusion. One study included only participants who were aged ≥ 50 years but did not give the age range.[166]

Seventy-three studies (72%) provided information on gender of the participants, with 3337 women and 2900 men. Fourteen studies provided data on participants' race. Almost all studies reported the diagnosis; populations varied: healthy volunteers,[158] those diagnosed with OHT[155] or glaucoma[196] or a mixed population.[131]

Included studies compared the reference standard tonometer GAT (Haag Streit, Koeniz, Switzerland) with eight different types of tonometer: DCT (PASCAL); RT (Icare); TonoPen; Medtronic Solan, Jacksonville, FL, USA (incorporating Xomed); or Intermedics Intraocular Inc., Pasadena, CA, USA]; Ocuton S; Perkins (Kowa HA-2, Kowa, Japan), NCT (Canon USA Inc., Lake Success, NY, USA; Keeler Ltd., Windsor, UK; NIDEK Co. Ltd., Gamagori, Japan; Reichert Ophthalmic Instruments, Buffalo, NY, USA; or Topcon Corporation, Tokyo, Japan); ORA; transpalpebral tonometer, including the PPT (Proview Eye Pressure Monitor) and the TGDc-01, also known as the Diaton tonometer.

All but three[104,179,187] of the included studies reported sufficient data (or data were provided by the authors when contacted) to be included in the agreement meta-analysis. A total of 27, 20, 17 and 37 studies provided data on recordability, acceptability, practicality and reliability, respectively.
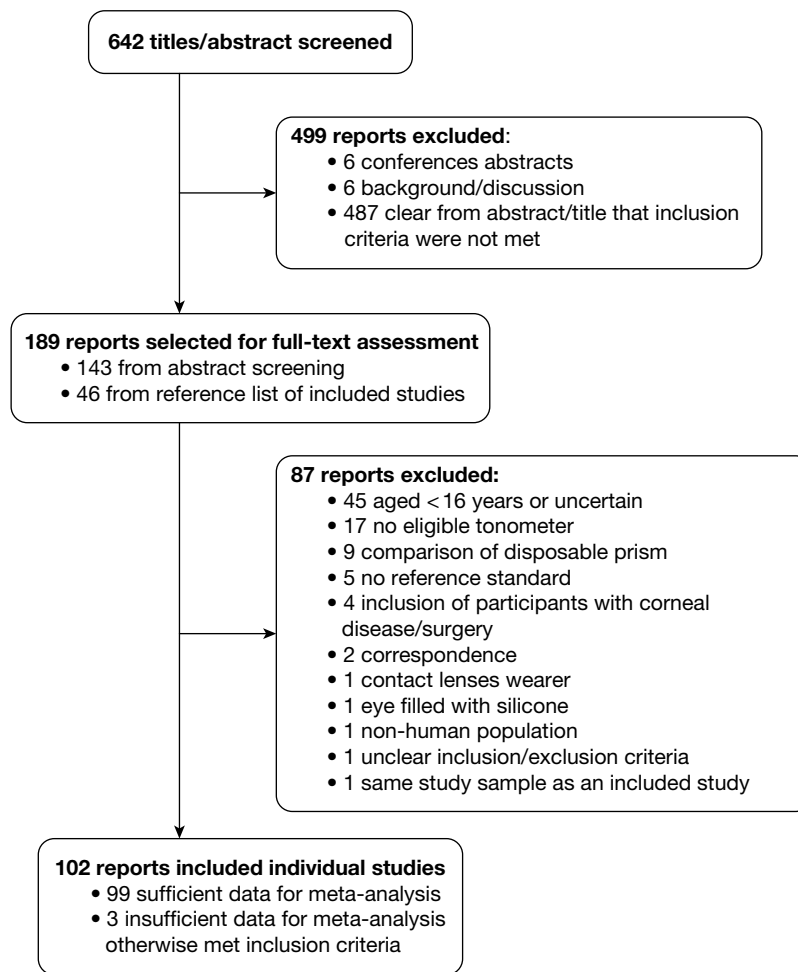
FIGURE 23 Flow diagram of the selection process.

The 87 reports that were excluded at the full-text assessment stage as they failed to meet one or more of the inclusion criteria in terms of study participants, study design, candidate tonometers and reference standard are listed in *Appendix 2*. The tonometers that were not included because they were not commercially available and/or were not suitable for clinical practice were the applanation resonance tonometer,[219] ocular blood flow instrument,[115] Schiotz tonometer,[220] SmartLens (Ophthalmic Development Company, Zurich, Switzerland) tonometer[221] and pneumotonometer.[222]

Summary of agreement across tonometers

Agreement of tonometers with Goldmann applanation tonometer

In total, 99 studies (125 paired comparisons) provided enough data on agreement to be included in the meta-analysis. Comparison across tonometers is difficult given the indirect nature of the analysis. A summary of the main analyses for all candidate tonometers is provided in *Tables 9* and *10*. Full results of all of the meta-analyses for the candidate tonometers are given in *Appendix 4*. The percentage of results that would be within 2 mmHg based on the main analysis mean difference and random error (assuming a normal distribution) is also presented. Based on the analysed studies the expected difference did appear to vary across tonometers, with NCT and Tonopen having the smallest estimated difference and Ocuton S the largest. There was substantial uncertainty for most of the tonometers. In terms of the estimated random error, results varied, with Perkins having marginally (over NCT) the smallest expected random error and Ocuton S the largest. For all tonometers, the 95% LoA stretched from at least 3 mmHg less to 3 units higher

TABLE 9 Pooled estimates and summary 95% LoAs of IOP (mmHg unless otherwise stated)

Comparator	No. of Studies	Mean difference[a]	95% CI	Random error[b]	95% CI	95% LoA	% within 2/3 mmHg
DCT	32	1.8	1.4 to 2.2	2.4	2.1 to 2.6	−2.9 to 6.5	48/67
NCT	26	0.2	−0.1 to 0.6	2.1	1.8 to 2.3	−3.8 to 4.3	66/85
Ocuton S	3	2.7	−1.2 to 6.6	3.5	2.4 to 4.6	−4.1 to 9.6	33/48
ORA	12	1.5	0.9 to 2.2	2.8	2.5 to 3.1	−3.9 to 7.0	46/65
Perkins	4	−1.2	−2.8 to 0.4	2.1	1.3 to 2.8	−5.2 to 2.8	59/79
RT	14	0.9	0.4 to 1.4	2.6	2.1 to 3.2	−4.3 to 6.1	52/72
TonoPen	14	−0.2	−1.0 to 0.5	3.1	2.5 to 3.7	−6.2 to 5.8	48/67
Transpalpebral	20	−0.5	−1.3 to 0.3	3.3	2.8 to 3.7	−6.9 to 5.9	46/64

a Mean difference is mean comparator value minus mean GAT value.
b Random error is the estimated SD of the differences.

TABLE 10 Pooled estimates with 95% prediction intervals of IOP (mmHg unless otherwise stated)

Comparator	No. of Studies	Mean difference[a]	95% prediction interval	Random error[b]	95% prediction interval
DCT	32	1.8	−0.4 to 4.0	2.4	1.1 to 3.6
NCT	26	0.2	−1.4 to 1.9	2.1	0.8 to 3.3
Ocuton S	3	2.7	−4.0 to 9.4	3.5	1.7 to 5.3
ORA	12	1.5	−0.6 to 3.7	2.8	1.6 to 4.0
Perkins	4	−1.2	−4.4 to 2.0	2.1	0.6 to 3.6
RT	14	0.9	−0.9 to 2.7	2.6	0.6 to 4.7
TonoPen	14	−0.2	−3.0 to 2.5	3.1	0.9 to 5.3
Transpalpebral	20	−0.5	−3.8 to 2.8	3.3	1.2 to 5.4

a Mean difference is mean comparator value minus mean GAT value.
b Random error is the estimated SD of the differences.

with Ocuton S and transpalpebral with the widest intervals. With regard to the percentage of measurements that were within 2 mmHg of the GAT value, for most tonometers this value was approximately 50%; Ocuton S at 33% had the lowest value and NCT and Perkins with 66% and 59%, respectively, had the highest. Corresponding values for the percentage within 3 mmHg of the GAT value were 48%, 85% and 79%, respectively, for Ocuton S, NCT and Perkins.

Prediction intervals

Substantial heterogeneity was observed in estimates between studies for most tonometers. The 95% prediction intervals for the mean difference and random error are shown in *Table 10*. The values illustrate the impact of the heterogeneity between individual study estimates of the mean difference: −4.0 to 9.4 mmHg for Ocuton S, whereas for NCT the range of values was only from −1.4 to 1.9 mmHg. For most tonometers except NCT and Perkins, a difference of > 2 mmHg bar was observed. Similarly, the random error 95% prediction intervals illustrate the difference in the level of variability between studies.

Results by tonometer

Dynamic contour tonometer

Quality assessment

Thirty-four studies representing 3726 participants (4933 eyes) compared DCT with GAT. *Figure 24* summarises the quality assessment for these studies. All 34 studies (100%) specified the
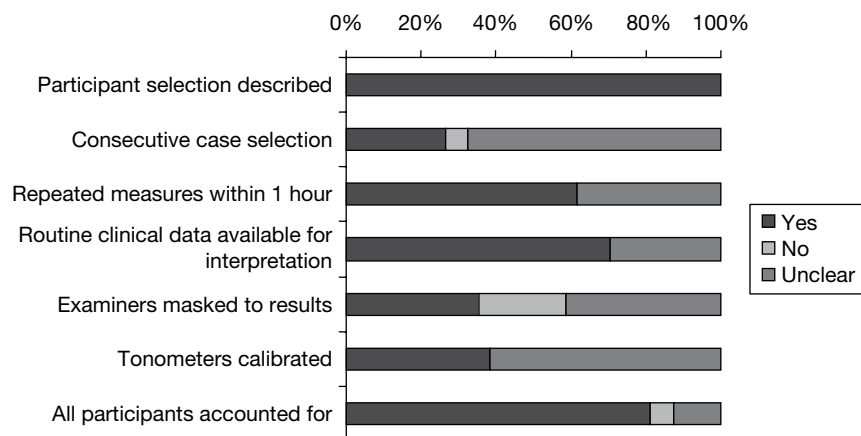
FIGURE 24 Dynamic contour tonometer: summary of quality assessment.

selection criteria. In nine studies (26%), cases were selected consecutively. A total of 21 studies (62%) reported individual measures taken within 1 hour. In 24 studies (71%) the same clinical data were available for interpretation as would be in clinical practice. In total, 12 studies (35%) reported whether the examiner(s) were masked to the results. In 13 studies (38%), the tonometers were reported as calibrated. Almost all of the studies (31; 91%) included all participants approached in the analysis or stated a reason why not. Only three studies[144,148,218] met all of the quality criteria.

Agreement between dynamic contour tonometer and Goldmann applanation tonometer

Thirty-two studies provided sufficient data to include in the meta-analysis. The full results of the agreement analyses of DCT and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was 1.8 mmHg (95% CI 1.4 mmHg to 2.2 mmHg) with a corresponding random error of 2.4 mmHg (95% CI 2.1 mmHg to 2.6 mmHg). For both analyses there was evidence of a large amount of heterogeneity, with very large I^2 values (97% and 95%, respectively), and this can be seen in *Figures 25* and *26*, in which the forest plots are presented. Based on the main analysis the expected mean difference is 1.8 mmHg (95% LoA –2.9 mmHg to 6.5 mmHg).

Recordability
Six studies[134,142,161,189,192,210] provided information on the recordability of the DCT. The data are shown in *Appendix 5*. Individual studies varied in size from 63 to 211 observations. Recordability was high varying from 93% to 100% across the studies.

Acceptability and practicality
Eight studies[123,132,143,144,163,184,210,215] reported on the acceptability and/or practicality of the DCT. The data are shown in *Appendix 6*. For the five studies reporting acceptability, comments were favourable for DCT, and in one study,[215] which measured preference, a substantial proportion (36; 34%) expressed a preference for DCT over GAT, with 55 (52%) having no preference. Only three studies reported on practicality, with all three reporting difficulty with its use: the tonometer was 'not easy to use'[123] and 'entailed a learning curve'[144] or extra measurements were needed.[210]

Reliability
Ten studies[104,123,129,142,150,155,159,181,210,218] reported on either inter- and/or intraobserver reliability of DCT. The data are shown in *Appendix 7*. A variety of reliability measures [LoA, coefficient of variance (CoV), concordance correlation coefficient (CCC), intracluster correlation (ICC) and RC] and types of observers (student, 'experienced' ophthalmologist, optometrist and technician)

DCT study		Mean difference (95% CI)	% weight
Papastergiou 2008[201]		0.20 (–0.30 to 0.70)	3.19
Lanza 2008[143]		2.10 (1.70 to 2.50)	3.25
Johannesson 2008[210]		1.18 (0.79 to 1.57)	3.25
Herdener 2008[159]		0.79 (0.41 to 1.17)	3.26
Eichenbaum 2007[132]		–0.10 (–0.41 to 0.21)	3.29
Salvetat 2007[144]		3.20 (2.73 to 3.67)	3.21
Medeiros 2007[134]		0.39 (0.06 to 0.72)	3.28
Pepose 2007[135]		3.30 (2.65 to 3.95)	3.08
Martinez-de-la-Casa 2006[189]		4.40 (3.98 to 4.82)	3.24
Barleon 2006[161]		0.40 (0.02 to 0.78)	3.26
Ku 2006[215]		1.95 (1.54 to 2.36)	3.24
Kotecha 2005[155]		0.70 (0.21 to 1.19)	3.19
Pache 2005[163]		1.06 (0.54 to 1.58)	3.17
Doyle 2005[123]		0.06 (–0.96 to 1.08)	2.73
Duba 2004[186]		1.90 (1.26 to 2.54)	3.08
Milla 2009[192]		5.43 (4.49 to 6.37)	2.81
Ceruti 2009[148]		2.61 (2.38 to 2.84)	3.32
Realini 2009[141]		1.60 (1.20 to 2.00)	3.25
Detorakis 2010[203]		4.75 (2.18 to 7.32)	1.36
Morita 2010[130]		–0.20 (–0.58 to 0.18)	3.26
Halkiadakis 2009[202]		2.10 (1.62 to 2.58)	3.20
Vandewalle 2009[184]		3.10 (2.46 to 3.74)	3.09
Fogagnolo 2010[218]		2.39 (2.11 to 2.67)	3.31
Sullivan-Mee 2009[142]		2.60 (2.15 to 3.05)	3.22
Pelit 2009[151]		2.47 (2.12 to 2.82)	3.27
Vasconcelos De Moraes 2009[213]		1.16 (0.18 to 2.14)	2.77
Roszkowska 2009[150]		2.00 (1.57 to 2.43)	3.23
Kotecha 2010[129]		2.36 (1.93 to 2.79)	3.23
Oncel 2009[212]		1.60 (1.07 to 2.13)	3.16
Punjabi 2006[131]		1.70 (1.47 to 1.93)	3.32
Liu 2006[171]		1.11 (0.88 to 1.34)	3.33
Pourjavan 2007[181]		1.20 (0.60 to 1.80)	3.12
Overall		1.80 (1.41 to 2.19)	100.00

–7.32   0   7.32

FIGURE 25 Meta-analysis of mean difference between DCT and GAT (main analysis – random effects).

were used, with variable numbers of measurements taken (two or three up to six). Substantial (±3 mmHg or more) interobserver variation was observed in the 95% LoA for two of the studies with the same lead author,[129,155] with a narrower interval for one study (±2 mmHg).[142]

Non-contact tonometer

Twenty-eight studies enrolling 2222 participants (2868 eyes) compared NCT with GAT. *Figure 27* summarises the quality assessment for these studies.

Quality assessment

Selection criteria were specified in 24 studies (86%). In only nine studies (32%) were cases consecutive. A total of 19 studies (68%) reported individual measures taken within 1 hour. In 18 studies (67%), the same clinical data were available for interpretation as would be in clinical practice. In total, 15 studies (54%) reported whether or not the examiner(s) were masked to the results. In nine studies (30%) the tonometers used were calibrated. All bar one study (96%) included all participants approached in the analysis, or stated a reason why not. Only one comparison[156] met all quality criteria specified.

DCT study		Random error (95% CI)	% weight
Papastergiou 2008[201]		1.72 (1.37 to 2.07)	3.20
Lanza 2008[143]		2.24 (1.95 to 2.53)	3.28
Johannesson 2008[210]		2.46 (2.18 to 2.74)	3.29
Herdener 2008[159]		1.37 (1.10 to 1.64)	3.30
Eichenbaum 2007[132]		2.58 (2.36 to 2.60)	3.35
Salvetat 2007[144]		2.40 (2.07 to 2.73)	3.22
Medeiros 2007[134]		2.25 (2.01 to 2.49)	3.34
Pepose 2007[155]		2.68 (2.22 to 3.14)	3.04
Martinez-de-la-Casa 2006[189]		2.60 (2.30 to 2.90)	3.27
Barleon 2006[161]		2.75 (2.48 to 3.02)	3.30
Ku 2006[215]		2.15 (1.86 to 2.44)	3.28
Kotecha 2005[155]		2.86 (2.51 to 3.21)	3.20
Pache 2005[163]		2.67 (2.30 to 3.04)	3.17
Doyle 2005[123]		2.60 (1.88 to 3.32)	2.57
Duba 2004[186]		2.04 (1.59 to 2.49)	3.04
Milla 2009[192]		4.76 (4.10 to 5.42)	2.68
Ceruti 2009[148]		2.04 (1.88 to 2.20)	3.40
Realini 2009[141]		2.00 (1.71 to 2.29)	3.28
Detorakis 2010[203]		6.42 (4.60 to 8.24)	1.08
Morita 2010[130]		1.07 (0.80 to 1.34)	3.30
Halkiadakis 2009[202]		2.43 (2.09 to 2.77)	3.22
Vandewalle 2009[184]		3.11 (2.66 to 3.56)	3.05
Fogagnolo 2010[218]		2.63 (2.44 to 2.82)	3.37
Sullivan-Mee 2009[142]		2.51 (2.20 to 2.83)	3.24
Pelit 2009[151]		2.08 (1.83 to 2.33)	3.32
Vasconcelos de Moraes 2009[213]		3.17 (2.48 to 3.87)	2.62
Roszkowska 2009[150]		1.82 (1.52 to 2.12)	3.26
Kotecha 2010[129]		2.17 (1.87 to 2.47)	3.26
Oncel 2009[212]		2.11 (1.73 to 2.49)	3.16
Punjabi 2006[131]		2.69 (2.52 to 2.86)	3.40
Liu 2006[171]		0.89 (0.73 to 1.05)	3.40
Pourjavan 2007[181]		2.20 (1.78 to 2.62)	3.09
Overall		2.38 (2.15 to 2.60)	100.00

−8.24   0   8.24

FIGURE 26 Meta-analysis of random error between DCT and GAT (main analysis – random effects).
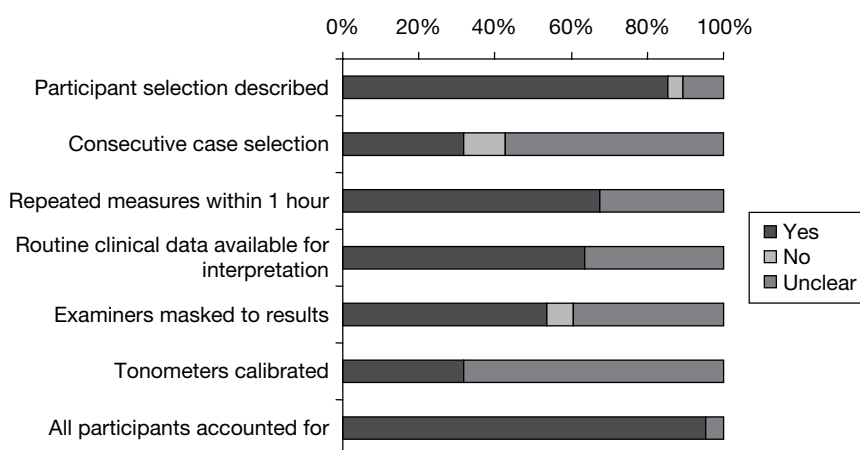
FIGURE 27 Non-contact tonometer: summary of quality assessment.

Agreement between non-contact tonometer and Goldmann applanation tonometer

Twenty-six studies provide sufficient data to be included in the meta-analysis. The full results of the agreement analyses of NCT and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was 0.2 mmHg (95% CI –0.1 mmHg to 0.6 mmHg) with a corresponding random error of 2.1 mmHg (95% CI 1.8 mmHg to 2.3 mmHg). For both analyses, there was evidence of a large amount of heterogeneity with very large I^2 values (95% for both), and this can be seen in *Figures 28* and *29*, in which the forest plots are presented. Based on the main analysis, the expected mean difference is 0.2 mmHg (95% LoA –3.8 mmHg to 4.3 mmHg).

Recordability

Four studies[147,156,182,194] provided information on the recordability of NCT tonometers. The data are shown in *Appendix 5*. Individual studies varied in size from 45 to 100 observations. Recordability was very high (96–100%) for all bar one study (76%).[147]

Acceptability and practicality

Three studies[156,173,194] reported on the acceptability and practicality of NCT tonometers. The data are shown in *Appendix 6*. In one study,[173] 'approximately 50%' of participants appeared to prefer NCT 2000 to Pulsair 2000 or GAT. In another study,[194] 11% of participants expressed anxiety about the NCT Pulsair EasyEye, necessitating a 5-minute period between measurements to allow patients to 'calm down'. Two studies[173,194] found NCT tonometers to be faster than GAT, with NCT 2000 favoured over Pulsair 2000 for ease of use and speed. One study[156] found Pulsair 2000 (mean of 2 minutes) to be faster than GAT and Ao MkII (both with a mean time of 3 minutes).

NCT study	Mean difference (95% CI)	% weight
Papastergiou 2008[201]	0.00 (–0.55 to 0.55)	3.76
AlMubrad 2008[193]	0.30 (–0.04 to 0.64)	4.02
Ogbuehi 2008[194]	0.10 (–0.27 to 0.47)	3.99
Lafaut 2007[122]	2.31 (1.69 to 2.93)	3.66
Garcia Resua 2005[191]	3.21 (2.67 to 3.75)	3.78
Lam 2004[175]	0.10 (–0.74 to 0.94)	3.30
Naruse 2004[178]	–1.60 (–2.05 to –1.15)	3.90
Jorge 2003[199]	0.09 (–0.89 to 1.07)	3.08
Jorge 2002[200]	0.12 (–0.25 to 0.49)	3.99
Garzozi 2001[206]	0.13 (–0.03 to 0.29)	4.16
Popovich 1997[139]	–0.30 (–0.52 to –0.08)	4.12
Cho 1997[173]	0.20 (–0.39 to 0.79)	3.71
Mackie 1996[156]	1.04 (0.29 to 1.79)	3.46
Van de Velde 1995[182]	0.43 (0.07 to 0.79)	4.00
Boles Carenini 1992[147]	–0.50 (–0.89 to –0.11)	3.97
Chauhan 1988[128]	–1.36 (–1.72 to –1.00)	4.01
Regine 2006[149]	1.37 (0.93 to 1.81)	3.91
Murase 2009[180]	0.60 (0.20 to 1.00)	3.96
Ogbuehi 2006[195]	0.20 (–0.18 to 0.58)	3.98
Cheng 2005[176]	0.46 (0.18 to 0.74)	4.08
Pelit 2009[151]	1.39 (0.93 to 1.85)	3.89
Oncel 2009[212]	–0.10 (–0.59 to 0.39)	3.84
Gupta 2006[196]	–0.93 (–1.46 to –0.40)	3.79
Ko 2005[204]	1.00 (0.55 to 1.45)	3.89
Tonnu 2005[75]	–0.70 (–1.18 to –0.22)	3.86
Hansen 1995[217]	–0.92 (–1.38 to –0.46)	3.88
Overall	0.24 (–0.09 to 0.58)	100.00

FIGURE 28 Meta-analysis of mean difference between NCT and GAT (main analysis – random effects).

NCT study		Random error (95% CI)	% weight
Papastergiou 2008[201]		1.92 (1.53 to 2.31)	3.77
AlMubrad 2008[193]		1.40 (1.16 to 1.64)	4.00
Ogbuehi 2008[194]		1.60 (1.34 to 1.86)	3.97
Lafaut 2007[122]		2.80 (2.36 to 3.24)	3.69
Garcia Resua 2005[191]		2.09 (1.71 to 2.47)	3.79
Lam 2004[175]		2.40 (1.80 to 3.00)	3.37
Naruse 2004[178]		1.37 (1.05 to 1.69)	3.90
Jorge 2003[199]		3.30 (2.61 to 3.99)	3.17
Jorge 2002[200]		1.11 (0.85 to 1.37)	3.97
Garzozi 2001[206]		1.02 (0.90 to 1.13)	4.11
Popovich 1997[139]		2.30 (2.14 to 2.46)	4.08
Cho 1997[173]		1.40 (0.99 to 1.81)	3.73
Mackie 1996[156]		3.60 (3.07 to 4.13)	3.51
Van de Velde 1995[182]		1.83 (1.58 to 2.08)	3.98
Boles Carenini 1992[147]		1.65 (1.37 to 1.93)	3.95
Chauhan 1988[128]		1.83 (1.58 to 2.08)	3.99
Regine 2006[149]		2.24 (1.93 to 2.55)	3.90
Murase 2009[180]		1.70 (1.42 to 1.98)	3.94
Ogbuehi 2006[195]		1.50 (1.23 to 1.77)	3.96
Cheng 2005[176]		1.56 (1.36 to 1.76)	4.05
Pelit 2009[151]		2.72 (2.40 to 3.04)	3.88
Oncel 2009[212]		1.94 (1.60 to 2.29)	3.85
Gupta 2006[196]		3.10 (2.72 to 3.48)	3.80
Ko 2005[204]		3.02 (2.70 to 3.35)	3.89
Tonnu 2005[75]		2.52 (2.18 to 2.86)	3.86
Hansen 1995[217]		2.52 (2.19 to 2.85)	3.88
Overall		2.07 (1.82 to 2.32)	100.00

−4.13   0   4.13

FIGURE 29 Meta-analysis of random error between NCT and GAT (main analysis – random effects).

Reliability

Nine studies[122,149,152,156,175,193–195,217] reported a reliability measure. The data are shown in *Appendix 7*. None reported on interobserver reliability. A variety of reliability measures [including CoV, LoA, mean (SD), RC and variance of the difference between the middle reading and the average of the first and last readings (Varmid)] and types of observers (ophthalmologist and optometrist) was used with variable numbers of measurements taken (three to four). Three studies reported substantial 95% LoA for intraobserver reliability and one a substantial RC value.[152,193,195,217]

Ocuton S

Three studies enrolling 173 participants (258 eyes) compared Ocuton S with GAT. *Figure 30* summarises the quality assessment for these studies.

Quality assessment

In all three studies (100%) the selection criteria were specified, the same clinical data were available for interpretation as would be in clinical practice and all participants approached in the analysis were included or a reason was stated why not. On the other hand, none of the studies reported calibration of the tonometers used. Only the study by Sacu and colleagues[126] reported selection of the cases consecutively and reported individual measures taken within 1 hour. In the studies by Marchini and colleagues[146] and Wells[170] the examiner(s) were masked to the results. None of the studies met all criteria specified.

Agreement between Ocuton S and Goldmann applanation tonometer

All three studies[126,146,170] provided sufficient data to be included in the meta-analysis. The full results of the agreement analyses of Ocuton S and GAT are given in *Appendix 4*. Under the main

FIGURE 30 Ocuton S: summary of quality assessment.

analysis the pooled mean difference was 2.7 mmHg (95% CI –1.2 mmHg to 6.6 mmHg) with a corresponding random error of 3.5 mmHg (95% CI 2.4 mmHg to 4.6 mmHg). For both analyses, there was evidence of a large amount of heterogeneity with very large I^2 values (96% and 88%, respectively), and this can be seen in *Figures 31* and *32*, in which the forest plots are presented. Based on the main analysis the expected mean difference is 2.7 mmHg (95% LoA –4.1 mmHg to 9.6 mmHg).

Recordability

Two studies[126,170] provided information on the recordability of Ocuton S. The data are shown in *Appendix 5*. Individual studies varied in size from 68 to 85 observations. Recordability was reasonably high in both studies (82% and 94%).

Acceptability and practicality

Only one study[146] reported on the acceptability and practicality of Ocuton S. The data are shown in *Appendix 6*. Of the 80 participants, 62 (78%) reported no issue, 14 (18%) reported a foreign body sensation and four (5%) complained of burning. Examiners were taught how to use the tonometer by carrying out at least three practice examinations each.

Reliability

Two studies[146,170] reported on intraobserver and interobserver reliability for Ocuton S. The data are shown in *Appendix 7*, with the kappa statistic given for two observers (both intra- and inter-reliability) in one study and a very large RC of 9.2 mmHg reported in the second study.

Ocular response analyser
Quality assessment

Twelve studies enrolling 867 participants (1147 eyes) compared ORA with GAT. *Figure 33* summarises the quality assessment for these studies.

All 12 studies (100%) specified the selection criteria and included all participants approached in the analysis or stated a reason why not. None of the studies selected cases consecutively; it was unclear in most of the cases (92%). Half of the studies (50%) reported details on individual measures were taken within 1 hour. In nine studies (75%) the same clinical data were available for interpretation as would be in clinical practice. Five studies (42%) reported whether the examiner(s) were masked to the results. Only two studies (17%) reported calibration of the tonometers. No studies met all of the specified criteria.

Ocuton S study	Mean difference (95% CI)	% weight
Sacu 2004[126]	2.40 (1.39 to 3.41)	33.27
Marchini 2002[146]	6.30 (5.29 to 7.31)	33.27
Wells 2003[170]	−0.46 (−1.33 to 0.41)	33.46
Overall	2.74 (−1.16 to 6.64)	100.00

−7.31   0   7.31
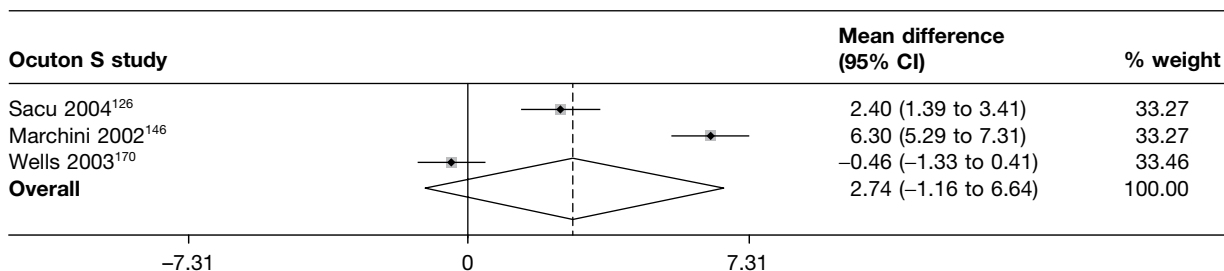
FIGURE 31 Meta-analysis of mean difference between Ocuton S and GAT (main analysis – random effects).

Ocuton S study	Random error (95% CI)	% weight
Sacu 2004[126]	4.13 (3.41 to 4.85)	32.93
Marchini 2002[146]	2.30 (1.59 to 3.01)	32.96
Wells 2003[170]	3.98 (3.36 to 4.60)	34.11
Overall	3.48 (2.36 to 4.59)	100.00

−4.85   0   4.85
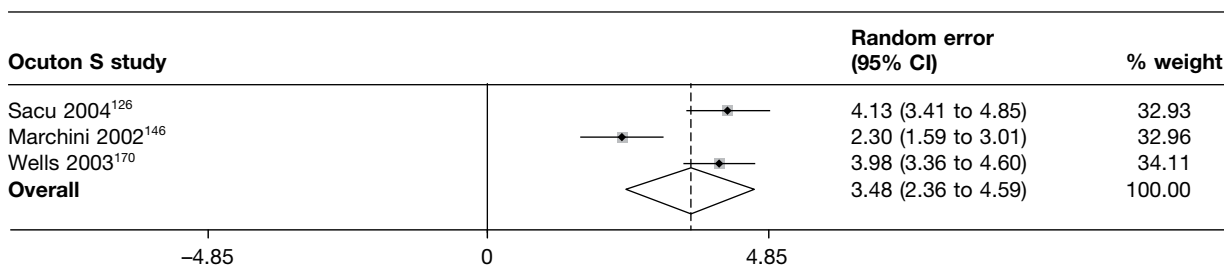
FIGURE 32 Meta-analysis of random error between Ocuton S and GAT (main analysis – random effects).
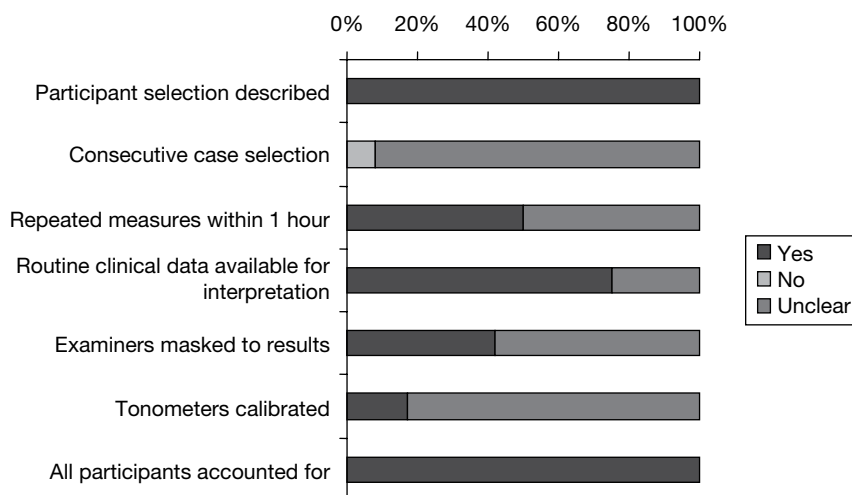
FIGURE 33 Ocular response analyser: summary of quality assessment.

Agreement between ocular response analyser and Goldmann applanation tonometer

Twelve studies[110,111,129,130,135,136,142,174,183,184,212,214] provide sufficient data to include in the meta-analysis. The full results of the agreement analy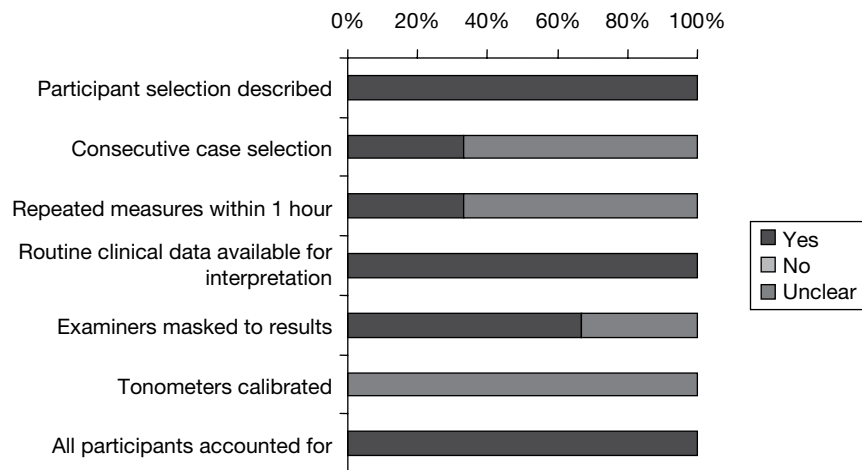ses of ORA and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was 1.5 mmHg (95% CI 0.9 mmHg to 2.2 mmHg), with a corresponding random error of 2.8 mmHg (95% CI 2.5 mmHg to 3.1 mmHg). For both analyses, there was evidence of a large amount of heterogeneity with very large I^2 values (93% and 89%), and this can be seen in *Figures 34* and *35*, in which the forest plots are presented. Based on the main analysis the expected mean difference is 1.5 mmHg (95% LoA −3.9 mmHg to 7.0 mmHg).

Recordability

Two studies[142,185] reported on recordability with 62 of 63 and 49 of 50 participants (98%) having a valid measurement.
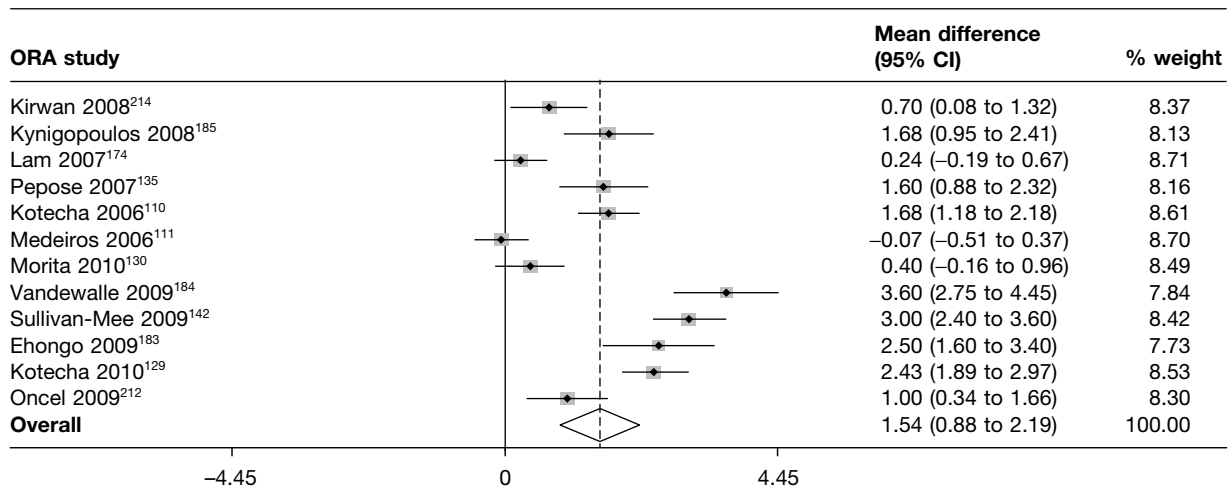
FIGURE 34 Meta-analysis of mean difference between ORA and GAT (main analysis – random effects).
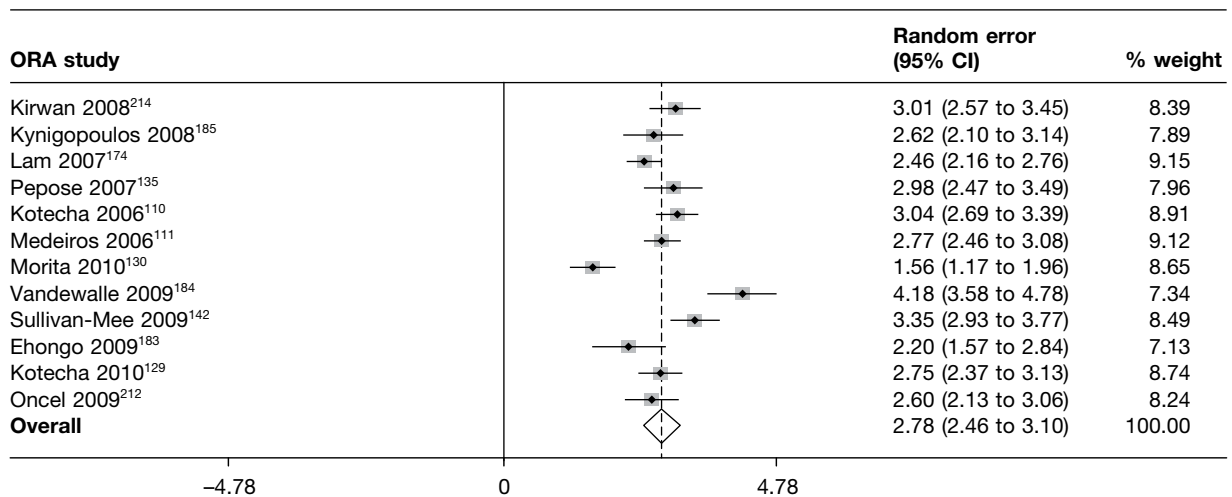
FIGURE 35 Meta-analysis of random error between ORA and GAT (main analysis – random effects).

Acceptability and practicality

One study[184] reported on acceptability and none reported on practicality. Discomfort was reported to be higher for ORA than for RT and DCT.

Reliability

Five studies[110,129,142,183,185] reported a reliability measure. The data are shown in *Appendix 7*. Two studies[129,142] reported on interobserver reliability, showing substantial LoA ±3 or 4 mmHg. A variety of reliability measures (CCC, CoV, LoA, mean difference and RC) and types of observers (ophthalmologist and optometrist) was used with variable numbers of measurements taken (three up to eight).[193,195] One study[142] reported a substantial 95% LoA for intraobserver reliability.

Perkins

Four studies[132,166,191,200] enrolling 433 participants (506 eyes) compared Perkins with GAT. *Figure 36* summarises the quality assessment for these studies.

Quality assessment

All four studies[132,166,191,200] (100%) specified the selection criteria and included all participants assessed in the analysis or stated a reason why not. Three[132,191,200] of the four studies (75%)
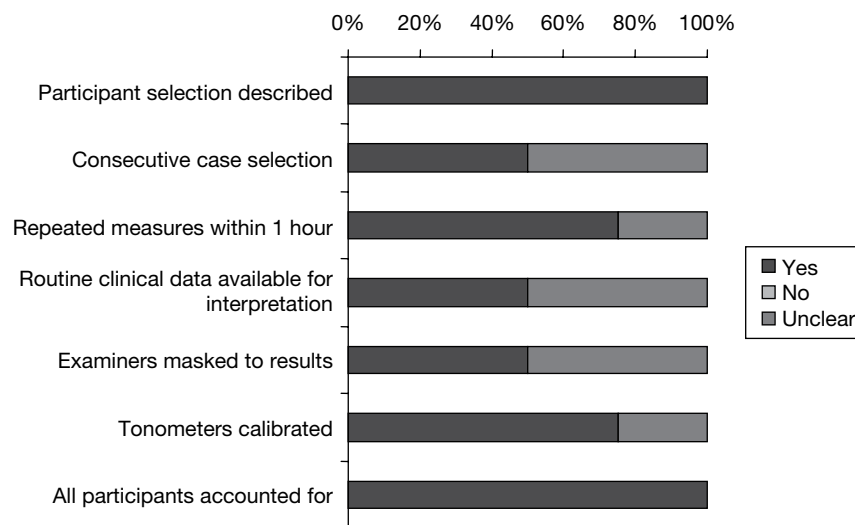
FIGURE 36 Perkins tonometer: summary of quality assessment.

reported individual measures taken within 1 hour, and three studies[132,166,200] (75%) reported calibration of the tonometers. In two studies (50%)[132,166] participants were selected consecutively and the same clinical data were available for interpretation as would be in clinical practice. The examiner(s) were masked to the results in two studies (50%).[166,200] None of the studies met all of the specified criteria.

Agreement between Perkins and Goldmann applanation tonometer

Four[132,166,191,200] studies provide sufficient data to include in the meta-analysis. The full results of the agreement analyses of Perkins and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was –1.2 mmHg (95% CI –2.8 mmHg to 0.4 mmHg) with a corresponding random error of 2.1 mmHg (95% CI 1.3 mmHg to 2.8 mmHg). For both analyses, there was evidence of a large amount of heterogeneity, with large I^2 values (99% and 97%), and this can be seen in *Figures 37* and *38*, in which the forest plots are presented. This finding is driven by one study[166] that had both a much larger mean difference and a much larger random error than the other three studies (see *Appendices 3* and *4* for more information on this study). Based on the main analysis the expected mean difference is –1.2 mmHg (95% LoA –5.2 mmHg to 2.8 mmHg).

Recordability

No studies reported on the recordability of the Perkins tonometer.

Acceptability and practicality

Two studies[132,166] reported on the acceptability of the Perkins tonometer, one[166] of which also reported on its practicality. Perkins was faster to use than GAT (96 vs 120 seconds). The data are shown in *Appendix 6*. No difficulties were observed but it was noted that the patients were 'not new to the practice' in one study.[132]

Reliability

None of the studies reported on the reliability of the Perkins tonometer.

Rebound tomometer

Fourteen studies[106,107,145,154,160,165,167,184,189,190,207,209–211] representing 1239 participants (1792 eyes) compared RT with GAT. *Figure 39* summarises the quality assessment for these studies.

Perkins study		Mean difference (95% CI)	% weight
Eichenbaum 2007[132]		0.00 (−0.22 to 0.22)	25.32
Garcia Resua 2005[191]		−0.42 (−0.85 to 0.01)	24.99
Jorge 2002[200]		0.34 (−0.03 to 0.71)	25.10
Jackson 1995[166]		−4.80 (−5.40 to −4.20)	24.58
Overall		−1.20 (−2.83 to 0.44)	100.00

−5.4   0   5.4
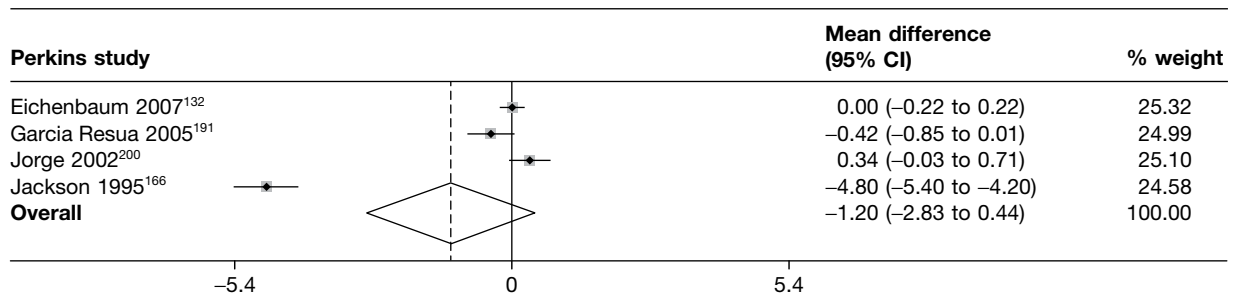
FIGURE 37 Meta-analysis of mean difference between Perkins and GAT (main analysis – random effects).

Perkins study		Random error (95% CI)	% weight
Eichenbaum 2007[132]		1.80 (1.65 to 1.95)	25.74
Garcia Resua 2005[191]		1.68 (1.37 to 1.99)	24.97
Jorge 2002[200]		1.13 (0.87 to 1.39)	25.22
Jackson 1995[166]		3.70 (3.27 to 4.13)	24.06
Overall		2.06 (1.29 to 2.82)	100.00

−4.13   0   4.13
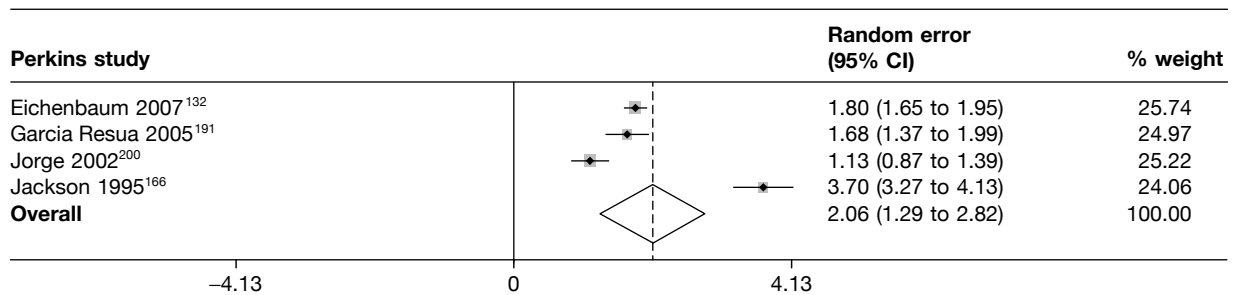
FIGURE 38 Meta-analysis of random error between Perkins and GAT (main analysis – random effects).
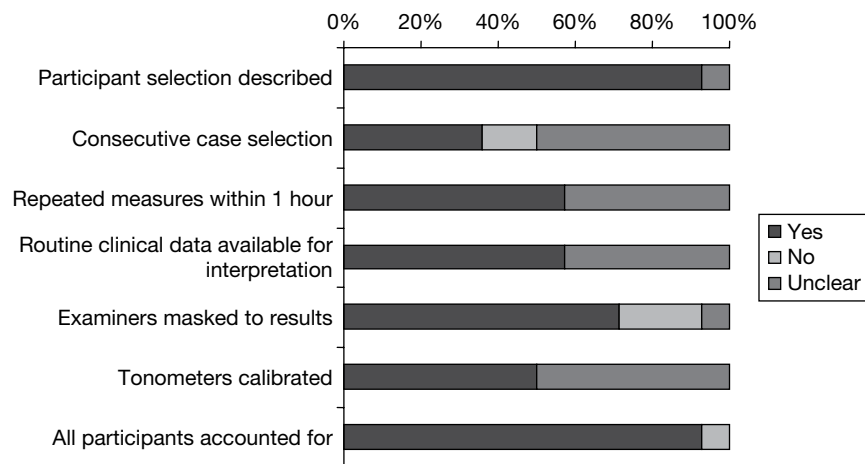
FIGURE 39 Rebound tonometer: summary of quality assessment.

Quality assessment

Nearly all of the studies[106,107,145,160,165,167,184,189,190,207,209–211] (13, 93%) specified the selection criteria. In five studies[145,165,189,190,209] (36%), cases were selected consecutively. In eight studies[106,145,154,165,190,209–211] (57%), individual measures taken within 1 hour were reported and the same clinical data were available for interpretation as would be in clinical practice. Ten studies[106,107,145,154,165,167,184,207,209,211] (71%) reported whether or not the examiner(s) were masked to the results. In seven studies[106,107,145,165,207,209,211] (50%), the tonometers used were calibrated. Almost all studies[106,107,145,154,165,167,184,189,190,207,209-211] (13, 93%) included all participants approached in the analysis or stated a reason why not. Only two studies[145,165] met all of the specified criteria.

Agreement between rebound tonometer and Goldmann applanation tonometer

Fourteen studies[106,107,145,154,160,165,167,184,189,190,207,209–211] provided sufficient data to be included in the meta-analysis. The full results of the agreement analyses of RT and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was 0.9 mmHg (95% CI 0.4 mmHg to 1.4 mmHg), with a corresponding random error of 2.6 mmHg (95% CI 2.1 mmHg to 3.2 mmHg). For both analyses, there was evidence of a large amount of heterogeneity, with very large I^2 values (94% and 98%), and this can be seen in *Figures 40* and *41*, in which the forest plots are presented. Based on the main analysis the expected mean difference is 0.9 mmHg (95% LoA −4.3 mmHg to 6.1 mmHg).

Recordability

Four studies[167,189,190,210] provided information on the recordability of RT. The data are shown in *Appendix 5*. Individual studies varied in size from 36 to 146 observations. Recordability was very low in one study (18; 50%)[190] or 100% in the others.

RT study	Mean difference (95% CI)	% weight
Rehnman 2008[209]	1.52 (0.64 to 2.40)	6.33
Johannesson 2008[210]	2.02 (1.52 to 2.52)	7.31
Pakrou 2008[167]	0.40 (−0.09 to 0.89)	7.31
Ruokonen 2007[160]	1.80 (1.66 to 1.94)	7.83
Sahin 2007[211]	0.43 (−0.21 to 1.07)	6.97
Martinez-de-la-Casa 2006[189]	1.40 (0.96 to 1.84)	7.42
Nakamura 2006[107]	1.40 (0.76 to 2.04)	6.97
Brusini 2006[145]	−1.00 (−1.51 to −0.49)	7.27
Davies 2006[154]	0.50 (−0.20 to 1.20)	6.80
Martinez-de-la-Casa 2005[190]	1.80 (1.35 to 2.25)	7.39
Fernandes 2005[106]	1.34 (0.75 to 1.93)	7.10
van der Jagt 2005[207]	0.60 (−0.04 to 1.24)	6.97
Vandewalle 2009[184]	0.60 (−0.08 to 1.28)	6.87
Abraham 2008[165]	−0.16 (−0.58 to 0.26)	7.46
Overall	0.91 (0.41 to 1.40)	100.00

FIGURE 40 Meta-analysis of mean difference between RT and GAT (main analysis – random effects).

RT study	Random error (95% CI)	% weight
Rehnman 2008[209]	3.00 (2.38 to 3.62)	6.80
Johannesson 2008[210]	3.10 (2.75 to 3.45)	7.21
Pakrou 2008[167]	3.00 (2.65 to 3.35)	7.21
Ruokonen 2007[160]	1.10 (1.00 to 1.20)	7.40
Sahin 2007[211]	2.55 (2.10 to 3.00)	7.08
Martinez-de-la-Casa 2006[189]	2.72 (2.41 to 3.03)	7.25
Nakamura 2006[107]	2.19 (1.74 to 2.64)	7.08
Brusini 2006[145]	3.50 (3.14 to 3.86)	7.19
Davies 2006[154]	2.33 (1.83 to 2.83)	7.01
Martinez-de-la-Casa 2005[190]	2.80 (2.48 to 3.12)	7.24
Fernandes 2005[106]	2.03 (1.62 to 2.44)	7.13
van der Jagt 2005[207]	3.32 (2.87 to 3.77)	7.08
Vandewalle 2009[184]	3.32 (2.84 to 3.80)	7.04
Abraham 2008[165]	2.15 (1.85 to 2.45)	7.27
Overall	2.65 (2.08 to 3.21)	100.00

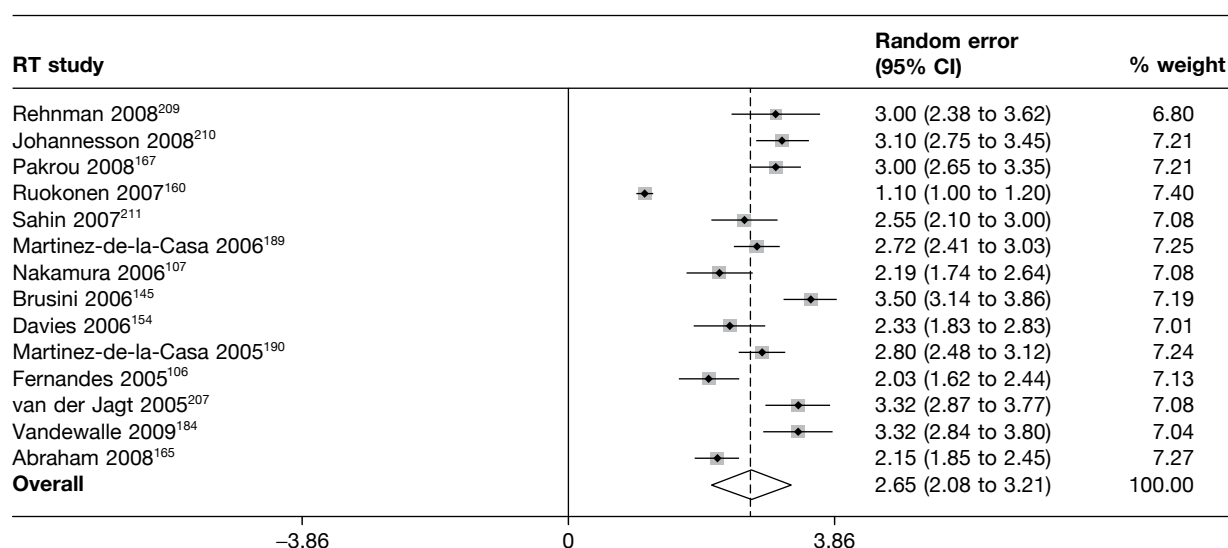**FIGURE 41** Meta-analysis of random error between RT and GAT (main analysis – random effects).

Acceptability and practicality

Seven studies[160,167,184,207,209–211] reported on acceptability and/or practicality for RT. The data are shown in *Appendix 6*. Of the four studies[167,184,207,209] reporting acceptability, comments were favourable for RT, with one study[167] that measured preference showing a large proportion (28; 74%) in favour of RT over GAT (2; 5%), with 18 (21%) showing no preference. Six studies[160,167,207,209–211] reported on practicality, with three reporting its use as 'easy', two noting that additional measurements were required and one that there was no need for anaesthesia and that it used disposable tips.

Reliability

Four studies[154,165,190,210] reported on either inter- and/or intraobserver reliability for RT. The data are shown in *Appendix 7*. A variety of reliability measures (correlation coefficient, CoV, LoA, 'intrasubject variation coefficient' and RC) and types of observers (student, 'experienced' ophthalmologist and optometrist) was used with variable numbers of measurements taken (two up to six). Only one study[190] reported the interobserver correlation coefficient, which was 0.82 for a small number of eyes ($n = 12$). Substantial intraobserver variation was observed in the LoA measures for one study,[154] with a substantial RC for two studies.[165,210]

TonoPen

Fifteen studies[75,107,138,140,144,153,157,166,169,188,191,197,206,207,216] enrolling 1413 participants (1950 eyes) compared TonoPen with GAT. *Figure 42* summarises the quality assessment for these studies.

Quality assessment

All but one study[75,107,138,140,144,157,166,169,188,191,197,206,207,216] (93%) specified the selection criteria. In six studies[144,153,157,166,206,216] (40%), cases were selected consecutively. Nine studies[75,138,140,144,157,169,188,191,216] (60%) reported individual measures taken within 1 hour and 10 studies[75,138,140,144,166,169,188,206,207,216] (61%) that the same clinical data were available for interpretation as would be in clinical practice. Nine studies[107,140,144,153,157,166,169,207,216] (60%) reported whether or not the examiner(s) were masked to the results. In eight studies[75,107,140,144,153,157,166,169] (53%), the tonometers used were calibrated. Almost all studies[107,138,140,144,153,157,166,188,191,197,206,207,216] (13, 87%) included all participants approached in the analysis or stated a reason why not. The Salvetat[144] study met all of the specified criteria.
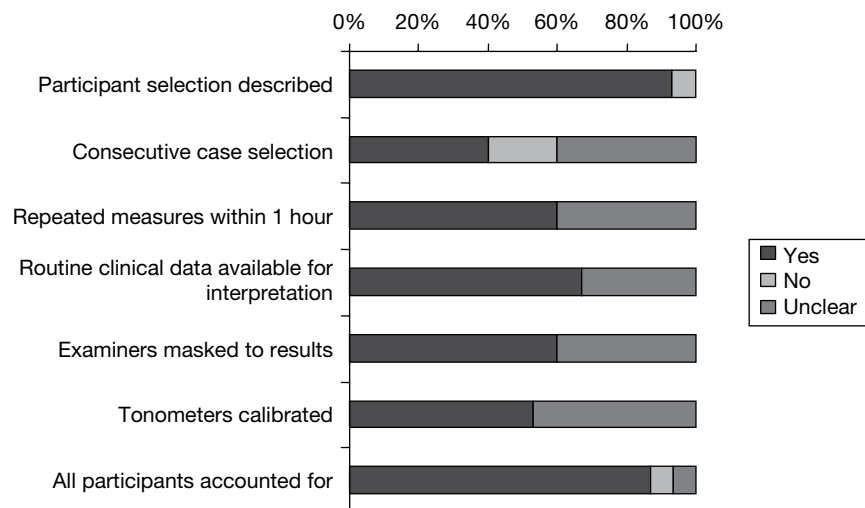

FIGURE 42 TonoPen: summary of quality assessment.

Meta-analysis of agreement between TonoPen and Goldmann applanation tonometer

Fourteen studies[76,138,140,144,153,157,166,169,188,191,197,206,207,216] provided sufficient data to be included in the meta-analysis. The full results of the agreement analyses of TonoPen and GAT are given in *Appendix 4*. Under the main analysis, the pooled mean difference was –0.2 mmHg (95% CI –1.0 mmHg to 0.5 mmHg), with a corresponding random error of 3.1 mmHg (95% CI 2.5 mmHg to 3.7 mmHg). For both analyses, there was evidence of a large amount of heterogeneity, with very large I^2 values (97% and 98%), and this can be seen in *Figures 43* and *44*, in which the forest plots are presented. One study[166] had a much larger mean difference than the other studies, as was the case for the Perkins mean difference for this study. Based on the main analysis the expected mean difference is –0.2 mmHg (95% LoA –6.2 mmHg to 5.8 mmHg).

TonoPen study	Mean difference (95% CI)	% weight
Salvetat 2007[144]	–0.50 (–1.38 to 0.38)	6.82
van der Jagt 2005[207]	0.80 (0.11 to 1.49)	7.07
Garcia Resua 2005[191]	3.54 (2.90 to 4.18)	7.13
Horowitz 2004[169]	–0.41 (–0.86 to 0.04)	7.32
Li 2004[138]	–1.60 (–1.96 to –1.24)	7.38
Bandyopadhyay 2002[197]	1.00 (0.69 to 1.31)	7.41
Garzozi 2001[206]	0.11 (–0.07 to 0.29)	7.48
Schipper 2000[188]	0.40 (–0.34 to 1.14)	7.00
Christoffersen 1993[216]	–0.43 (–0.86 to –0.00)	7.33
Kooner 1992[140]	–0.70 (–1.17 to –0.23)	7.30
Carrim 2009[157]	0.54 (–0.37 to 1.45)	6.78
Jackson 1995[166]	–5.80 (–6.57 to –5.03)	6.97
Bafa 2001[153]	0.60 (–0.24 to 1.44)	6.88
Tonnu 2005[75]	–0.61 (–1.25 to 0.03)	7.13
Overall	–0.22 (–0.97 to 0.54)	100.00

−6.57　　　0　　　6.57

FIGURE 43 Meta-analysis of mean difference between TonoPen and GAT (main analysis – random effects).

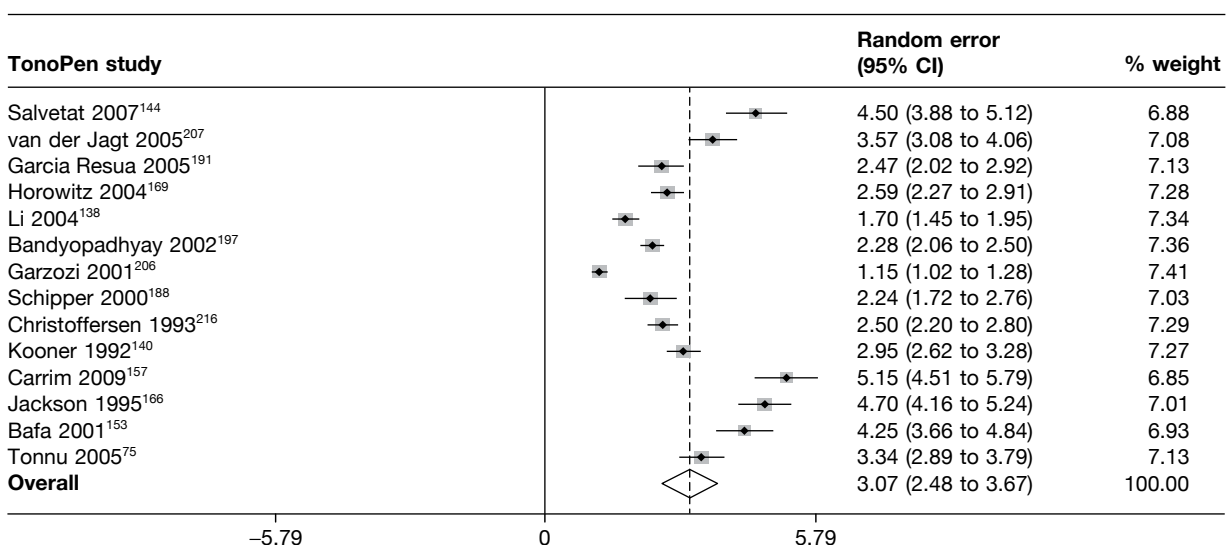| TonoPen study	Random error (95% CI)	% weight
Salvetat 2007[144]	4.50 (3.88 to 5.12)	6.88
van der Jagt 2005[207]	3.57 (3.08 to 4.06)	7.08
Garcia Resua 2005[191]	2.47 (2.02 to 2.92)	7.13
Horowitz 2004[169]	2.59 (2.27 to 2.91)	7.28
Li 2004[138]	1.70 (1.45 to 1.95)	7.34
Bandyopadhyay 2002[197]	2.28 (2.06 to 2.50)	7.36
Garzozi 2001[206]	1.15 (1.02 to 1.28)	7.41
Schipper 2000[188]	2.24 (1.72 to 2.76)	7.03
Christoffersen 1993[216]	2.50 (2.20 to 2.80)	7.29
Kooner 1992[140]	2.95 (2.62 to 3.28)	7.27
Carrim 2009[157]	5.15 (4.51 to 5.79)	6.85
Jackson 1995[166]	4.70 (4.16 to 5.24)	7.01
Bafa 2001[153]	4.25 (3.66 to 4.84)	6.93
Tonnu 2005[75]	3.34 (2.89 to 3.79)	7.13
Overall	3.07 (2.48 to 3.67)	100.00

−5.79　　　0　　　5.79

FIGURE 44 Meta-analysis of random error between TonoPen and GAT (main analysis – random effects).

Recordability

Three studies[197,207,216] provided information on the recordability of the TonoPen tonometer. The data are shown in *Appendix 5*. Individual studies varied in size from 103 to 208 observations. Recordability was high in all three: 90–100%.

Acceptability and practicality

Four studies[144,166,197,207] reported on acceptability and practicality for TonoPen. The data are shown in *Appendix 6*. All four reported favourably on acceptability, with one study,[197] which measured preference, showing a large proportion (64; 32%) preferring TonoPen over GAT (30 ; 15%), and 100 (49%) having no preference. Two studies[144,207] reported that TonoPen was 'easy to use' and two studies[166,197] reported the time to undertake measurement. Bandyopadhyay and colleagues[197] reported a mean (SD) of 54 (18) seconds for TonoPen compared with 15 (4) seconds for GAT, wherease Jackson and colleagues[166] reported means of 50 and 120 seconds for TonoPen and GAT, respectively.

Reliability

Three studies[152,153,169] reported a reliability measure. The data are shown in *Appendix 7*. Only intraobserver data were reported. A variety of measures were reported (CoV, ICC, mean difference, LoA and RC). One study[169] reported a mean difference of 0.7 (95% LoA −3.3 to 3.6) for TonoPen compared with 0.74 (95% LoA −2.3 to 3.7) for GAT. Tonnu and colleagues[152] reported an RC of 4.3 mmHg and 2.2 mmHg for TonoPen and GAT, respectively. The third study[153] stated that the CoV was < 5% for 97/99 eyes whereas 2/99 eyes had a CoV between 5% and 10%.

Transpalpebral tonometer

Twenty studies[124,125,127,133,136–138,158,162,164,168,172,176–178,191,198,205,207,208] enrolling 1509 participants (2091 eyes) compared transpalpebral tonometer with GAT.

Quality assessment

Figure 45 summarises the quality assessment for these studies. A total of 19 studies[124,125,127,133,136–138,158,164,168,172,176–178,191,198,205,207,208] (95%) specified the selection criteria and in six studies[124,125,172,176,178,205] (30%) cases were selected consecutively. In total, 10 studies[124,127,137,138,168,176,178,191,198,208] (50%) reported individual measures taken within 1 hour. In 12 studies (60%) the same clinical data were available for interpretation as would be available in clinical practice[124,127,133,137,138,158,162,168,172,177,205,207] and it was reported that the examiner(s) were masked to the results.[124,125,158,162,168,172,176–178,205,207,208] In seven studies[125,127,172,177,178,205,208] (35%) the tonometers used were calibrated. Almost all
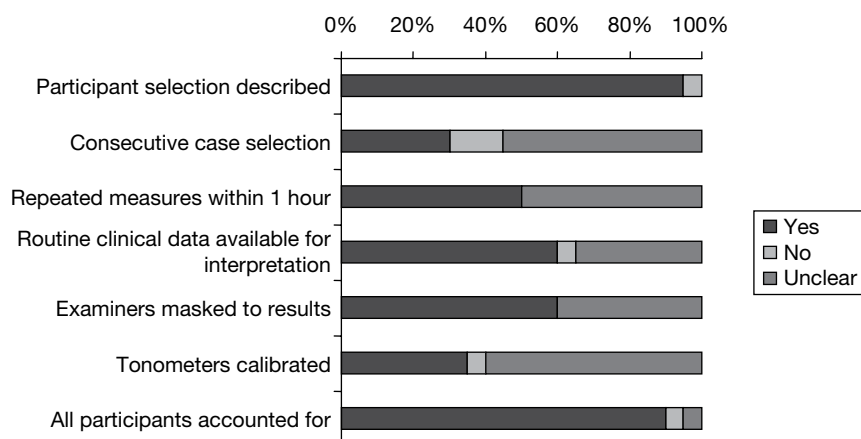

FIGURE 45 Transpalpebral tonometer: summary of quality assessment.

studies[124,127,133,136–138,158,164,168,172,176–178,191,198,205,207,208] (18, 90%) included all participants approached in the analysis or stated a reason why not. No studies met all of the specified criteria.

Meta-analysis of agreement between transpalpebral tonometers and Goldmann applanation tonometer

Twenty studies[124,125,127,133,136–138,158,162,164,168,172,176–178,191,198,205,207,208] provide sufficient data to be included in the meta-analysis. The full results of the agreement analyses of transpalpebral and GAT are given in *Appendix 4*. Under the main analysis the pooled mean difference was –0.5 mmHg (95% CI –1.3 mmHg to 0.3 mmHg) with a corresponding random error of 3.3 mmHg (95% CI 2.8 mmHg to 3.7 mmHg). For both analyses, there was evidence of a large amount of heterogeneity, with very large I^2 values (98% and 97%), and this can be seen in *Figures 46* and *47*, in which the forest plots are presented. Based on the main analysis, the expected mean difference is –0.5 mmHg (95% LoA –6.9 mmHg to 5.9 mmHg).

Recordability

Nine studies[124,127,137,138,168,174,176,177,198] reported on recordability, with all but one study[137] (recordability of 76%) having very high recordability percentages (91–97%). Seven studies[127,137,168,172,176,177,198] reported the reason, i.e. why participants could not see the phosphene. The data are shown in *Appendix 5*.

Acceptability and practicality

Eight studies[125,133,136–138,172,205,207] reported on acceptability and/or practicality. Approaches varied between studies although the data were generally favourable. Two studies[133,136] reported on practicality from a clinical perspective, with one criticising the subjectivity of patient perception of the phosphene. One study[207] suggested that the tonometer was 'easy to use'. The data are shown in *Appendix 6*.

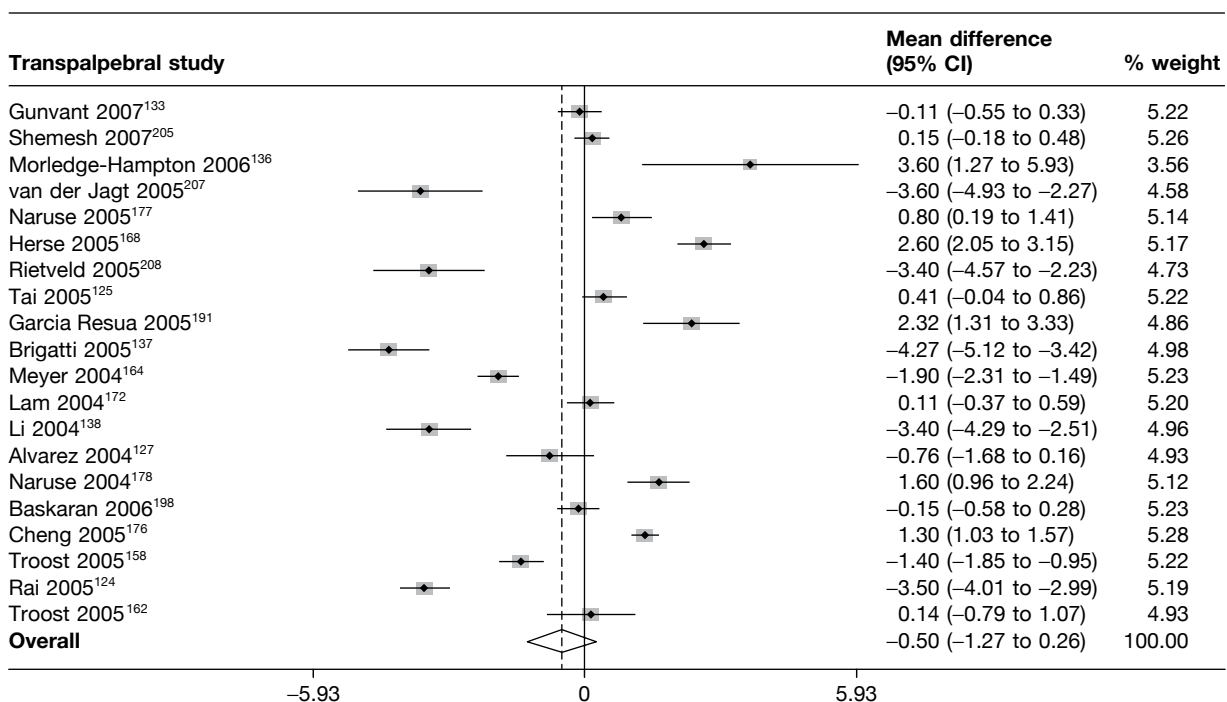| Transpalpebral study	Mean difference (95% CI)	% weight
Gunvant 2007[133]	–0.11 (–0.55 to 0.33)	5.22
Shemesh 2007[205]	0.15 (–0.18 to 0.48)	5.26
Morledge-Hampton 2006[136]	3.60 (1.27 to 5.93)	3.56
van der Jagt 2005[207]	–3.60 (–4.93 to –2.27)	4.58
Naruse 2005[177]	0.80 (0.19 to 1.41)	5.14
Herse 2005[168]	2.60 (2.05 to 3.15)	5.17
Rietveld 2005[208]	–3.40 (–4.57 to –2.23)	4.73
Tai 2005[125]	0.41 (–0.04 to 0.86)	5.22
Garcia Resua 2005[191]	2.32 (1.31 to 3.33)	4.86
Brigatti 2005[137]	–4.27 (–5.12 to –3.42)	4.98
Meyer 2004[164]	–1.90 (–2.31 to –1.49)	5.23
Lam 2004[172]	0.11 (–0.37 to 0.59)	5.20
Li 2004[138]	–3.40 (–4.29 to –2.51)	4.96
Alvarez 2004[127]	–0.76 (–1.68 to 0.16)	4.93
Naruse 2004[178]	1.60 (0.96 to 2.24)	5.12
Baskaran 2006[198]	–0.15 (–0.58 to 0.28)	5.23
Cheng 2005[176]	1.30 (1.03 to 1.57)	5.28
Troost 2005[158]	–1.40 (–1.85 to –0.95)	5.22
Rai 2005[124]	–3.50 (–4.01 to –2.99)	5.19
Troost 2005[162]	0.14 (–0.79 to 1.07)	4.93
Overall	–0.50 (–1.27 to 0.26)	100.00

FIGURE 46 Meta-analysis of mean difference between transpalpebral tonometer and GAT (main analysis – random effects).

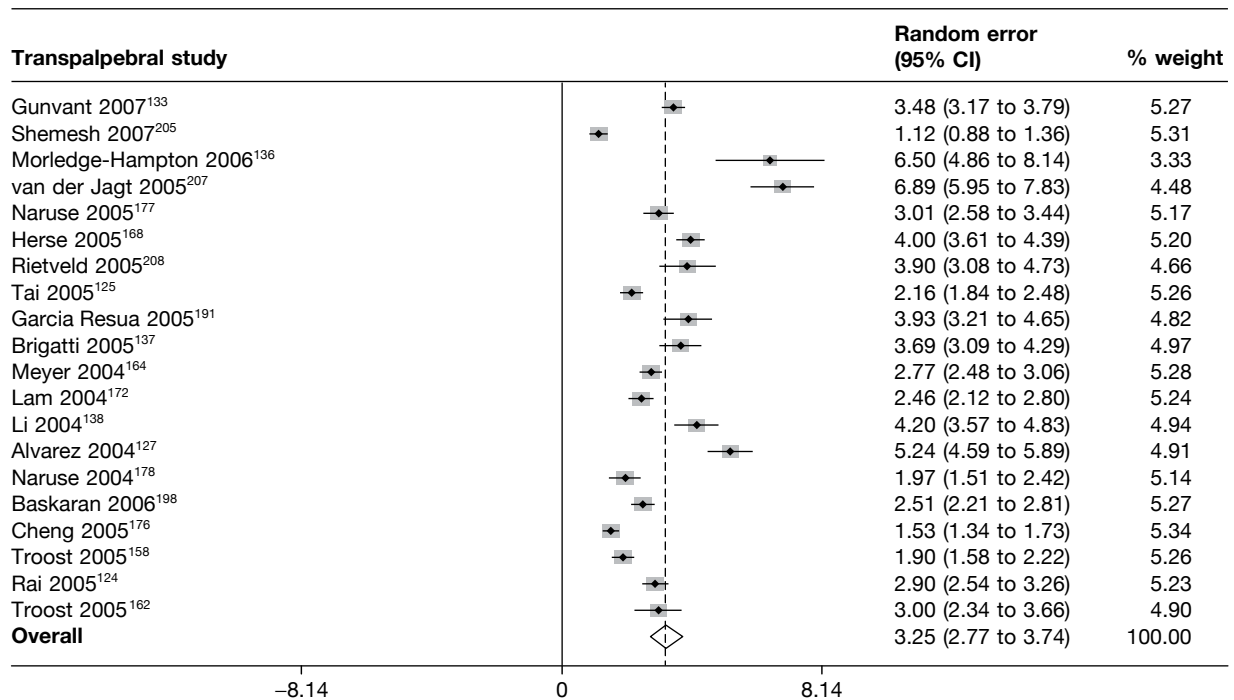| Transpalpebral study		Random error (95% CI)	% weight
Gunvant 2007[133]		3.48 (3.17 to 3.79)	5.27
Shemesh 2007[205]		1.12 (0.88 to 1.36)	5.31
Morledge-Hampton 2006[136]		6.50 (4.86 to 8.14)	3.33
van der Jagt 2005[207]		6.89 (5.95 to 7.83)	4.48
Naruse 2005[177]		3.01 (2.58 to 3.44)	5.17
Herse 2005[168]		4.00 (3.61 to 4.39)	5.20
Rietveld 2005[208]		3.90 (3.08 to 4.73)	4.66
Tai 2005[125]		2.16 (1.84 to 2.48)	5.26
Garcia Resua 2005[191]		3.93 (3.21 to 4.65)	4.82
Brigatti 2005[137]		3.69 (3.09 to 4.29)	4.97
Meyer 2004[164]		2.77 (2.48 to 3.06)	5.28
Lam 2004[172]		2.46 (2.12 to 2.80)	5.24
Li 2004[138]		4.20 (3.57 to 4.83)	4.94
Alvarez 2004[127]		5.24 (4.59 to 5.89)	4.91
Naruse 2004[178]		1.97 (1.51 to 2.42)	5.14
Baskaran 2006[198]		2.51 (2.21 to 2.81)	5.27
Cheng 2005[176]		1.53 (1.34 to 1.73)	5.34
Troost 2005[158]		1.90 (1.58 to 2.22)	5.26
Rai 2005[124]		2.90 (2.54 to 3.26)	5.23
Troost 2005[162]		3.00 (2.34 to 3.66)	4.90
Overall		3.25 (2.77 to 3.74)	100.00

−8.14   0   8.14

FIGURE 47 Meta-analysis of random error between transpalpebral tonometer and GAT (main analysis – random effects).

Reliability

Eight studies[124,127,136,137,158,168,172,177] reported a reliability measure. The data are shown in *Appendix 7*. One study reported on interoberver reliability.[158] A variety of reliability measures (correlation coefficient, CoV, LoA, mean difference and RC) and types of observers (ophthalmologist, patient and technician) was used with variable numbers of measurements taken (3 up to 15). Two studies[137,168] reported a wide 95% LoA for intraobserver reliability and one study[158] reported a very wide 95% LoA for interobserver reliability. One study[127] had an RC of 5.1 mmHg and another of 4.2 mmHg.[168]

Discussion

Overview

This review was conducted to systematically evaluate the properties of a range of IOP measurement devices (tonometers) and their suitability to replace the current reference standard (i.e. GAT) for surveillance of patients with OHT in a primary-care setting. We identified 102 studies comparing eight candidate tonometer types against GAT. Studies included a variety of individuals, both patients and healthy subjects and patients with treated and untreated OHT and glaucoma. Poor reporting limited the assessment of the quality of the included studies. Most studies did not provide sufficient details to assess all of the criteria. In particular, many studies did not state whether or not cases were consecutive, whether or not the tonometers were calibrated, if examiners were masked to results and the time period within which repeated measurements were taken. Despite comprehensive guidance being available on how to undertake measurement of IOP in clinical studies,[223] a substantial proportion of studies appear to adopt suboptimal practice, or at least reporting of such aspects.

Agreement

The results of this study suggest that, compared with GAT, NCT was the tonometer with the most agreement for measuring IOP. Almost two-thirds of measurements with NCT were estimated to be within 2 mmHg of the GAT value and > 80% were within 3 mmHg. Next best was the Perkins tonometer (59% and 79% of measurements, respectively), which was not surprising because it is also an applanation tonometer based on the Goldmann principle. Perkins has the same advantages and limitations as GAT, the only substantial difference being that Perkins is a portable instrument. Other tonometers had about half or less of measurement differences within 2 mmHg or about two-thirds within 3 mmHg. Ocuton S appeared to have the lowest agreement with GAT, with only a third of measurements within 2 mmHg and half within 3 mmHg.

Recordability

Recordability was reported for all tonometer categories except for Perkins. Disappointingly, many studies did not explicitly state the number of participants in whom a measurement was attempted, as opposed to the number in whom a measurement was successfully taken. In general, reported recordability was moderate to very high, with most studies reporting values of ≥ 90%. For one RT study of only 36 participants, recordability was low, at 50%. For NCT, Ocuton S and transpalpebral tonometer, values in the range of 70–90% were observed in a single study, which could be considered problematic if representative of a monitoring scenario.

Acceptability and practicality

Acceptability data were available for all eight tonometers. Because of the variable nature of reporting it is difficult to provide a precise summary. Overall, the data suggest that the candidate tonometers were as acceptable to patients as GAT, if not more so. Marchini and colleagues[146] reported short-term side effects (foreign body sensation or burning) associated with Ocuton S in a sizeable number of participants. Ogbuehi and AlMubrad[194] reported some anxiety about the procedure, with time breaks of 5 minutes used between measurements. Patient preferences were stated for DCT, TonoPen, RT and NCT over GAT in four studies that assessed this.[167,173,197,215]

Reliability

Some reliability data were reported for all except the Perkins tonometer; however, data were available for both inter- and intraobserver reliability for only five of the eight tonometers. The number of measurements used to determine reliability varied across studies, as did the statistical measure used. Greater consistency of reporting would be an aid. Either LoA or RC should, in our opinion, always be reported. Nevertheless, there was a clear suggestion of sizeable inter- and intraobserver variability for all seven tonometers for which data were available. It is worth noting that GAT reliability, although often lower than the candidate tonometer values, was also usually sizeable. This would to some extent explain the scale of heterogeneity observed in the agreement meta-analyses, although the use of repeated measurements for both GAT and the candidate tonometer should have lessened the impact.

Clinical issues

Tonometers that could potentially be used to measure IOP in a monitoring setting were eligible for this review. They are based on different principles, and were categorised as contact and non-contact. Another possible classification would be those that are automated compared with those that require investigator judgement. A subgroup of technologies has been designed to enable users to measure their own IOP.

Although GAT is the device currently used in secondary care, and recommended by NICE to diagnose and monitor OHT, it is relatively complex to use, and is influenced by the investigator's experience.[1] There are other potential limitations of GAT. Applanation tonometry is based on

the Imbert–Fick law, which states that the force to applanate (flatten) the anterior corneal surface is equal to the true IOP *multiplied by* the applanated area at the posterior corneal surface.[224] However, although the applanated area of the anterior surface of the cornea is constant the CCT and the rigidity and resistance to a deforming force vary among individuals. Previous corneal refractive surgery for myopia can lead to measurement errors with GAT by thinning and fattening the cornea and changing its rigidity. The tear film can also affect IOP readings with GAT. The surface tension of the tear film attracts the prism to the cornea, and may cancel the force needed to flatten the cornea. The precise effect of the tear film on IOP readings is not fully understood.[25] GAT implicitly assumes a CCT value in its calculation. Some of the candidate tonometers (e.g. ORA and DCT) purport to compensate for the patient's CCT and other corneal properties. Finally, GAT has been shown to underestimate IOP compared with manometry.[114] Together, the above issues seem to have led, at least in some studies, to GAT measurements having substantial imprecision, which in turn will have contributed to the substantial random error observed between candidate tonometers and GAT measurements. Reliability data for GAT support this understanding, as noted above.

Two independent experts suggested that a difference of ≥ 2 mmHg would be clinically relevant and might influence the clinical management of patients with OHT. To help in the interpretation of the variability results we explored the proportion of measurements with a difference of ≤ 2 mmHg or 3 mmHg. Others may view a larger difference as acceptable.[90]

Methodological issues

We chose to include only studies that compared a candidate tonometer against GAT, which we used as a reference standard. In principle, this should have provided some consistency across comparisons, although the results perhaps suggest that this standard, although widely accepted, is somewhat variable in implementation. Implicitly, any contrast between studies is an indirect comparison and suffers from the limitations of such approaches, such that the observed difference may reflect at least to some degree the differences between the studies (e.g. population and observers) that contribute to each comparison. As a result, comparisons between candidate tonometers should be made cautiously.

Although LoA studies are very common, methods for the meta-analysis of such studies have been proposed only fairly recently.[118] The approach outlined here was to generate a summary or 'pooled' LoA. The mean difference and random error (SD) of individual studies were pooled and 95% LoA generated from pooled estimates. This provides a summary of the evidence across multiple studies that assess the same comparison. Our approach, following that of the original LoA approach for primary studies[225] and the meta-analysis approach of Williamson and colleagues,[118] assumes an underlying normal distribution for the differences to generate the LoA. This is often at least approximately true for the differences and may still be a reasonable approach even when the assumption is violated in individual studies.[120]

An important finding of the review was the large-scale heterogeneity between the results of individual studies that assess the same comparison. This was unlike the original example in which this methodology was used.[117] Sensitivity analyses and subgroup analyses were undertaken to seek to identify sources of heterogeneity although with little light generated. The scale of the heterogeneity can perhaps be most clearly seen in the prediction intervals, which illustrate a plausible range of values for the mean difference and random error of individual studies for each comparison. For only one of the eight tonometers (NCT) the prediction intervals did not include a difference of ≥ 2 mmHg. Between-study heterogeneity, which reflected differences in study population and methodology, was clearly substantial. Nevertheless, in the absence of clarity of factors that contribute to the IOP measurement the prediction intervals show the uncertainty and the scale of difference that is plausible if a new study were to be carried out.

There were a number of limitations in the reporting of individual studies that limited the extent to which we could accurately represent the evidence. A number of studies included more than one eye per participant, which resulted in clustering of intraindividual data.[90] Where possible, we used only one eye per participant, but we chose not to exclude the study if this was not possible. The scale of the clustering effect is uncertain, but our sensitivity analyses, which included only studies with one eye per participant, did not suggest a substantial impact on the estimate of the random error. Four studies did not even provide sufficient information to estimate the mean difference and the within-person difference for all the tonometers, even when allowing imputation of the correlation coefficient. A number of reports did not report any measure of variability for within-person differences and it was necessary to impute a correlation. The sensitivity analyses showed that the results were not unduly influenced by this assumption.

Another limitation for conducting a systematic review was that studies used varying numbers of observations both for the candidate tonometer and for the reference standard. In some cases it was not clear how many measurements had been taken because of ambiguous reporting. We sought to address the impact of variation in the number of measurements in two ways. First, we used the SMD as the effect size given that, in theory, such variability could be addressed this way. Second, where an estimate of the within-person SD was known, we used this value to adjust the variance to match the situation with a single observation of both the candidate tonometer and GAT. If this value was not reported for a particular study but it was reported for another within the same comparison, then the value from the other study was used. Although this did change the estimate of the random error, the magnitude of the change was not large and it did not suggest that it was the primary cause of observed inconsistency in estimates.

Clinical factor analyses

Prespecified analyses looked at the impact of IOP, CCT, refractive surgery and examiner type on the LoA. Because of the magnitude of observed heterogeneity, an additional subgroup analysis investigating the impact of manufacturer and excluding low-quality studies was undertaken. All of the subgroup analyses reflect the limitations in the reported data, although less markedly so for manufacturer, which was generally well reported. Overall, the analyses were inconclusive. The scale of heterogeneity made any subgroup analysis susceptible to spurious differences when only a small number of studies were in a group. In a few studies there were suggestions that IOP was an influencing factor in the observed differences, although this was not a universal finding and appeared to vary by tonometer. Refractive surgery appears to impact on the difference for the transpalpebral tonometer. There was no clear pattern regarding examiner and manufacturer. It was not possible to undertake a subgroup analysis of CCT even on the basis of a crude dichotomy of the group for any of the studies. The exclusion of (based on reporting) lower-quality studies similarly did not provide clarity in this regard, although this perhaps reflects the substantial amount of non-reporting of key information in the studies. Ideally, IPD would have been available to form subgroups, for example according to IOP and CCT. This would allow more informative investigation of the influence of these factors at a review level.

Further research

There is a need for a reporting standard tailored for method comparison studies of tonometers building on recent work in this area.[226] Reporting is inconsistent, and basic let alone desirable data are often not presented. The quality assessment highlighted a lack of reporting of key study characteristics and the fact that the issues of clustering of eyes with participants and the number of observations used is regularly ignored. Furthermore, an in-depth study of factors that could influence the pressure measurements is needed for the reference standard and candidate tonometers. This could either be a large primary study or potentially take the form of an IPD meta-analysis.[118] Given the level of heterogeneity, it may be the case that a systematic review of LoA studies requires very focused study inclusion criteria akin to those recently proposed for

diagnostic test accuracy.[227] Further studies evaluating the agreement of the Perkins and Ocuton S tonometers would also be beneficial given that only a small number of studies have been carried out. Finally, further evaluation of the role of GAT as the default tonometer in clinical practice is warranted.

Conclusions

A variety of tonometers are used to evaluate IOP, and GAT is the current reference standard. The NCT or Perkins tonometer appears preferable if the aim is to achieve as close to a GAT measurement as possible. However, the findings cast doubt on the validity of GAT as the default standard. Consistent use of the same tonometer during clinical follow-up is arguably almost as important as the choice of tonometer.

Chapter 6

Optimal frequency of monitoring of intraocular pressure and visual fields

Introduction

Clinical management of individuals with OHT is based on the measurement and interpretation of initial IOP levels as well as monitoring to detect the onset of OAG, hereafter referred to as glaucoma. The initial monitoring is to establish whether or not treatment is required to lower the IOP and, subsequently, monitoring is to establish the response to treatment.

It is widely accepted that, in addition to IOP measurement, monitoring to detect glaucoma should include a measure of the visual field, commonly by SAP, and an evaluation of structural changes in the optic nerve and RNFL to detect the tissue loss that is characteristic of glaucoma.

When any quantity (such as an individual's IOP) is measured over time, the observed value can be thought of as a product of the true underlying value and its variability. This variability is the result of genuine short-term fluctuations such as biological variability and also the inherent imprecision in the method of measurement. Additionally, over time, the true IOP level may change, and it is this change in the underlying level of IOP that a monitoring programme is seeking to detect. The short-term variation (noise) in the IOP measurement can mask the 'true' IOP level and thereby inhibit detection of a change in IOP. The long-term change will vary between individuals and this can be described as long-term between-individual variability. The majority of individuals being monitored may present with a chronic 'stable' condition, and therefore the signal can be thought as the divergence of a few individuals from the population average. The same principles apply to monitoring of the visual fields to detect glaucoma, in which the variability of visual field indices can be described. To determine the optimal frequency of monitoring it is necessary to quantify both short- and long-term variability in measurement of the test parameter for individuals in the target population.[228]

The signal-to-noise ratio is a useful indicator of the magnitude of long-term variability in relation to the noise. A high signal-to-noise ratio means that the changes observed are likely to be larger than the noise and hence represent a true change, whereas a low ratio indicates that the change observed is likely to be caused mainly by the noise. Clinical decisions need to be based on a true change (high signal-to-noise ratio) instead of being driven by spurious short-term changes (noise).

In choosing the monitoring frequency, there is a trade-off between predominantly measuring noise and the early detection of those at increased risk (signal). Previous work in other chronic conditions suggests that, in the short term, the signal can be relatively weak in comparison with the noise.[228,229] In the context of monitoring subjects with OHT, decisions on long-term management based on short-term measurement could lead to over- or undertreatment of OHT. Hence, quantifying the noise and the signal levels present will help inform the optimal frequency for monitoring.

Choosing monitoring tests and criteria

When choosing monitoring tests and criteria four factors should be considered:[228]

1. Clinical validity: the monitoring pathway (including test types and frequency of testing) should be either a measure of the clinically relevant outcome or a good predictor of the clinically relevant outcome.
2. Responsiveness: the test should change promptly in response to changes in the underlying value (e.g. after therapy).
3. Signal-to-noise ratio: the test should differentiate clinically important changes over time from background measurement variability (short-term biological fluctuations and technical measurement error).
4. Practicality: ideally the test should be non-invasive, cheap and simple to perform; test results should be immediately available and suitable for patient self-monitoring.

Although, in terms of clinical validity, measuring the visual field (by SAP) is the widely accepted test to detect the clinical outcome of glaucoma, IOP is also valid as it is the measure on which treatment decisions are made. In relation to responsiveness, clinical trials have shown that it is possible to detect a rapid reduction in IOP levels after treatment.[32] In addition, IOP is a measure for which it is possible to determine its monitoring characteristics (signal and noise), although these have not been determined previously. Finally, in terms of practicality, both IOP and visual fields are generally measured during the same session, making them equally practical.

There is currently insufficient evidence to guide clinicians with respect to the optimal frequency of monitoring necessary to identify clinically significant changes in IOP or the optimal frequency and spacing of testing required to detect clinically significant changes in the visual field over a relevant time period, which could be up to 5 years.

The aim of the study was to quantify the signal-to-noise ratio of IOP and visual field measures (MD and PSD), thereby informing a monitoring pathway. Data from two randomised placebo-controlled trials were used to estimate the short-term variability (noise) and long-term true change and variability (signal).

Objectives

The objectives were to:

1. determine the optimal monitoring criteria (tests, frequency of monitoring, and determinants of progression to glaucoma) by:
    i. estimating the 'noise' in IOP and visual field measurements (measurement error and biological variation)
    ii. estimating the 'signal' in IOP and visual field measurements (change over time plus long-term variability)
    iii. identifying the optimal number of measurements required to quantify the true IOP
    iv. estimating the signal-to-noise ratio in IOP and visual field measurements given different risk of OAG
2. provide estimates to inform the economic modelling evaluation of the impact of different plausible surveillance regimes for monitoring OHT (see *Chapter 8*)
3. identify future research needs.

Methods

Our methods involved the secondary analysis of data from randomised placebo-controlled trials involving participants with OHT with repeated monitoring over a long-term follow-up. These trials should provide suitable data for quantifying the signal-to-noise ratio of IOP and visual fields. The placebo group consists of a homogeneous group having long-term follow-up. This highly controlled group can then be used to estimate how IOP changes over time, estimate the 'noise' and determine the impact that the small number of individuals with rapid IOP changes has over the population (variation component of the 'signal'). The basis of this analysis relies on estimating the within-person variability, the between-person variability and the natural trend of IOP as well as the variability of other clinically relevant measures of glaucoma (visual field variables MD and PSD) within specified periods. These estimates are determined by fitting statistical models to IOP measurements over time for a group of individuals. Two statistical models were used to obtain these estimates: (1) a direct method and (2) a linear random-effects model, both of which are discussed in *Modelling variability*. A more detailed description of possible models of IOP or visual field parameter changes over time, taking into account both measurement change and long-term true change, is provided in *Appendix 8*.

Data sets used in the analysis

Data from two RCTs carried out at Moorfields Eye Hospital, London, and Rotterdam Eye Hospital, the Netherlands, were used for the analyses; a description of the data sets is provided in *Chapter 3*.

Moorfields data set

The Moorfields cohort included data from participants randomised between 1992 and 1996. The following variables were available for each participant and used in this analysis:

1. *Baseline characteristics:* participant identifier (ID), observer (clinician measuring the IOP), age, gender, date after randomisation, eye (under study), IOP, MD, PSD and medication.
2. *Conversion to glaucoma:* eye and visit date of conversion to glaucoma.
3. *Repeated measures at each 4-monthly visit (1–14):* visit date (recoded as days from baseline visit), IOP, MD, PSD and medication.

Rotterdam data set

The Rotterdam cohort included data on participants randomised between 1997 and 2001. The following variables were available for each participant and used in this analysis:

1. *Baseline characteristics:* participant ID, date after randomisation, IOP and medication.
2. *Conversion to glaucoma:* repeated measures at each 6-monthly visit (0–16).
3. *Visit number, IOP and medication:* as the number of days between visits (or dates) was not recorded, this was estimated by determining the total number of days from baseline to final follow-up for each participant divided by the total number of visits. This defined the average days between visits per participant, which was used for the analysis.

Both data sets included data on both eyes. In the Moorfields data set, one eye per patient was selected for analysis; this was based on the eye that was chosen in the original study (indicated in the data set). In the Rotterdam data set, one eye per participant was selected randomly for analysis. In both data sets, only IOP measurements obtained with GAT were included in the analysis. The analyses use information from the placebo group only, which was formed by participants who had been allocated to receive either placebo or no medication within the trial period. Participants in the placebo group who were put on treatment during the

study period (crossovers) were treated as censored and contributed information only during the non-medicated period with their data treated as missing for the rest of the study (see *Imputation assumptions*).

Modelling variability

As the data arose from placebo-controlled RCTs, the apparent change in IOP and visual field measurements over time can be thought of as made up of three components:

1. the *long-term (true) change* in IOP and visual field measures of the whole group
2. *short-term variability*, that is, a combination of analytic variability and a week-to-week biological fluctuation around a stable average
3. *long-term variability*, that is, a variation in long-term true change between individuals (as would be seen with the theoretical average of a large number of measurements per individual).

All three components were modelled using two different methods: (1) a direct method and (2) a linear random-effects model. For monitoring purposes, the short-term variability is equal to the 'noise' in the measurement whereas the combination of the average true long-term change and the long-term variability are equal to the 'signal'.

Direct method

The average change of the group can be estimated from the group average at each time point. Therefore, a direct method for estimating the long-term variability uses the variance of the differences between the baseline value and each subsequent time point *i*. Taking IOP as an example, this can be calculated as:

$$\sum(IOP_{\text{time } i} - IOP_{\text{baseline}})^2 \qquad\qquad \text{[Equation 8]}$$

The time *i* corresponds to the different follow-up times at which data were recorded – 4-monthly for the Moorfields RCT and 6-monthly for the Rotterdam RCT – with a maximum of 10 years' follow-up for either trial. The baseline IOP was taken as the first visit measurement for the Moorfields data set and the second visit for the Rotterdam data set, to reduce the potential noise created by the initial response to therapy as well as any potential 'regression to the mean' effect. The variability of change is used as a proxy for variation from the true mean at different time points. Therefore, using linear extrapolation on the variation of change it is possible to estimate short-term variation in IOP (noise), which would be proportional to the variance at time point 0.[228] The long-term variability at different time points (signal) can be estimated by subtracting the short-term variability from the variability of the change.

The direct estimation method was used to explore the variability of IOP and visual field variables, specifically MD and PSD.

Linear random-effects model

In addition, the three variability components of IOP were estimated using a linear random-effects model. The main assumption of this model is that the IOP (or some transformation of IOP) for each individual can be adequately described as having a linear change (increase/decrease) over time; importantly, it allows this rate of change to vary between individuals. This model can be written as:

$$\text{Transf}(IOP_i) = \alpha + \beta \times (\text{days}_0) \qquad\qquad \text{[Equation 9]}$$

for each individual, where α is the baseline level (in the transformed IOP scale) and β is the change in IOP for every day that passes from the first measurement (days$_0$ = number of days since first visit). The observed IOP will have a measurement error (ε) and this model can be extended to the whole population using a linear random-effects model on all relevant parameters so that for an individual j at a given time i:

$$\text{Transf}\left(\text{I0P}_{ij}\right) = \alpha_j + \beta_j \times \left(\text{days}_0\right) + \varepsilon_{ij} \qquad \text{[Equation 10]}$$

$$\alpha_j \sim N\left(a, \sigma^2{}_\alpha\right), \beta_j \sim N\left(b, \sigma^2{}_\beta\right), \varepsilon_{ij} \sim N\left(0, \sigma^2{}_W\right)$$

where $\sim N(x, y)$ refers to a normal distribution with a mean x and a variance y. From this model, the average true long-term change is equal to the mean of $\beta = b$; the short-term variability or 'noise' is equal to the variance of the $\varepsilon = \sigma^2{}_w$, whereas the long-term variability is equal to the variance of the $\beta s = \sigma^2{}_\beta$. The two main assumptions in this model are the linearity of the association and the normality in the distributions of α, β and ε. To satisfy these assumptions, the raw IOP data were transformed, using a Box–Cox or power transformation to normalise the complete set of IOP measurements, at each time point and over all the assessment points combined.

Descriptive statistics were used to identify both the degree of normality of the IOP data as well as the presence of possible outliers. Shapiro–Wilk normality tests were carried out to further test the assumption of normality at each assessment point. Allowing for multiple tests adjustment, a 10% significance level was chosen, below which normality was rejected.

Variability of estimates to baseline intraocular pressure and other covariates

Bland–Altman plots were used to investigate the relationship between variability and mean IOP values. First, the mean versus SD was plotted for each subject to check for differences in variance based on an overall mean (as the measures for IOP seemed to be very stable over time for the majority). Second, the mean of the baseline and first visit measurements was plotted against the difference between baseline and first visit to check for differences in variance based on initial IOP measure.

Estimates of the variability components were obtained after stratification by IOP baseline level (threshold at 26 mmHg) to explore potential differences. Relevant covariates were included as part of the linear random-effects model to determine potential association with IOP.

Imputation assumptions

Missing data make the accurate estimation of a change in a measurement variable and the quantification of the variability problematic. Additionally, the number of visits per participant in both trial data sets was variable, with right-sided censoring at high values of IOP. The high IOP values are the most informative with regards to monitoring IOP. Imputation methods are therefore needed to incorporate these right-censored values into the analysis.

The analyses were conducted under different imputation assumptions for the missing times between visits, and for the missing IOPs. For the missing times between visits, imputed values were obtained for the independent variable, time, using a regression imputation method. In this method, each imputed value is predicted from the regression of the observed values at the time the value is missing on all observed data at each previous time point.

For missing IOP values, three approaches were assessed: non-imputation (missing at random), median imputation and last value carried forward (LVCF). Under the first approach, the analyses

were carried out without any imputation of missing data. For the second approach, the median value was imputed when IOP values were missing. Briefly, if IOP_{ij} is the missing IOP value for patient *i* at visit *j*, then the difference between the averages of all observed values for patient *i* and all observed values at visit *j* is added to each observed value at visit *j*. This generates a set of values whose median will be the imputed IOP_{ij}. The median imputation method has the characteristic of returning values close to the individual's mean. This was not the obvious pattern for individuals with a tendency to reach very high IOP values and thus leaving the placebo group. To eliminate this effect, the imputed values for individuals with last observed values higher than their mean were excluded from the analysis. Fourteen per cent of the placebo group were in this situation. The linear random-effects model was fitted to the remaining data.

The third imputation approach applied to missing IOP data was the LVCF. In this approach the last observed value is used to replace all subsequent missing data, that is, visits for which no IOP measurement was available. If there was an observed measurement available before and after a missing measurement, the average of these IOP values (before and after) was imputed for the missing value.

Based on the different imputation assumptions, three different versions of the data sets were created for the transformed IOP data: (1) non-imputed IOP data, (2) median imputation data with points regressing to the mean removed and (3) LVCF imputed data.

Model validation

Estimates obtained using the above methods allow modelling of the number of patients whose 'true' IOP level exceeds predefined threshold(s). The average change in IOP per year and the estimated within-person variability at different times (e.g. 1, 2 and 5 years) allow us to estimate the proportion whose true value would, and would not, have drifted beyond a specific level. The signal-to-noise ratio can also be computed at different time points to identify the time point at which the signal of the measurement is greater than the noise or measurement error. These models were used to simulate the proportions of individuals with a given predicted change at different time points, which were compared against the observed change from baseline IOP for both data sets (Moorfields and Rotterdam). Graphical methods were used to explore divergence from normality on the slopes and noise estimates as was the Shapiro–Wilk normality test.

External validation

Having estimated the parameters using the Moorfields data set, the Rotterdam data set was used for external validation of these models. For the Rotterdam data, the eyes were selected for inclusion randomly. For values of IOP entered as a range, the mid-point in the range was used. After running descriptive statistics for IOP, the extreme values found were replaced with the maximum IOP observation for the particular patient, excluding the outlier. Up to 16 visits within the trial period were considered for this analysis. The same imputation methods and Box–Cox transformation as for the Moorfields data were used. The validation process consisted of predicting the transformed Rotterdam IOP values using each estimated Moorfields model. The proportion of times that the observed value fell in the estimation interval given by:

$$PTT_{it} \pm 1.96 \times sqrt(Var(PTT_{it})), \text{ where } Var(PTT_{it}) = \sigma_{TT}^2 + \beta^2 t^2 + \sigma_w^2 \qquad \text{[Equation 11]}$$

was calculated, where PTT is the predicted transformed Rotterdam IOP, TT represents the transformed IOP and σ^2_{TT} is equivalent to σ^2_α in Equation 10. High values indicate greater accuracy of the corresponding model.

Results

Description of included data

Moorfields data set

The placebo group in the Moorfields data set includes data on 153 participants (61% male) with a mean (SD) age of 58.8 (10.9) years and an age range of 29–90 years. The majority were of white ethnicity (76%). The mean (SD) baseline IOP was 24.5 (3.5) mmHg with a range of 16–36 mmHg.

A total of 14 approximately 4-monthly visits in the period 1992–8 were included in this analysis. As the visit number (i.e. follow-up) increased, the number of participants with data steadily decreased. This ranged from 4.6% missing values at visit 2 to 63.4% missing values at visit 14. The mean IOP measurement over time for the placebo group is shown in *Figure 48*. A slight reduction can be observed over time, contrary to initial expectations.

Figure 49 shows the distribution of IOP at each of the 14 visits. The normality test results are also shown and normality was rejected at four time points (visits 4, 5, 6 and 9). A single outlier was identified (IOP = 48 mmHg, visit 4). This outlier was not included in further analyses.

Figure 50 shows the IOP measurements after the data were transformed (Box–Cox transformation with $\lambda = 0.26$) and corresponding normality test results. Normality was rejected for only one visit (visit 2, $p = 0.002$) suggesting that the *transformation resulted in data that were more suitable for methods (such as the linear random-effects model) that assumed normality*.

Handling of missing data
Intraocular pressure

The change in IOP from baseline at each visit (visits 2–14) was calculated for the raw and imputed data sets. Graphs of the mean and variance of the change in IOP from baseline as a function of time are shown in *Appendix 8*. Bland–Altman plots (*Figure 51*) show that the variability decreases with high mean IOP. This could be owing to censoring of those with higher IOP.

A total of 50 out of the 153 patients in the data set had crossed over from placebo to treatment, representing 33% of this group. Once a patient crossed over to treatment, his or her measured IOP was not included in the analysis. In these patients, the data points after crossover were replaced using LVCF. A single case presented several crossovers during the trial period, starting as a placebo case. This patient was included in the analysis and a sensitivity analysis run later to investigate the effect of the presence of this patient on the estimated parameters, with their inclusion having no significant impact. Stratification of the placebo group by baseline IOP resulted in 59 (39%) patients with IOP higher than the risk value (26 mmHg), of whom 30 (51%) were included among the crossover group mentioned above.
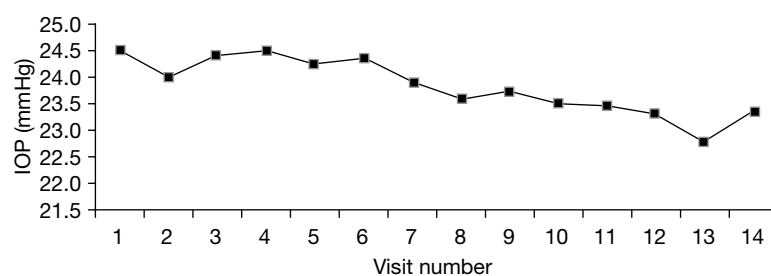

FIGURE 48 Mean IOP values over follow-up period (Moorfields data – placebo group only).
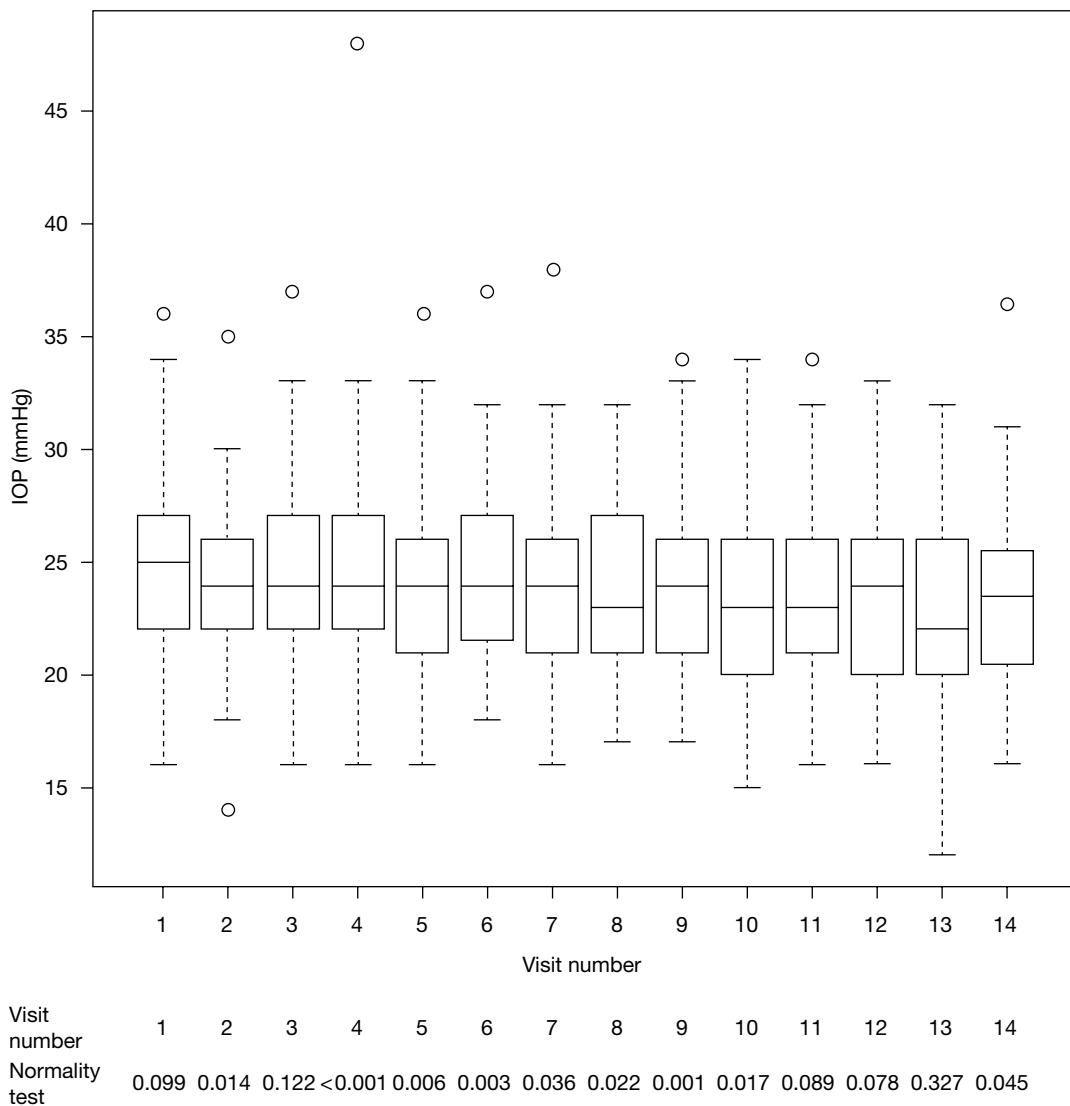
Visit number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Normality test	0.099	0.014	0.122	<0.001	0.006	0.003	0.036	0.022	0.001	0.017	0.089	0.078	0.327	0.045

FIGURE 49 Intraocular pressure measurements over time (Moorfields placebo group – original IOP data).

Visual fields

For visual fields, the proportion of missing MD measurements ranged from 10% at baseline to 60% at visit 14, whereas PSD was not recorded for every patient at all time points. The PSD variable was measured at baseline in 75 (49%) patients. The mean (SD) MD at baseline was −0.39 (1.52). The mean and variance of MD change scores as functions of time show that up to visit 11 the mean and variance of change had an increasing trend (figure not shown). As for PSD, this was a highly skewed variable at baseline. Its mean (SD) is given by 2.20 (1.05 dB), with a median (range) of 1.98 (0.08–7.82 dB) and a skewness value of 2.85.

Rotterdam data set

The Rotterdam placebo data set includes data on 132 participants (51% male). The age range was 24–81 years, with a mean (SD) of 56.1 (11.0) years. Only those of white ethnicity were eligible to participate in the RCT. The mean (SD) baseline IOP was 25.7 (2.5) mmHg with a range of 18–35 mmHg.

A total of 23 visits were available; however, the second visit was taken as the baseline visit to account for regression to the mean, leaving only 21 visits for the initial analysis. The percentage
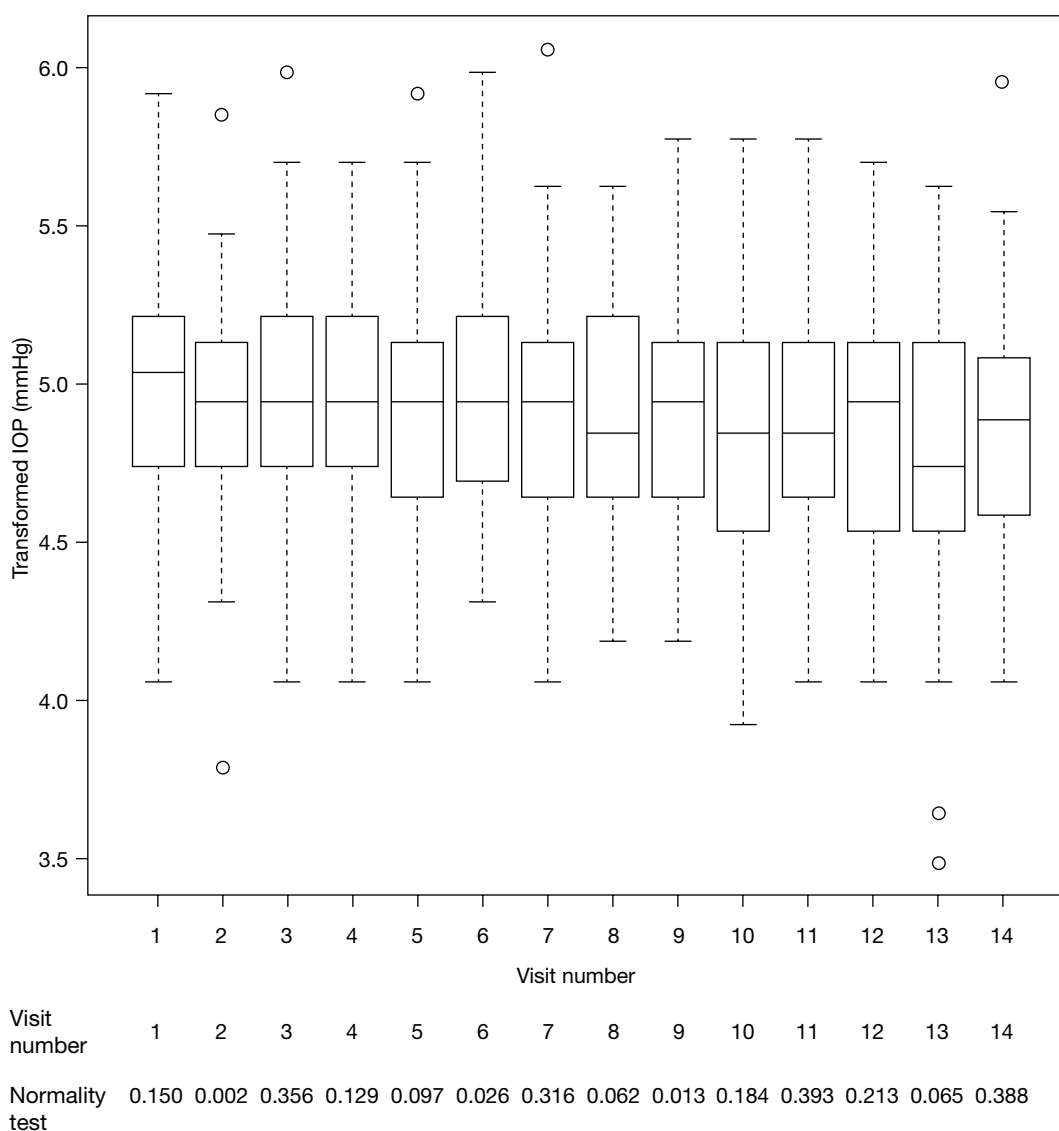
FIGURE 50 Intraocular pressure measurements over time (Moorfields placebo group – transformed incomplete IOP data).

of missing data at each visit, illustrated in *Appendix 9*, increased over time, as expected, with a peak at visit 10 of 38%, and then a slow increase from 25% to 36% between visits 10 and 15, to be followed by a sharp increase to a final 85%. We used data up to visit 15.

The IOP values in the placebo group from baseline (equivalent to visit 2 in the original data set) show a steady increase up to visit 7 and then a decline in mean IOP (*Figure 52*). Bland–Altman plots showed no association between variability and mean IOP level (*Figure 53*).

Short-term and long-term intraocular pressure variability

For both estimation methods (direct method and linear random-effects model), three analyses making different assumptions about the missing data were carried out. All six analyses used the transformed IOP as the dependent variable and time as the independent variable. The chosen Box–Cox transformation implied that the true baseline IOP at time i was given by:

$$IOP_i = (0.26 \times TT_i + 1)^{3.84} \qquad \text{[Equation 12]}$$
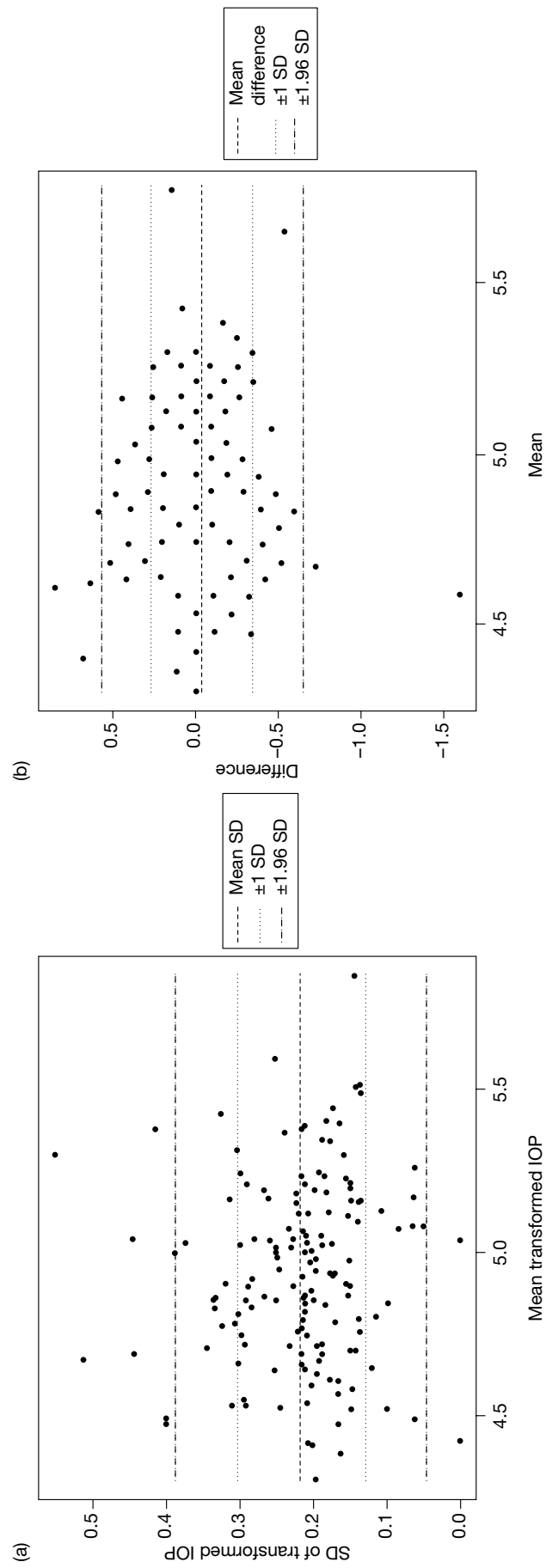
FIGURE 51 Bland–Altman plots using transformed Moorfields IOP data: (a) mean vs SD of transformed IOP per subject; (b) mean vs difference of transformed IOP at baseline and visit 1.
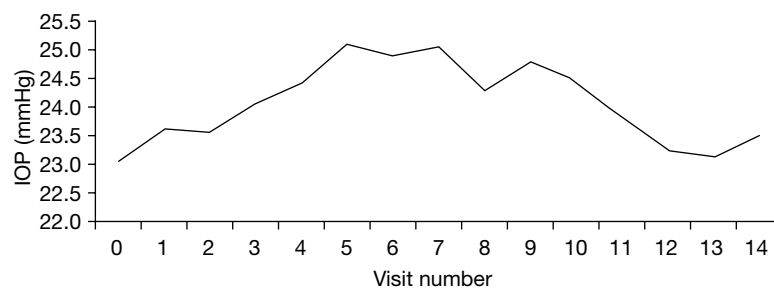
FIGURE 52 Mean IOP values over follow-up period (Rotterdam data – placebo group only).

where TT_i is equivalent to the transformed IOP at time point *i* (or α in our notation in *Modelling variability*).

From an exploratory analysis of the data it became apparent that estimates from the direct method would be less stable than those obtained using the linear random-effects model. Such estimates would also have been difficult to incorporate into an economic model, and therefore the following results focus on the linear random-effects model. For completeness, the estimates obtained using the direct method are reported in *Appendix 9. Table 11* reports the estimates for the linear random-effects model under three missing data assumptions. Under these assumptions, missing data were imputed using either LVCF or the median imputation method or there was no imputation (i.e. missing at random assumption). The last scenario is included for completeness of presentation as it allows us to observe the effect of the sample size increment in the variances compared with the previous two models.

There is a difference in the estimates of the noise and the variance of the random slope (signal) in these three models. The LVCF model was chosen out of the three for further validation and reporting as it has the smallest estimate of the noise (0.036) with the highest estimate of the variance for the random slope (3.6×10^{-8}). High variability in the random slope means that individuals have a greater chance of showing a divergence from the overall population whereas a small noise would also mean that any change has a greater chance to be real (as opposed to random error). Therefore, this model should give the most conservative estimates for monitoring purposes (monitor more often). All three models estimate the average change over time to be close to 0 (-7.0×10^{-6} for model 1). This is equivalent to a decrease in IOP of approximately 0.08 mmHg in a 3-year interval (baseline of 25 mmHg to 24.92 mmHg at 3 years).

Signal-to-noise ratio

Table 12 presents the signal-to-noise ratio ($\sigma_\beta^2 t^2 / \sigma_w^2$) at 3 and 6 months, and 1, 2 and 3 years obtained under the models for IOP in the placebo group. According to these ratios, the LVCF linear random-effects model estimates a greater signal than noise after 3 years. To explore the impact of having multiple measurements of IOP at a single visit and using the average as an estimate for the true IOP, we generated the equivalent signal-to-noise ratios based on the average IOP of two and three observations (this assumes conditional independence between the IOP measures of an individual in one visit). These values are presented in *Table 12* and show that, for the average of two measures, the signal-to-noise ratio would reach a value of 1 within 2 years, while for the average of three measures the signal-to-noise ratio reaches a value of 1 between 1 and 2 years from baseline.

True- and false-positives

We explored the impact of the noise in the observed IOP measurement given a defined threshold of IOP (26 mmHg) for different baseline levels (21, 25 and 28 mmHg). The calculation of true and
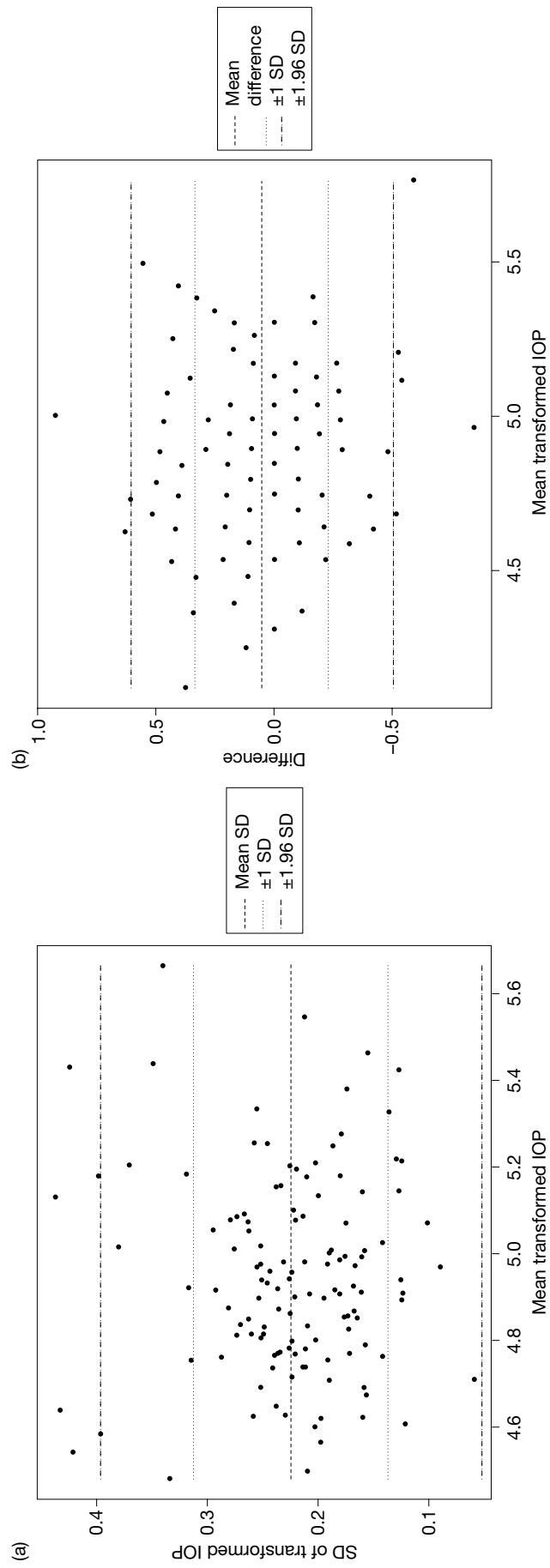
FIGURE 53 Bland–Altman plots using transformed Rotterdam IOP data: (a) mean vs SD of transformed IOP per subject; (b) mean vs difference of transformed IOP at baseline and visit 1.

TABLE 11 Linear random-effects model outputs for transformed IOP (Moorfields placebo group)

Baseline distribution of TT_i	Distribution of the random slope β	Correlation of TT_i and β	Noise estimate σ_w^2
Assumption 1: using LVCF to impute missing data			
$N(4.96, 0.104)$	$N(-7.0 \times 10^{-6}, 3.6 \times 10^{-8})$	0.07	0.036
Assumption 2: using the incomplete median imputed data			
$N(4.94, 0.107)$	$N(-1.6 \times 10^{-5}, 1.0 \times 10^{-8})$	0.21	0.039
Assumption 3: using incomplete (non-imputed) data			
$N(4.97, 0.106)$	$N(-3.6 \times 10^{-5}, 2.3 \times 10^{-8})$	0.15	0.048

TT_i is the transformed IOP following Box–Cox transformation at time *i*.
$N(x,y)$ represents a normal distribution with mean of *x* and variance of *y*.

TABLE 12 Signal-to-noise ratio estimates at several time points – based on the linear random-effects model estimation

Model	3 months	6 months	1 year	2 years	3 years
Placebo group					
LVCF	0.0	0.0	0.1	0.5	1.2
(Incomplete) median imputation	0.0	0.0	0.0	0.1	0.3
Transformed not imputed IOP	0.0	0.0	0.1	0.3	0.6
Several measures at a single time point[a]					
LVCF – average of two observations	0.0	0.1	0.3	1.1	2.4
LVCF – average of three observations	0.0	0.1	0.4	1.6	3.6

a Equivalent to multiplying the signal-to-noise ratio by the number of measures.

false-positives indicates that measuring IOP in shorter than 1-year periods results in identifying only noise. However, for the purposes of clinical management it is more relevant to explore the potential changes as defined by the model, that is, the expected change in IOP (both true and observed) for an average individual in the population and for one having an extreme change in IOP. *Table 13* presents these IOP values for different baseline levels.

Based on the results obtained from the LVCF linear random-effects model (see *Table 11*), we can estimate the number of observed IOP measures that differ from their true IOP by ≥ 2 mmHg and explore how different measurement regimes, using the mean of two, three, four or five measures (assuming independence), improve on these percentages. These estimates are presented in *Table 14*.

For comparison, linear random-effects models were fitted over the treatment group using the transformed not imputed data. Model 1T in *Table 15* was obtained by regression of the transformed IOP on time. In this case, the distribution of the random slope was similar to the one obtained with model 1 in the placebo group. The mean of the baseline transformed IOP was slightly lower but its variance was much larger than the one estimated by any of the three models in the placebo group. The same was true for the noise estimate.

TABLE 13 Examples of median and extreme monitoring IOP scenarios given baseline IOP of 21, 25 and 28 mmHg using the LVCF model

| | IOP (mmHg) | | | | | |
	Baseline	4 months	1 year	2 years	3 years	4 years
Median β (-1.1×10^{-5})						
True	21.0	21.0	21.0	20.9	20.9	20.9
Observed	22.4	19.3	21.7	19.5	20.0	24.7
True	25.0	25.0	25.0	24.9	24.9	24.8
Observed	25.2	27.6	23.9	26.3	25.3	25.7
True	28.0	28.0	28.0	27.9	27.9	27.8
Observed	27.7	27.4	31.3	27.4	28.2	27.8
Extreme β (6.1×10^4)						
True	21.0	21.7	23.2	25.5	28.0	30.6
Observed	22.9	21.1	22.2	25.2	25.6	31.3
True	25.0	25.8	27.5	30.1	32.9	35.9
Observed	30.2	25.8	30.4	28.9	28.8	34.9
True	28.0	28.9	30.7	33.5	36.6	39.8
Observed	25.0[a]	27.3	29.5	37.1	36.3	37.7

a   Low value randomly obtained showing that error in measurement could be either above or below the true IOP level.

TABLE 14 Percentage of patients with differences between real and observed IOP ≥ 2 mmHg

| Initial IOP (mmHg) | Missing data assumption | Number of measures[a] | | | | |
		1	2	3	4	5
21.0	LVCF	28.3	11.4	6.3	3.2	1.7
	Pseudo	28.9	14.1	6.4	4.2	2.4
	No imputation	31.9	18.1	10.6	4.6	2.5
25.0	LVCF	34.7	16.8	9.7	5.7	2.9
	Pseudo	35.3	19	9.8	6.2	4.4
	No imputation	37.2	25	14.8	7.4	4.3
28.0	LVCF	38.4	20.1	11.8	8.7	4.5
	Pseudo	39.1	22	12.7	8.5	6.8
	No imputation	41.7	29.7	17.7	10.7	6.9

a   For two, three, four and five measures the comparison is between the mean of all observed measures and the 'real' IOP.

TABLE 15 Linear random-effects model outputs for transformed IOP (Moorfields – treatment group)

Baseline distribution of TT_i	Distribution of the random slope β	Correlation of TT_i and β	Noise estimate σ_w^2
Model 1T: using incomplete (non-imputed) data and leaving out covariates			
$N(4.71, 0.160)$	$N(-7.3 \times 10^{-5}, 3.2 \times 10^{-8})$	-0.46	0.058

TT_i is the transformed IOP following Box–Cox transformation at time *i*.
$N(x,y)$ represents a normal distribution with mean of *x* and variance of *y*.

Validation of intraocular pressure models

Internal validation – Moorfields data

The purpose of these models is to predict the optimal monitoring frequency, particularly for inclusion into the economic model. For this reason it is necessary to explore the accuracy of these models on the prediction of change at a population level. To do this, we simulated the percentage of patients with a given predicted change from baseline at different time points using data from the LVCF model and compared these proportions against the observed change from baseline IOP for the Moorfields data set (placebo group only). These are presented in *Tables 16* and *17* and show that the model underestimates large changes in IOP ($\geq 5\,\text{mmHg}$) for all time points.

The histograms in *Figure 54* present the empirical (histogram) and estimated (dotted line) distributions of the residuals and slopes obtained from the (LVCF) transformed IOP Moorfields data. They show that the normality assumption is adequate for the residuals, but the corresponding slopes-estimated model underestimates the proportion of extreme values observed. *Appendix 9* gives a table comparing the cumulative distributions from the model (normal distribution) and the empirical distribution for the slopes.

Converting these estimates to the original IOP scale (*Figure 55a*) shows that the noise term for a single measure is likely to be in the region of 0.6–1.4 mmHg (normal range). Therefore, assuming two measurements (e.g. baseline and follow-up) with additive errors and taking the upper level of the normal range for the error (1.4 mmHg), a difference of 3 mmHg between two readings can be due only to noise (no real change). *Figure 55b* presents a histogram of a series of simulated differences between two values with no real change (difference owing to noise only). The probability of having a difference of $\leq 3\,\text{mmHg}$ is 73% (dotted lines in figure).

External validation – Rotterdam data

The distribution of the placebo IOP data from the Rotterdam data set after applying the same Box–Cox transformation as for the Moorfields data set ($\lambda = 0.26$ for all visits) resulted in a normal distribution in 12 out of the 16 visits. Normality was rejected at visits 1 ($p = 0.002$), 2 ($p = 0.003$), 4 ($p = 0.004$) and 13 ($p = 0.004$).

TABLE 16 Moorfields observed data – percentage of patients with a given change from baseline ($IOP_t - IOP_0$)

Time from baseline	0–1 mmHg	1–2 mmHg	2–3 mmHg	3–4 mmHg	4–5 mmHg	≥ 5 mmHg
4 months	13.7	26	18.5	18.5	11.0	12.3
1 year	10.3	20.6	20.6	16.9	15.4	16.2
2 years	8.9	26.8	17.9	15.2	13.4	17.9
3 years	6.1	20.4	21.4	16.3	11.2	24.5

TABLE 17 Simulated data – percentage of patients with a given change from baseline ($IOP_t - IOP_0$)

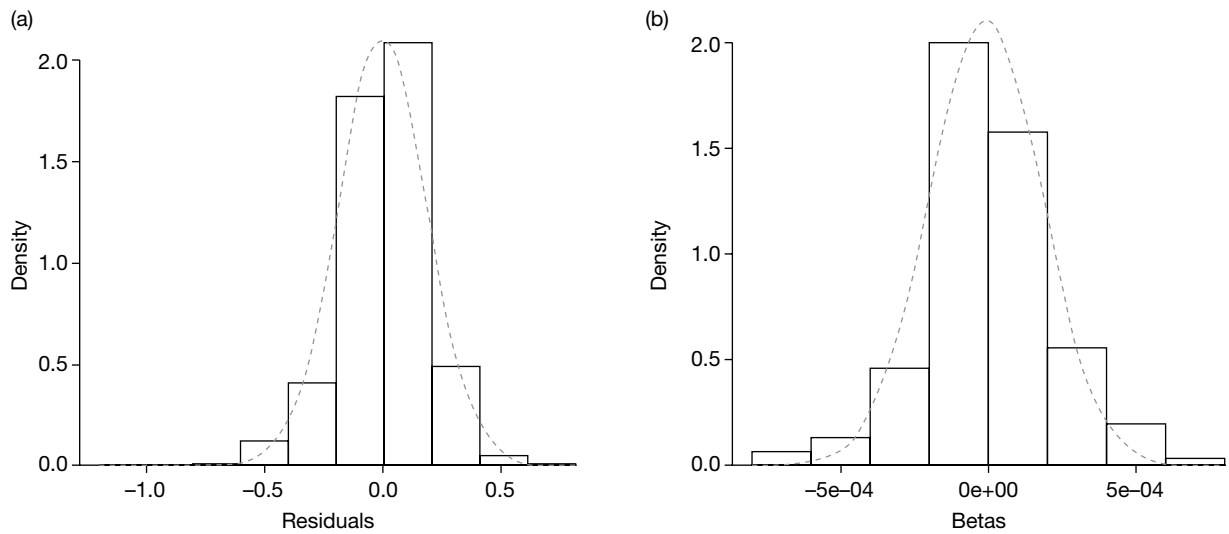| Time from baseline	0–1 mmHg	1–2 mmHg	2–3 mmHg	3–4 mmHg	4–5 mmHg	≥ 5 mmHg
4 months	28.1	24.6	18.6	13.9	7.1	7.7
1 year	26.3	25.4	19.3	12.5	7.6	8.9
2 years	26.2	23.7	17.5	11.0	10.2	11.4
3 years	21.8	21.7	17.3	15.5	8.9	15.1

(a)

(b)

FIGURE 54 Histograms presenting the distributions of residuals and slopes from the raw data and simulated data (Moorfields data): (a) histogram of the mixed-model residuals; (b) histogram of slopes from LVCF (transformed) IOP data.
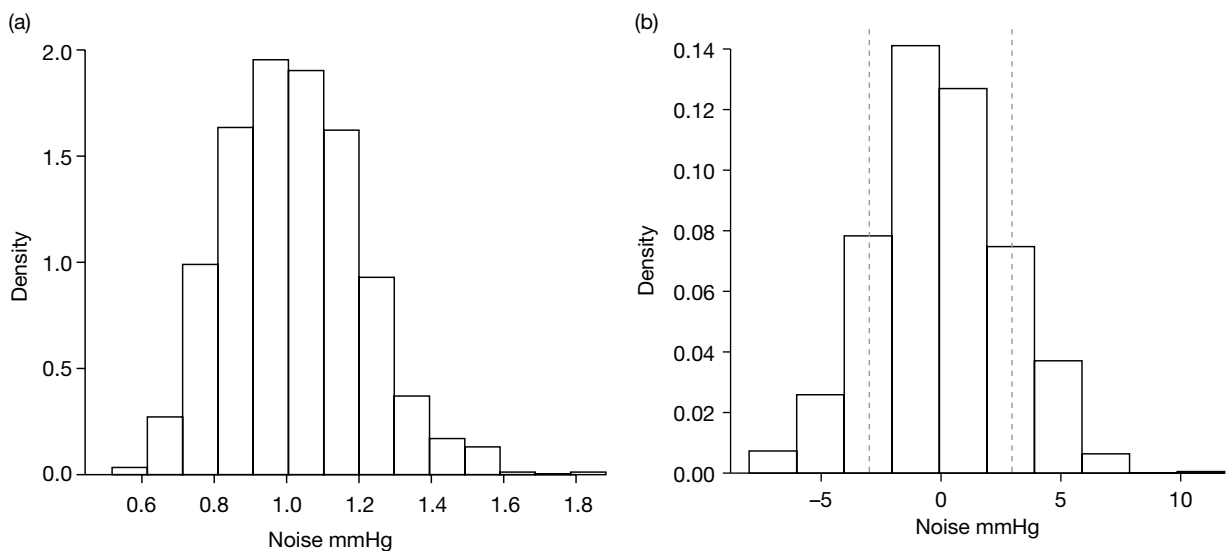
(a)

(b)

FIGURE 55 Histograms presenting (a) the distribution of the noise in the IOP scale and (b) the difference between two measurements attributed to noise only (no real change).

For each model (i.e. LVCF, reduced median imputation, no imputed data), 98% of the time the observed TT_{it} in the Rotterdam file fell within the 95% CI defined by *Equation 11*. This would suggest that, in terms of predictive ability, the model would accurately predict the observed variability in the Rotterdam data.

As with the internal validation with the Moorfields data set, we calculated the proportions of observed change from baseline IOP for selected differences. These are presented in *Table 18*. Reassuringly, these percentages appear to be similar to those obtained in the observed Moorfields data set and suggest that the variance is similar across populations. As for the Moorfields data, the model consistently underestimates the proportion of individuals with large observed changes.

Impact of baseline intraocular pressure and covariates

Impact of baseline intraocular pressure

We explored the variability of estimates according to different IOP baseline levels. Two separate models were obtained based on IOP levels at baseline. A low IOP level group was defined as those individuals with observed baseline IOP < 26 mmHg whereas a high IOP level was defined as ≥ 26 mmHg. All estimates were obtained using a linear random-effects model and LVCF as the imputation method. *Table 19* shows that the estimates are very similar for both groups with a smaller noise estimate in the high IOP level. The impact of this is that the signal-to-noise ratio reaches a value of 1 earlier in the high IOP level group (*Table 20*).

As with the model for the full data, we explored whether the assumption of normality on the errors and the slopes was appropriate. The histograms in *Figure 56* present the empirical and estimated distributions of the slopes for the two strata (IOP < 26 mmHg and IOP ≥ 26 mmHg) obtained from the (LVCF) transformed IOP Moorfields data. They show that the corresponding slopes estimated by the model underestimate the proportion of extreme values observed (heavier tails), as was observed in the full model.

TABLE 18 Rotterdam observed data – percentage of patients with a given change from baseline ($IOP_t - IOP_0$)

	0–1 mmHg	1–2 mmHg	2–3 mmHg	3–4 mmHg	4–5 mmHg	≥ 5 mmHg
6 months	14.0	27.1	26.4	10.1	7.0	15.5
1 year	8.5	27.4	19.7	20.5	13.7	10.3
2 years	9.7	31.0	15.9	14.2	10.6	8.0
3 years	8.7	17.5	23.3	10.7	14.6	18.4

TABLE 19 Linear random-effects model outputs for transformed IOP, stratified by IOP at baseline (Rotterdam placebo group)

Baseline distribution of TT_i	Distribution of the random slope β	Correlation of TT_i and β	Noise estimate σ_w^2
Model 1a: low IOP (baseline IOP < 26 mmHg, LVCF imputation)			
$N(4.83, 0.076)$	$N(-1.0 \times 10^{-6}, 2.34 \times 10^{-8})$	0.07	0.037
Model 1b: high IOP (baseline IOP ≥ 26 mmHg, LVCF imputation)			
$N(5.28, 0.075)$	$N(-1.9 \times 10^{-5}, 5.29 \times 10^{-8})$	0.07	0.032
Model 1: using LVCF full sample			
$N(4.96, 0.104)$	$N(-7.0 \times 10^{-6}, 3.6 \times 10^{-8})$	0.07	0.036

TT_i is the transformed IOP following Box–Cox transformation at time *i*.
$N(x,y)$ represents a normal distribution with mean of *x* and variance of *y*.

TABLE 20 Signal-to-noise ratio estimates, stratified by IOP at baseline (Rotterdam placebo group)

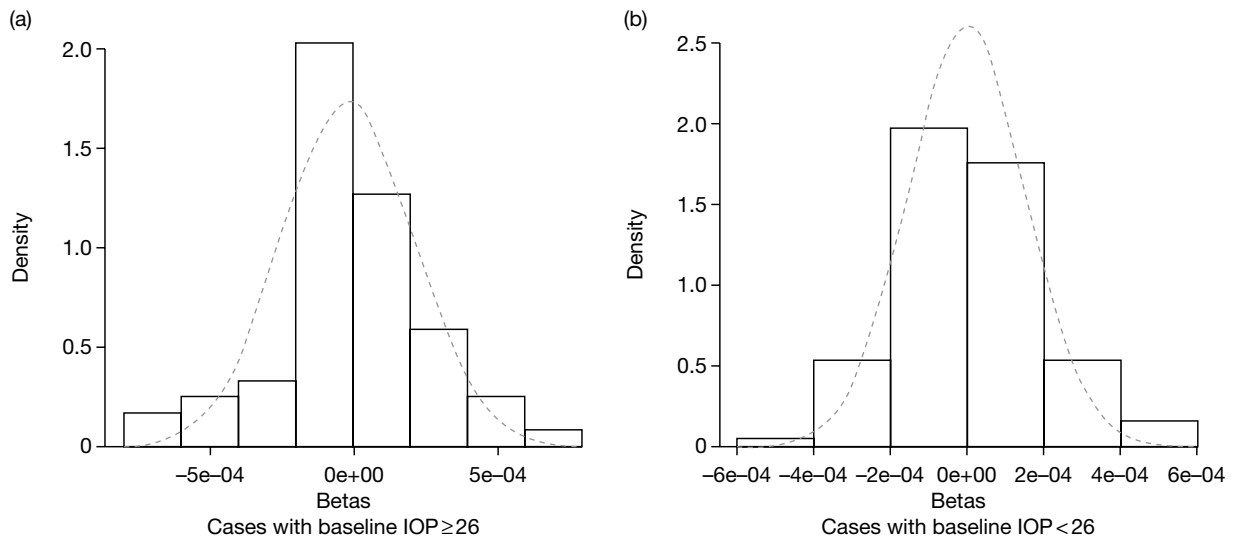| Model	3 months	6 months	1 year	2 years	3 years
$IOP_{baseline} < 26$ mmHg	0.0	0.0	0.1	0.3	0.8
$IOP_{baseline} \geq 26$ mmHg	0.0	0.1	0.2	0.9	2.0
Full sample	0.0	0.0	0.1	0.5	1.2

FIGURE 56 Histograms for the empirical and estimated distributions of slopes based on the Moorfields' (LVCF) transformed IOP data for two strata: (a) $IOP_{baseline} \geq 26$ mmHg; (b) $IOP_{baseline} < 26$ mmHg.

The histograms in *Figure 57* present the empirical and estimated distributions of the residuals for the two strata (IOP < 26 mmHg and IOP ≥ 26 mmHg) obtained from the (LVCF) transformed IOP Moorfields data. They show that the assumption of normality appears adequate.

Impact of important covariates

The covariates age and observer are regarded clinically as potential modifiers of IOP measurement. The effect of these clinically significant covariates was analysed by fitting a separate model for each covariate. None of these covariates was statistically significant for the models fitted over the placebo group. In the treatment group, the coefficient for age was not significantly different from 0, whereas investigator had a significant effect (*Table 21*). Also, this model returned a smaller estimated variability [i.e. $Var(TT_i)$, $Var(\beta)$ and σ_w^2 decreased].

Models for the visual field parameters: mean deviation and pattern standard deviation

Because of limitations in the available data, the variability of MD was explored only by fitting a model using the direct method. The variability of PSD could not be explored as only baseline data were available. The direct method returned an estimated mean (SD) for MD at baseline of –0.39 (1.52) with a noise estimate of $\sigma_w^2 = 1.15$. This method gives an estimate of the mean and variance of change at every visit. On three occasions the estimate of this variance was negative (*Table 22*). This is owing to the additive formula for estimating such variance and the lack of linearity of this variance over time (R^2 of the regression of variance of change on time = 0.24).

Discussion

Recent research has highlighted how the variability of the measures used for monitoring chronic conditions impacts on the usefulness of these measures.[230–232] This research suggests that, in most cases, measurements are carried out too frequently and that most of the change identified is owing to biological (day-to-day) variability and inherent measurement error and therefore not a real change in the health status of the individual. Clinical decisions such as changes to medication titration based on such measurements are potentially detrimental to a patient's health. Quantifying this variability is therefore a necessary step for optimal monitoring.[232]
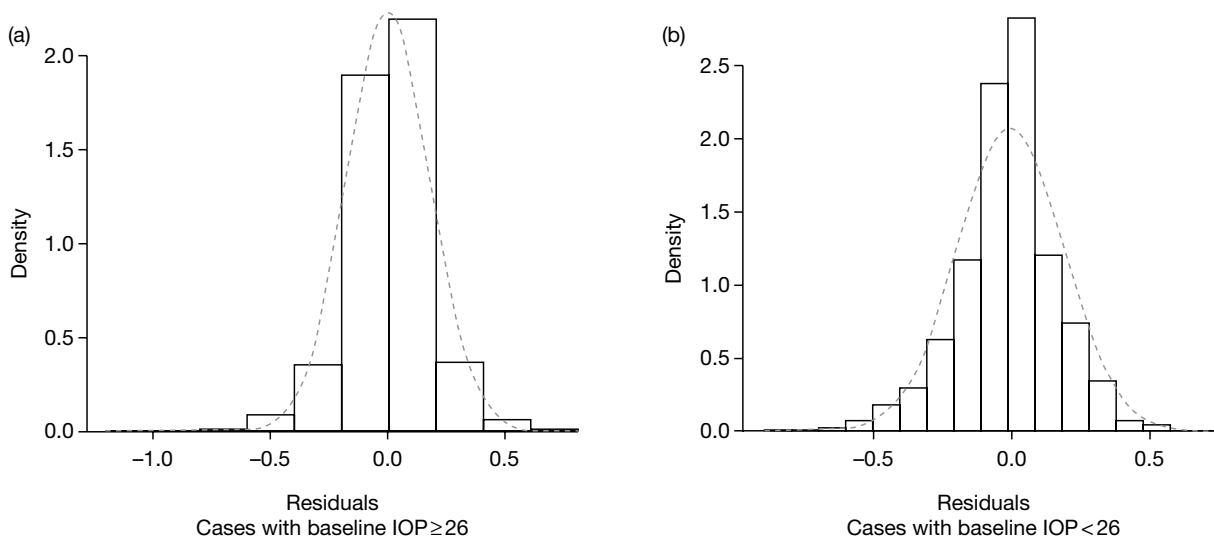
FIGURE 57 Histograms for the empirical and estimated distributions of residuals based on the Moorfields' (LVCF) transformed IOP data for two strata: (a) $IOP_{baseline} \geq 26$ mmHg; (b) $IOP_{baseline} < 26$ mmHg.

TABLE 21 Estimates of effect of covariates (treatment group)

Baseline distribution of TT_i	Distribution of the random slope β	Correlation of TT_i and β	Noise estimate σ_w^2
Model 2T: using incomplete (non-imputed) data and including investigator as a covariate			
$N(4.77, 0.149)$	$N(-8.9 \times 10^{-5}, 2.6 \times 10^{-8})$	−0.30	0.053

For investigator: only one statistically significant comparison was observed: investigator Garway-Heath (reference category) vs investigator Tan (coefficient = 0.15, standard error = 0.05, p = 0.004).
TT_i is the transformed IOP following Box–Cox transformation at time *i*.
$N(x,y)$ represents a normal distribution with mean of *x* and variance of *y*.

TABLE 22 Direct method outputs for MD

Characteristic	Visit number, t[a]												
	2	3	4	5	6	7	8	9	10	11	12	13	14
Mean change	0.01	0.03	0.10	0.07	0.11	−0.01	0.01	0.10	0.10	0.09	0.12	−0.19	−0.11
Variance of change	−0.11	−0.13	−0.24	1.71	0.30	0.40	0.57	1.41	0.49	0.78	0.67	1.70	0.77
Months	4	8	12	16	20	24	28	32	36	40	44	48	52
Signal-to-noise ratio	−0.1	−0.1	−0.2	1.5	0.3	0.4	0.5	1.2	0.4	0.7	0.6	1.5	0.7

Baseline: mean = −0.39, variance = 1.52, noise = 1.2.
a   Visit frequency every 4 months.

For individuals with high risk of developing glaucoma, there are several factors that are monitored with regularity, including IOP, visual function measured by visual fields and structural changes in the RNFL and the optic nerve. The monitoring frequency should be determined by the outcome that has greatest bearing on treatment decisions, which in this context is monitoring the visual fields and the optic nerve and RNFL to detect early progression to glaucoma. However, quantifying changes in visual fields and the RNFL/optic nerve is characteristically difficult, with no unified single measure for either. In the present study it was not possible to make conclusions about the long-term variability for the visual field variables PSD and MD. For PSD it would be necessary to identify the subjects with this measurement at follow-up; these data were not available in the data sets used. Neither data set had repeated measures of the RNFL or optic nerve head that were available for this analysis.

Thus, the focus of this chapter has been almost entirely on IOP and on estimating the 'noise' (biological variability and measurement error) and the signal (variability and change beyond the noise) in IOP measurement. This chapter presents estimates for the signal and noise of IOP measurement based on data from the placebo arm of the Moorfields RCT[32] evaluating the effectiveness of a treatment for OHT. These data have the advantage that individuals were followed over a long period with no treatment and therefore IOP variability is not affected by medication changes. The main assumption made is that the repeated IOP measures in an individual allow us to determine his/her real IOP. In simplistic terms, the true IOP can be determined from the average of repeated IOP measures over time.

The models used to describe the behaviour of IOP and estimate the relevant parameters follow those previously reported.[233] The use of linear random-effects models allows long-term prediction of IOP; these were used for the economic model. Different strategies for handling missing data can be incorporated. As the completeness of the information decreases with time, the choice of strategy for handling missing data impacts on the estimates of the signal and the noise. The most appropriate models for the transformed IOP were those based on the imputation methods LVCF and (incomplete) median imputation, in that order. This was because these methods are conservative in potentially underestimating the noise and hence leading to recommendations of more frequent monitoring.[233]

Summary of main findings

The main findings are summarised below:

- Measured IOP could be incorrect (different from true IOP) by ≥ 2 mmHg at baseline about 30% of the time. This is consistent with repeatability coefficients of around 2 mmHg observed for two measurements of IOP on the same occasion using GAT as reported in *Chapter 5*.[131,154,157,210]
- The level of IOP is relatively stable but a change ≥ 5 mmHg can be observed at 3 years in about 25% of participants.
- The model estimates that point estimates of IOP changes may be due to noise (measurement error and biological variation). Based on the normal range of the noise (0.6–1.4 mmHg for a single measure), with two measurements (baseline and follow-up) with additive errors, and taking the upper level of the normal range for the error (1.4 mmHg), observed changes of the order of 3 mmHg could be explained as noise.
- For the majority of individuals the true change in IOP is likely to be smaller than the measurement error plus biological variability (noise).
- Assuming independence in the measurements, using the average of two measures at baseline increases the signal-to-noise ratio such that one would expect to be able to detect true change using the average of two measures at 2 years. If the average of three baseline measures of IOP were taken, true change might be detectable with repeat testing between 1 and 2 years. This

approach of reducing noise is based on the assumption that the measurement errors across the repeated IOPs are not associated (e.g. all errors going in the same direction). There were insufficient data to inform whether or not this assumption holds for repeat measures at the same session.

■ Individuals with higher baseline IOP (≥ 26 mmHg) are more likely to have a true change in IOP at around 2 years, whereas for individuals with lower IOP the model suggests that a true change in IOP would be unlikely within 3 years.

■ Participant age or observer undertaking the IOP measurement did not affect the findings for the main analysis. Observer was statistically associated with the variability in the treatment group, providing some evidence on variability associated with the person undertaking the tonometry.

■ As IOP increased fluctuations in IOP also increased. These fluctuations are likely to be physiological.[96]

■ As part of the internal validation of the model, the noise error term in the intervention group was found to be larger than that for the placebo group. This may reflect the variability of the treatment effect or variable participant adherence with treatment.

■ As part of the external validation, the model predicted the observed variability in the Rotterdam data.

Strengths and limitations

Our analysis focused on a single risk factor, IOP, and used statistical methods to simultaneously model the between- and within-subject variability over time. However, the following limitations are noted. First, the model estimates obtained were based on data from the placebo group of a single study. However, we used data from the intervention arm to provide internal validation as well as data from the second RCT (Rotterdam) as part of an external validation of the models. In both cases the behaviour of the model and the prediction of future IOP values suggest that the estimates were adequate.

A second limitation is that none of the models chosen appeared to capture those participants with a rapid clinically significant change. The ideal data set would include individuals with no treatment (or stable treatment) over a long follow-up period during which the IOP was measured regularly. However, as a high IOP triggers the clinical decision of initiating or modifying treatment such data are not available. Our choice of model, the LVCF model, maintaining higher values for those censored, minimised this problem. However, the model is likely to have underestimated the small proportion of individuals with a large change in IOP. To adjust for this we used an empirical distribution that inflates the probability of a large change (which is used as part of the economic analysis reported in *Chapter 8*).

A third limitation is the use of data from a RCT, which means that these data are likely to be from a well-motivated and controlled group of participants, which could explain why the long-term IOP appears to be stable. Anecdotal evidence suggests that even within the trial there was operator variability, with some attenuation of IOP variation because of observers not using the first IOP observed but 'calibrating' their measurement based on previous IOP readings. This would mean that an extreme IOP at a particular visit might not have been taken at face value and multiple readings made. This would lead to an underestimation of the noise, in which case the signal-to-noise ratio would have been overestimated.

There are also several areas of further uncertainty. There are no clinical thresholds for IOP, and hence it is difficult to model the impact that noise has on clinical decisions. We have chosen a 2 mmHg difference as a clinically significant value based on the group collective experience. Modifying this threshold can again have implications depending on the need for more or less accurate estimates of IOP. The choice of tonometer is another uncertainty not addressed in this

chapter as the models were obtained assuming GAT as the reference standard tonometer. We speculate that the use of other tonometers is likely to increase variability (noise). Estimates of these have been presented in *Chapter 5* where the LoA between tonometers were explored. The standard current practice for IOP measurement is unclear as to whether the recording is from single or multiple measures and based on an average or the lowest or highest value recorded. The results and recommendations in this chapter assume that the measures were obtained as a single reading during the session.

Integration within the economic model (see Chapter 8)

The findings of this study informed the surveillance pathways compared within the economic modelling evaluation, namely:

1. Measure IOP more than once at a visit and take the average.
2. Use the signal-to-noise ratio (based on multiple measurements) to guide the frequency of IOP monitoring.
3. Individuals with more extreme changes in IOP were under-represented in the standard modelling; however, purposive oversampling of the extremes of the estimated distribution using the empirical distribution of the slopes (see *Appendix 9*, *Table 69*) partially corrected this.

Conclusions

For the majority of individuals the true change in IOP is likely to be smaller than the measurement error plus biological variability (noise). Observed changes, based on a single measurement, of the order of 3 mmHg can be explained as noise. Reducing the noise level by half, by taking an average of two or three measurements, will allow an estimation of IOP with enough precision to make clinical decisions based on a change in IOP of 2 mmHg. This could be obtained by using the average of repeated measures at a single visit. However, this approach of reducing noise is based on the assumption that the measurement errors across the repeated IOPs are not associated (e.g. all errors going in the same direction). There are insufficient data to inform whether or not this assumption holds for repeat measures at the same session. In addition, same session measurements will not include diurnal variation or short-term biological fluctuations. Diurnal variation can be addressed by standardising IOP measurement time. Short-term biological fluctuations can be addressed by averaging the IOP over repeat visits, but this may not be feasible in current eye care services.

Using the average of two measures of IOP at baseline at a single visit increases the signal-to-noise ratio such that one would expect to be able to detect true change, of the order of 2 mmHg, at 2 years. If the average of three baseline measures were taken, true change might be detectable with repeat testing between 1 and 2 years. These findings, together with the findings of *Chapter 5*, can be used to inform the development of a standardised algorithm for the measurement of IOP to reduce both measurement noise and operator variability.

Our findings are based on data from selected trial populations, and further research is required to validate the model in the general population of those with OHT.

Further research is required to determine the optimal interval of testing required to detect the onset of glaucoma. Standardisation to a single index for quantification of the visual field, as measured by SAP, is needed. This would enable modelling of the variability components of perimetry in the context of OHT to determine the optimal frequency of testing to detect glaucoma.

Chapter 7

Eliciting preferences for alternative monitoring services using a discrete choice experiment

Background

In choosing an optimal monitoring service for patients with OHT it is important to take account of individuals' preferences. This chapter used a discrete choice experiment (DCE) to investigate such preferences via a web-based online survey.

Objectives

The objectives of the DCE were:

- to investigate what characteristics of a monitoring service matter to individuals, as well as the relative importance of these characteristics
- to assess WTP for these individual characteristics as well as WTP for different configurations of monitoring services
- to feed the WTP estimates into an economic evaluation model (presented in *Chapter 8*).

Developing the discrete choice experiment

Discrete choice experiments are an attribute-based measure of benefit. The technique is based on the premise that goods or service can be described by their characteristics (attributes). Further, the extent to which an individual values those goods or that service depends on the levels of these attributes.[234,235] Participants are presented with a number of choices, each describing a series of attributes at different levels. The relative importance that individuals place on these attributes can be assessed by changing the levels of the attributes and asking participants to make their choice again. By including a price proxy (cost attribute) in the DCE, the monetary value of any given service can also be estimated. This monetary valuation, known as WTP, can be used within an economic evaluation as will be described in the next chapter. The process followed to develop the DCE is presented in *Figure 58* and described in detail below. Full ethical approval for this substudy was obtained from the North of Scotland Research Ethics Committee in June 2009 (REC reference number: 09/S0802/67).

Developing the attributes

The first stage of the DCE is to derive the attributes and associated levels. Although DCEs are a quantitative method to model preferences, qualitative methods are useful when defining attributes and levels (and developing the questionnaire).[236] A mixed methods approach was used to derive attributes and levels, integrating findings from an advisory panel and a focus group with relevant glaucoma health outcome attributes from an existing glaucoma utility measure.[4]
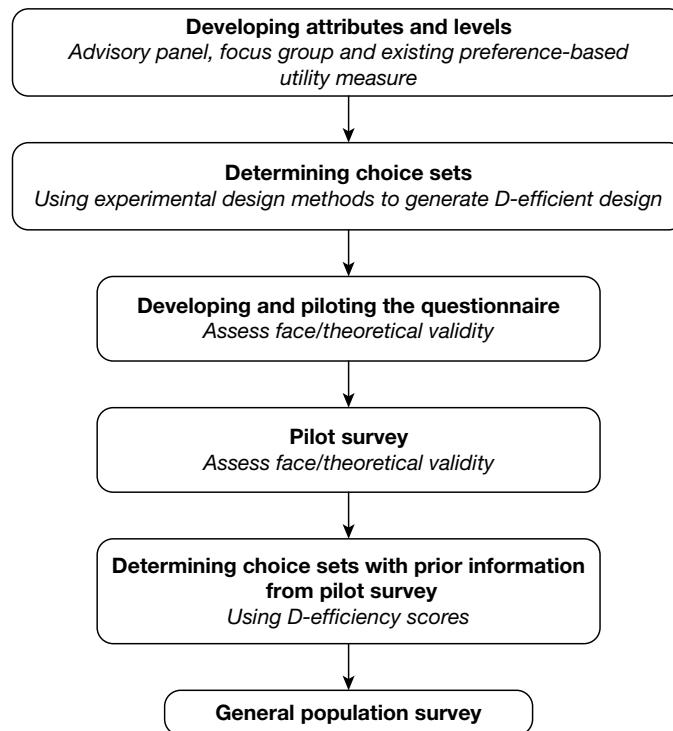
```
┌─────────────────────────────────────────────┐
│       Developing attributes and levels        │
│ Advisory panel, focus group and existing      │
│        preference-based utility measure        │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│            Determining choice sets             │
│ Using experimental design methods to generate  │
│              D-efficient design                │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│   Developing and piloting the questionnaire    │
│       Assess face/theoretical validity         │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│                 Pilot survey                   │
│       Assess face/theoretical validity         │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│  Determining choice sets with prior information │
│              from pilot survey                 │
│            Using D-efficiency scores           │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│          General population survey             │
└─────────────────────────────────────────────┘
```

FIGURE 58 Process for determining attributes and levels for the DCE.

Advisory panel

An advisory panel was convened to identify potential 'attributes' (characteristics influencing preferences for alternative monitoring services) and thus inform the framework (topic guide) for the planned focus group discussion with service users. Participants for the advisory panel were purposively selected to provide a broad UK perspective on managing OHT, but were not directly involved in the project. The panel included two service users [people under surveillance for OHT (SOH), Grampian, Scotland], one optometrist (academic optometrist, England), one non-clinical health service manager (ophthalmology service manager, secondary care, England), one specialist glaucoma nurse (secondary care, Scotland), one ophthalmologist (glaucoma specialist, England) and one community optometrist (Scotland). Six members of the local study team attended the advisory panel meeting (chief investigator, ophthalmologist, three economists and study research fellow). The meeting was recorded and transcribed. Minutes of the meeting with key points of agreement were circulated to participants following the meeting.

Discussion was driven by the participants on the desired elements of a monitoring service. Key themes (attributes) are summarised in *Table 23*.

Focus group

The findings from the advisory panel (see *Table 23*) were used to develop the topic guide for the focus group discussion with users of a service, the Grampian Glaucoma Referral and Monitoring Scheme Service. This service started in June 2004 and was developed to reduce unnecessary referrals to a hospital glaucoma clinic and to develop a community monitoring scheme for people at risk of developing glaucoma but not requiring treatment. Potential participants were identified from this scheme. Both men and women with OHT (who may or may not be undergoing treatment) or suspected of having OHT, of different ages and geographical locations within the Grampian area, were purposely sampled.

TABLE 23 Summary of attributes that the advisory panel felt were important in a monitoring service

Attribute	Discussion
Risk	Loss of vision causing, for example, loss of the ability to drive
Trust/confidence	Trust/confidence in who was seen. It was noted that patients are influenced by advice from their optometrist. Continuity of care would be preferable and this would contribute to building up a personal relationship and, therefore, an increase in trust
Time/cost of test	Some people might be more concerned about time commitments rather than cost; however, people value their sight and may be willing to give their time and money to maintain their sight
Place of testing	People would prefer not to have to go to a hospital for appointments because of parking problems and it would be easier if the location was on a bus route; however, people will go wherever they are told is best. Although location was important, overall confidence in a person was felt to be more important
Type of appointment	The merits of a fixed monitoring appointment arranged by the monitoring service rather than patient-arranged appointments were discussed. It was felt that a person's availability and flexibility would be affected by whether or not they were employed
Frequency of appointment	The number of visits was felt to be of lesser importance than the serious consequences (e.g. visual loss) that might result if not seen often enough. It was noted that in England 'low risk' individuals are recommended to seek testing about once a year in the community. In Scotland the community optometrist could potentially see the patient two to three times per year with visual field assessment and clinical assessment (measurement of IOP) if there was concern about the results of previous tests
Length of appointment	Length of appointment was felt to be important. It was noted that people might prefer to see a specialist nurse if it involved a shorter waiting time
Type of test	Type of test was felt to have an influence on the overall appointment. From a patient perspective it was noted that one might undergo any recommended test to avoid going blind

Participants were approached by post with an invitation letter from the consultant ophthalmologist and an information leaflet explaining the purpose of the study. In total, 93 people were invited and 11 agreed to participate in the study. Of these 11, six were able to attend on the suggested date for a focus group. In the event, five attended on the specified date, in September 2009. Summary statistics for the participants are presented in *Table 24.*

Pictorial cards were used to represent the most common tests for monitoring OHT (measuring IOP, looking at the back of the eye, measuring the field of vision and photographing the back of the eye) and to describe the different stages of progression of OHT and the likelihood of progression. This was based on a 5-year risk of conversion from OHT to glaucoma of 10%[64] and a 0.65 treatment effect[237] (i.e. treatment reduced the risk of conversion to 65% of the risk faced by those who were not treated) for those opportunistically detected with OHT.

Before the focus group, the topic guide and the stimulus material (pictorial cards) were piloted among six members of the Health Services Research Unit and Health Economics Research Unit at the University of Aberdeen. In general, the content of the topic guide was easily understood. Feedback regarding the content and format of the pictorial cards was incorporated to improve their value in promoting discussion among the group.

The discussion in the focus group was facilitated by the main researcher (ABP, who also acted as moderator), an economist (RH) and an ophthalmologist (AA) to explore participants' views on how a monitoring regime could be organised. Following initial introduction, the moderator (ABP) explored participants' experiences of how they were monitored and used the pictorial cards to introduce information about different characteristics that a monitoring service might have. Participants were encouraged to discuss their views and express their preferences about different characteristics of a monitoring programme, as well as discussing possible trading between attributes (i.e. how much of one attribute would they be willing to give up to get more of another). Throughout the discussion, the facilitators answered questions on current monitoring services in the UK, risk of progression of the disease and cost, as required.

TABLE 24 Summary statistics of focus group participants

Participant	Age (years)	Gender	Location	On treatment
P1	61	Male	Rural Aberdeenshire	No
P2	54	Male	Aberdeen city	Yes
P3	67	Male	Aberdeen city	No
P4	59	Female	Rural Aberdeenshire	No
P5	65	Female	Aberdeen city	No

The discussion was audiotaped and subsequently transcribed and analysed independently by two researchers (ABP, RH), using a modified framework approach.[238] After initial familiarisation with the transcripts, the data were grouped according to common themes that reflected the main research questions and key issues which transpired from the discussion. These themes emerging are described below.

Communication between the health-care provider and service users

Most participants expressed the view that more communication was needed in the current monitoring scheme. They believed that this was a vital characteristic for a monitoring service to be effective. Examples included the information given at the point of diagnosis and an awareness of the consequences of not being treated. Furthermore, participants believed that lay term explanations about additional steps in the care pathway were needed, including the reasons behind any tests and an explanation of what the tests involve:

> I didn't know anything about this hyperocular hypertension and I still don't know much about it either or this glaucoma thing.
>
> (P3)

> I find that with a lot of these tests the optician goes ahead and does it and doesn't really explain why they're doing it.
>
> (P5)

Location

When discussing where participants would like to be monitored, the two locations explored were their local optician and the hospital. Their preferred place for monitoring was influenced mainly by the equipment available and the skills of the staff performing the test. However, there was not a clear preference for the type of health-care professional that participants would like to be monitored by:

> I felt quite comfortable at Optician X. I felt they had the time because you're pushed for time when you have anything done at the hospital. I felt comfortable, relaxed and explained everything as well.
>
> (P4)

> She was very professional and if she hadn't been wearing that tag saying 'nurse' you would not know that she wasn't a – she was a sort of pillar rather than a full top medical person definitely. No issues at all.
>
> (P2)

Other factors that influenced individual preferences for location/provider were the time taken to attend the appointment, which varied depending on whether the person was working full-time or

was retired, accessibility of the location and parking facilities. Convenience for the operator was also highlighted as an influential factor:

> I'm quite happy to come into town. I just flash my bus pass.
>
> (P1)

> I'm willing to go just wherever it was convenient for the folk that were doing the examination for me.
>
> (P3)

Continuity of care was also perceived as an important aspect of monitoring:

> It would be good to see the same person.
>
> (P5)

Health outcomes and chance of developing glaucoma

Participants felt that preserving 'eyesight' was very important. For example, although participants did not particularly like the tests for measuring IOP, or the visual field test, both being integral to a surveillance programme, participants would still attend for monitoring tests.

> I would still have it … Of all the senses your eyesight is the most important one.
>
> (P1)

> If it's going to help things, of course you would.
>
> (P3)

> Eye sight is too precious.
>
> (P5)

The chance of developing glaucoma, particularly of losing vision, was identified as important. Participants perceived that risk increases when 'people do not turn up for an eye check', and cost played an important role (see *Willingness-to-pay explanation*):

> The ordinary person, you know, who thinks they don't need glasses or they don't bother about glasses – mainly because of the cost – there is a lot of people don't go to Optician Y and all that you know because you're paying £200 for a pair of glasses so there's a lot of people don't go and that's probably the people who end up the worst affected.
>
> (P3)

One participant expressed concern about the apparent lack of treatment effect and questioned the efficacy of the treatment provided in current practice as described in the quote below:

> What these figures are saying to me is that there's not a particularly effective treatment to prevent advanced glaucoma … Yes. I'm surprised at the small difference between not treated and treated in advance glaucoma.
>
> (P2)

For others, the lack of awareness of risk and its consequences if not treated was what was important to them and they concluded that more information at the point of diagnosis may have an impact on perception of risk. The participants believed that, by using pictorial cards representing the health risk information, people with OHT may gain a better understanding of the concept of risk and the consequences of monitoring:

Just show them that (referring to the risk card). Say, 'here's where you are, here's where you could be'.

(P2)

To fully explore the health outcome in the DCE this theme was expanded, building on our previous work with glaucoma patients,[4] to include further attributes that reflect a patient's perspective of the health impact of glaucoma. An existing utility measure, the Glaucoma Utility Instrument,[4] was refined. This instrument was developed based on focus group studies of patients with existing glaucoma. The risk of developing glaucoma was redefined as three attributes (risk of developing: glaucoma, severe glaucoma and visual impairment). The remaining two dimensions in the Glaucoma Utility Instrument, describing local and systemic side effects of treatment for OHT and glaucoma, were combined as one attribute (unwanted effects of treatment) to capture the consequences of treatment within the DCE.

Summary of attributes

The final attributes for the DCE are shown in *Table 25*. They include four attributes of health outcome from the Glaucoma Utility Index and two attributes of the monitoring process (communication and understanding; location). When a price proxy (cost attribute) is included in the DCE, WTP, a monetary measure of benefit, can be calculated for changes in individual attributes as well as for different configurations of a monitoring programme.[235,239,240] This is useful from a policy perspective, allowing estimation of the maximum amount of money individuals are willing to pay to participate in different monitoring programmes. We therefore also included a price proxy.

The next step was to define levels for each of these seven attributes.

Defining levels of the attributes

The levels for each attribute to be included in the DCE were informed by the focus group discussion, the literature, an existing economic model,[241] the existing glaucoma utility measure[4] and expert opinion within the study team.

Attributes 1, 2 and 3: risk of developing glaucoma, severe glaucoma and visual impairment

Based on the preliminary advisory panel and focus group discussions, a description of glaucoma, with a definition of each disease stage, was developed. The risk of developing glaucoma was also illustrated. This is shown in *Figure 59*. The levels for the risk of developing glaucoma were determined using an existing economic model, tailored for individuals with OHT[241] progressing to the different stages of glaucoma (glaucoma, severe glaucoma and visual impairment). All

TABLE 25 Attributes for the DCE

Attribute	
1	Risk of developing glaucoma
2	Risk of developing severe glaucoma
3	Risk of developing visual impairment
4	Unwanted effects of treatment
5	Communication and understanding
6	Location
7	Cost/price proxy[a]

a   Included to enable WTP, a monetary measure of benefit, to be estimated.

individuals under a monitoring regime were assumed to receive treatment. Those under no monitoring may be opportunistically diagnosed and treated. A 5-year risk of conversion to glaucoma was used for untreated individuals with OHT.[12] We assumed a treatment effect on risk of conversion, with treatment lowering the IOP for those with OHT and this reducing the risk of conversion to glaucoma compared with no treatment (HR 0.56, 95% CI 0.39 to 0.81). The risk of progression of glaucoma was also reduced by treatment[237] (HR 0.65, 95% CI 0.49 to 0.87). The economic model estimated the proportion of individuals with OHT likely to develop glaucoma, severe glaucoma and visual impairment each year and cumulatively over a 10-year follow-up. We used 95% CIs and two mid-points between these to obtain final levels.[4,237] The estimates produced in the economic model were reported as the number of people experiencing a particular level of severity over a 10-year period. Figures were developed for a hypothetical cohort of 10,000 individuals who started in the model with OHT. The values presented to DCE participants were rounded to improve understanding of risk by respondents whilst still maintaining representation of the model estimates (see *Table 26*).

Because individuals are known to have problems with understanding risk,[242] visual aids were used to help explain the concept of risk both in the focus group and subsequently in the DCE. Following Ancker and colleagues,[243] a graphical presentation was used in the DCE (*Figure 60*).

Attribute 4: unwanted effects of treatment
Levels were defined from the glaucoma-specific utility measure[4] and are detailed in *Table 26*.

Attribute 5: communication and understanding
There was some indication from the focus group that understanding the purpose of monitoring and feeling at ease were key components for monitoring. Therefore, following reflection on the findings of the focus groups and discussion within the study team four levels were developed for this characteristic:

- made me feel at ease and made sure I understood the purpose of monitoring
- made me feel at ease but did not make sure I understood the purpose of monitoring
- did not make me feel at ease but made sure I understood the purpose of monitoring and
- did not make me feel at ease and did not make sure I understood the purpose of monitoring.

Attribute 6: location
Defined as hospital eye clinic or local optician.

Attribute 7: cost per year (value)
To identify levels for the cost attribute a direct WTP question was included at the focus group stage. Here, participants were asked to consider the value that they would put on a monitoring service for people with raised eye pressure. Participants were presented with an example relating to the purchase of an article of clothing to help in understanding the concept of monetary value.

Willingness-to-pay explanation
You go to a shop and suddenly you see something that you really like, and think about buying it but suddenly you look at the price, say it cost £100 – too much – I don't want to pay that and you go home. Then you come back 2 or 3 weeks after that and the price of the good has changed – say, imagine, half price, £50 – and you really like it and you think '£50, I would pay £50 for that' and you buy it and you take it home. The economist will say that between £50 and £100 is the value that you put on that article.

Then, participants were asked how much they might be willing to pay for a service.

FIGURE 59 Describing the clinical condition. Images reproduced with permission from the Belgian Glaucoma Society (http://www.glaucoma.be/en/index.html).

Severe glaucoma

A condition where there is a marked restriction of the vision to the sides due to glaucoma, leading to some difficulties with daily activities.

Driving is not allowed (if glaucoma is severe in both eyes) because eyesight would not meet the required standards.

About 180 out of 10,000 people with ocular hypertension will develop severe glaucoma in 10 years if not treated.

Severe glaucoma

Following from the previous picture, there is more loss of vision to the sides.

Visual impairment

This is an advanced stage of glaucoma where one can still see to read as the central vision is not affected very much but a lot of vision to the sides is lost.

The main problem is difficulty getting around without assistance because the ability to see things at the sides is poor.

About 25 out of 10,000 people with ocular hypertension will develop visual impairment in 10 years if not treated.

Visual impairment

This is a more advanced stage of your condition, central vision remains but vision to the sides is lost.

FIGURE 59 Describing the clinical condition. Images reproduced with permission from the Belgian Glaucoma Society (http://www.glaucoma.be/en/index.html). (*continued*)
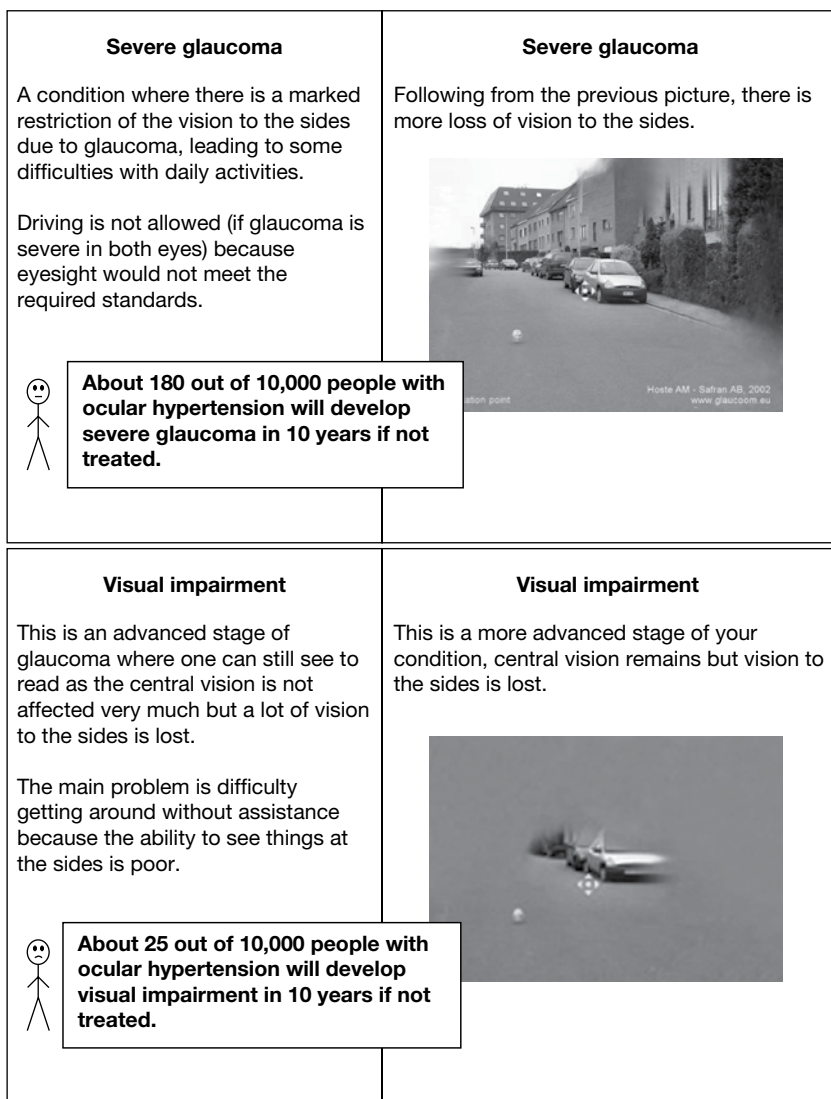
In other words:

Of 10,000 people with high eye pressure

Now

In 10 years' time…

About 1600 people will develop glaucoma

About 180 will develop severe glaucoma

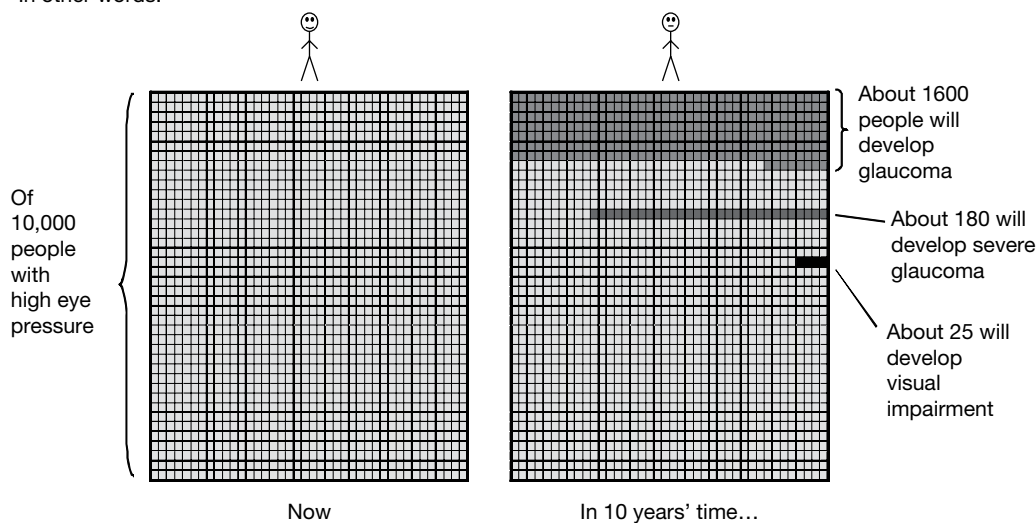About 25 will develop visual impairment

FIGURE 60 Graphical explanation of risk of developing glaucoma, severe glaucoma and visual impairment given to DCE participants.

Willingness-to-pay question

Now, I would like you to imagine the best monitoring programme. For some of you it is going to be at the hospital, being seen by an ophthalmologist, or it might be going to an optometrist every 6 months. Imagine the service that you really like, the service that you are really comfortable with. Then, obviously, the question is how much you might be prepared to pay for that type of service.

The monetary value of a service was a difficult concept to understand for the participants. Initially, participants expressed reluctance to pay for the service. However, after it was explained that this was an exercise to infer value, and that money is one way of looking at value, they engaged with the exercise.

Willingness to pay was based on previous experience with eye services or other health services in Scotland. Other factors considered when giving a value were a person's own health risk perception, financial circumstances, family history, previous experiences (good or bad) and current diagnosis (all suggesting engagement with the exercise). For example:

How much is the dentist? Is the dentist about £15? The same as the dentist.

(P4)

I would have paid it. I would have paid the £40 because my £40 gave me peace of mind.

(P3)

Something happened to my mother and in the light of today and prices and things, to cover National Health, I think I would be prepared to pay what about £40 for an hour's consultation, once a year.

(P5)

Well if you're wanting a figure, I would be comfortable in my wallet with £50.

(P1)

I think we're giving a kind of philosophical stun point here and the issue is not I think – to put a value on what you'd pay for a consultation then you need to know I guess you would maybe be prepared to pay a percentage of the cost of the consultation inclusive of everything involved, whatever that comes out to for a consultation. If it comes out to for a number £100 yes I would happily pay 10/20/30%, but without knowing the cost of the work then I can't really put a figure on it but perhaps the question should be 'would you pay or would you not pay' and I would pay – '…because of the situation I am in I would pay but it would be with extreme reluctance.

(P2)

Based on this discussion participants indicated values ranging from £15 (P4) to £50 (P1), with intermediate values of £30 (P2) and £40 (P3, P5). Such values provide a guide for levels of the cost attribute. However, given the small sample size of the focus groups, the levels for the cost attribute were extended beyond the maximum stated within the focus group, with an upper range of £70.

The final set of attributes and levels are summarised in *Table 26*.

Determining choice sets

The six attributes with four levels and one attribute with two levels resulted in 8192 possible choice sets (i.e. $4^6 \times 2^1$). Experimental design techniques were employed to reduce the number

TABLE 26 Summary of the monitoring attributes and levels

Attribute included in the DCE	Value for each level included in the DCE	No monitoring alternative
Risk of developing glaucoma		
Number of people out of 10,000 developing glaucoma in 10 years	740; 960; 1190; 1410	1600
Risk of developing severe glaucoma		
Number of people out of 10,000 developing severe glaucoma in 10 years	25; 60; 100; 130	180
Risk of developing visual impairment		
Number of people out of 10,000 developing visual impairment in 10 years	2; 6; 10; 15	25
Unwanted effects of treatment	None – means that you have not noticed any discomfort or difficulties	None
	Some – means that you have noticed occasional discomfort or difficulty	
	Quite a lot – means that you are aware of these discomforts or difficulties most of the time	
	Severe – means that you need or think you need additional treatment to control one or more of these difficulties	
Communication and understanding of information provided by the health professional	Made me feel at ease and made sure I understood the purpose of monitoring	Not applicable
	Made me feel at ease but did not make sure I understood the purpose of monitoring	
	Did not make me feel at ease but made sure I understood the purpose of monitoring	
	Did not make me feel at ease and did not make sure I understood the purpose of monitoring	
Location	Hospital eye clinic	No testing
	Local optician	
Cost per year (£)	15; 30; 50; 70	No cost

of choice sets to a more manageable number while still being able to infer WTP for all possible configurations. More specifically, SAS version 9.1.2 software[244] (SAS Institute Inc., Cary, NC, USA) was employed to generate a main-effects D-efficient design, ensuring that uncertainty around parameter estimates was minimised (by minimising the determinant of the covariance matrix).[245] This resulted in 32 choice sets.

An opt-out 'no monitoring' alternative was added to each of the 32 choice sets to allow individuals to choose the realistic option of not being monitored. The levels for the opt-out option were defined in consultation with experts in the field. An example of a choice set is shown in *Figure 61*. Respondents were presented with an example of a choice set before proceeding to the task themselves. The order of the choice questions presented for each participant was randomly generated.

Additional information in the questionnaire

In addition to the 32 choices derived from the experimental design, four rationality tests were added to test whether respondents were engaged in the experiment. The rationality tests applied Sen's contraction test.[246,247] Here, individuals were initially presented with a choice set involving three choices (monitoring service A, monitoring service B and no monitoring service). This choice was then repeated later in the questionnaire, with the choice set *contracted* to two options. These two options depended on what they chose initially, with the option they chose and one of the other options from the original three choices being presented. The individual should choose the same option when the choice set is *contracted*. The tests were included at four points within the questionnaire (choices 2 and 5, choices 11 and 15, choices 20 and 26 and choices 28 and

Which monitoring service would you choose? (please tick one box below)			
	Monitoring service A	**Monitoring service B**	**No monitoring service**
Number of people **out of 10,000** developing glaucoma in **10 years**	740	1410	1600
Number of people **out of 10,000** developing severe glaucoma in **10 years**	25	130	180
Number of people **out of 10,000** developing visual impairment in **10 years**	2	15	25
Unwanted effects of treatment	None	Severe	None
Communication and **understanding** of information provided by the **health professional**	Made me feel **at ease** and made sure I **understood** the purpose of monitoring	Did **not** make me feel **at ease** and did **not** make sure I **understood** the purpose of monitoring	Not applicable
Place of testing	Hospital eye clinic	Local optician	No testing
Cost per year	£15/year	£30/year	No cost
	Service A	**Service B**	**No monitoring**

(tick one box only)   ✓

FIGURE 61 Example of a DCE question.

34). This allowed testing of whether or not including a larger than average number of choices increased the cognitive burden or task complexity for respondents.[248–250]

If respondents 'fail' such tests, the question arises of what to do with them. Although deletion of such responses is common,[240] Lancsar and Louviere[251] argue that such deletion may be inappropriate because such responses may be valid, and random utility models are robust to errors made by individuals in forming and revealing preferences. Decisions on what to do with those who 'failed' were based on discussions once responses were analysed. Given that the rationality checks were not derived from the experimental design, the additional contracted questions were not included in the econometric analysis (because adding choices to the data matrix would compromise the statistical properties of the experimental design).

Before completing the DCE the following characteristics of the participants were collected: gender, age, region of UK, level of education (secondary school, college, university, none of the above and other), socioeconomic status (level of income per household and number of habitants) and experience of eye testing. Following the choices information was collected on general health, using the European Quality of Life-5 Dimensions (EQ-5D) (www.euroqol.org).

The full questionnaire is available from the authors.

Pre-piloting

The questionnaire was pre-piloted opportunistically among 10 members of staff from the University of Aberdeen. Comments on the format of the questionnaire were given in terms of the ordering of the different sections. One participant argued that having the 'easier' questions concerned with personal characteristics at the beginning of the questionnaire would help to ease respondents into the rest of the survey. Other useful comments related to the format were: 'useful to have an access link to the explanation of each characteristic and its levels throughout

the choice section in the same screen shot', 'provide a hyperlink to the example choice as a guide for answering the choices', 'highlight key words to emphasise meaning and capture attention'. Most people agreed that pictorial representation of disease stages together with the graphical risk explanation made information more understandable. Changes were made to the format of the survey accordingly.

Sample, pilot and main data collection

Given the poor response rates from unsolicited general population surveys,[252] a web-based online survey method was used to collect the data set, using a market research company (www.researchnow.co.uk/). This company contacted individuals registered with it, by email, and invited them to complete the DCE questionnaire. A financial incentive of £2 was promised for completion, credited to the respondent's personal account. Following discussions with the company, further modifications were made to the questionnaire to clarify meanings and to accommodate the online survey format. For example, respondents were given a separate icon to click for a reminder of definitions of attributes and levels in choices. Those invited to respond were aged > 18 years and living in the UK. The company's internal protocol was used to stratify by age and gender. Initially the questionnaire was piloted among a group of 183 respondents and provisional analysis conducted. Item response rate and the estimation using choice questions were analysed.

A novel component of this study was using the coefficients generated from the pilot work to further develop the experimental design. When deriving efficient designs there is often no prior knowledge about the potential values. It is thus assumed that parameters are zero. We made this assumption in our pilot work. A recent development is to use prior information, obtained from pilot work, to inform the design and improve statistical efficiency.[253–255] We obtained prior information from our pilot data to inform our final design, using SAS.[256] This changed the final set of choices respondents were presented with. The final version of the questionnaire is available from the authors. A sample size of 800 individuals was targeted for the main survey, providing a sufficient sample size to carry out subgroup analysis.

Econometric analysis

Responses to the choices within the DCE were analysed using a clustered conditional logit regression model, within Stata 11.0. The equation estimated is described in *Equation 13*.

$$
\begin{aligned}
V = {} & \beta_0 + \beta_1 \text{Risk\_earlyG} + \beta_2 \text{Risk\_SevereG} + \beta_3 \text{Risk\_visual\_impairment} \\
& + \beta_4 \text{Sideeffects\_some} + \beta_5 \text{Sideeffects\_quitealot} + \beta_6 \text{Sideeffects\_severe} \\
& + \beta_7 \text{CommUnder(at\_ease/not\_under)} + \beta_8 \text{CommUnder(not\_ease/under)} \\
& + \beta_9 \text{CommUnder(not\_ease/not\_under)} + \beta_{10} \text{Location} + \beta_{11} \text{Costperyear} + \varepsilon
\end{aligned}
\qquad \text{[Equation 13]}
$$

where utility from the monitoring service is represented by *V*, characterised by different combinations of the attribute levels (defined in *Table 26*), and ε is the error term (reflecting unobservable factors). β_0 is a constant reflecting the general preference for a monitoring service. If positive this would indicate a general preference for a monitoring service, and if negative a general preference not to have such a service (everything else being equal).

Dummy variables were employed for *side effects* (reference = none), *communication and understanding* (reference = not feeling at ease and not understood) and *location* (reference = community). The β parameters for these dummy variables indicate the marginal

value of a change from the reference category to a given level, that is, β_4 indicates the marginal value of moving from no unwanted treatment effects to some unwanted treatment effects.

The β parameters for the continuous variables, *risk* and *cost*, indicate the marginal value of a unit change in these attributes, that is, for the conversion to early glaucoma risk attribute the marginal value of reducing the number of people who convert to early glaucoma by 1 person in 10,000 over a 10-year time period. The parameter β_{11} on the cost attribute indicates the value of a £1 change in costs per year.

Willingness to pay

Drawing on *Equation 1*, the WTP for a marginal change in one of the attributes of a DCE is estimated as[235]:

$$-\left(\frac{\beta_i}{\beta_{\text{costs per year}}}\right) \qquad\qquad \text{[Equation 14]}$$

Thus, for example, from *Equation 13*, $-\beta_1/\beta_{11}$ indicates the WTP per year to reduce the number of people who convert to early stage glaucoma by 1 out of 10,000 over 10 years. Similarly, $-\beta_4/\beta_{11}$ indicates the WTP for a movement from no unwanted effects of treatment to some unwanted effects. WTP was estimated for a marginal change in all statistically significant attributes, and 95% CIs were obtained from bootstrapping with 1000 replications.[257]

Following estimation of the WTP for marginal changes in individual components of a monitoring service, overall WTP was estimated for two illustrative monitoring programmes: A (plausible best case scenario for monitoring, achieving the maximum treatment effect for OHT and OAG individuals)[4,237] and B (plausible worst case scenario for monitoring). Both, described in *Box 7*, were compared with no monitoring.

Following estimation for the aggregate model, subgroup analysis was conducted to investigate how preferences/WTP vary across different sections of the population. Preferences for monitoring are likely to vary according to age – age 50 years generally marks the onset of presbyopia (need for reading glasses) and is often the first experience of having a 'sight' test. However, it may be that as individuals get older they are less concerned about relatively low 10-year risks of conversion to glaucoma, severe glaucoma and becoming visually impaired as they

BOX 7 Descriptions of monitoring programmes A and B

Monitoring programme A

- Avoiding 860 per 10,000 individuals converting to early glaucoma in 10 years
- Avoiding 155 per 10,000 developing severe glaucoma over 10 years
- Avoiding 23 per 10,000 developing visual impairment over 10 years
- No unwanted effects of treatment
- Health-care professional makes one feel at ease and ensures understanding of the purpose of tests

Monitoring programme B

- Avoiding 190 per 10,000 individuals converting to early glaucoma in 10 years
- Avoiding 50 per 10,000 developing severe glaucoma over 10 years
- Avoiding 10 per 10,000 developing visual impairment over 10 years
- Severe unwanted effects of treatment
- Would not feel at ease or understand purpose of tests

believe it unlikely to affect them in their lifetime. Age was therefore investigated according to three groups (< 50, 50–65 and 65+ years).

Results

Characteristics of respondents

A total of 814 individuals responded to the questionnaire; *Table 27* shows their demographic characteristics. All individuals were adults, 49% were male and 50% were >45 years old. The sample was roughly evenly distributed in terms of education between those completing secondary, college and university level. Participants from England constituted 79% of the sample (645); 12% were from Scotland (97), 6% from Wales (48) and 3% from Northern Ireland (24).

In terms of experience of eye tests, 618 individuals (76%) had had their eyes tested within the previous 3 years, whereas 81 (10%), 40 (5%) and 54 (7%) had had them tested between 3 and 5 years previously, between 5 and 10 previously and >10 years previously, respectively. A total of 21 (3%) had never had an eye test.

Rationality tests

Table 28 shows the number of individuals who failed one or more of the rationality tests. Only three people failed all four consistency check questions. Given this, the regression analysis included all individuals.

Regression results

Table 29 shows the results for the conditional logit regression analysis. The positive constant implies a general preference for being monitored, suggesting that, everything else being equal, respondents would prefer to take part in a monitoring programme.

All coefficients, other than hospital setting, were significant at the 1% level, suggesting that they all influence preferences for monitoring.

The results show strong theoretical/face validity, with all regression coefficients having the expected sign and strength. For instance, participants were less likely to choose a monitoring programme with a higher risk of conversion to glaucoma, severe glaucoma and visual impairment (as indicated by the negative signs). Furthermore, the absolute size of a coefficient is higher the worse the attribute, that is, becoming visually impaired is worse than conversion to glaucoma, as indicated by the higher negative value (–0.65 compared with –0.03). Likewise, for unwanted effects of treatment, participants were less likely to choose a programme that would result in some, quite a lot or severe side effects than one with no unwanted effects, with severe side effects having the largest absolute level.

Given that the reference category for communication and understanding was 'not feeling at ease and did not understand', positive coefficients were expected as any movement to another level would be expected to be an improvement on the reference category. The results suggest that respondents do not differentiate between 'made me feel at ease but did not make sure I understood the purpose of the tests' and 'did not make me feel at ease but made sure I did understand the purpose of the tests' ($\chi^2 = 0.05$, $p = 0.83$). As expected, 'felt at ease and understood' was the most preferred level within this attribute (as indicated by the higher coefficient).

The cost attribute had the expected negative sign, suggesting that respondents prefer to pay less for a given monitoring service.

TABLE 27 Main survey sample demographic characteristics

Participant characteristics	*n*	%
Total sample size	814	100
Gender, male	400	49
Age (years)		
18–24	130	16
25–34	129	16
35–44	152	19
45–54	128	16
55–64	115	14
65+	160	20
Education (completed level)		
Secondary school	242	30
College	296	36
University	258	32
Other	18	2
UK region		
England	645	79
Scotland	97	12
Wales	48	6
Northern Ireland	24	3
Household income (£)		
<£19,999	240	29
Between £20,000 and £39,999	298	37
Between £40,000 and £59,999	131	16
Between £60,000 and £79,999	54	7
≥£80,000	91	11
General health		
Mean EQ-5D score (SD)	814	0.82 (0.23)

TABLE 28 Failure of rationality tests

Number of fails	Frequency	%	Cumulative percentage
0	556	68.30	68.30
1	195	23.96	92.26
2	49	6.02	98.28
3	11	1.35	99.63
4	3	0.37	100
Total	814	100	

Table 30 shows the WTP for individual attributes. Marginal WTP is defined as the monetary value of a unit change in a given attribute. This is calculated as the ratio of the attribute of interest divided by the negative of the coefficient on the cost attribute. So, everything else being equal, individuals were willing to pay £28 per year for a monitoring programme for OHT [0.50811/–(–0.01843)]. This positive value indicates a general preference to be monitored.

Health outcomes were defined as the number of people out of 10,000 with OHT developing glaucoma/severe glaucoma/visual impairment within a 10-year period. Respondents were willing to pay £0.03 per year for a 1 in 10,000 person reduction in the number of people converting

TABLE 29　Main survey multinomial logistic regression results (fixed effects)

Attribute	Coefficient	95% CI
Alternative specific constant (reference: no monitoring)	0.508[a]	0.3398 to 0.6764
OHT individual's 10-year risk of		
Conversion to early glaucoma	−0.001[a]	−0.0006 to −0.0004
Progressing to severe glaucoma	−0.001[a]	−0.0016 to −0.0007
Becoming visually impaired	−0.012[a]	−0.0159 to −0.0081
Unwanted treatment effects (reference: none)		
Some	−0.286[a]	−0.3308 to −0.2406
Quite a lot	−0.620[a]	−0.6919 to −0.5476
Severe	−1.094[a]	−1.1961 to −0.9919
Communication and understanding (reference: did not feel at ease and did not understand)		
Felt at ease and understood	0.861[a]	0.7690 to 0.9540
Felt at ease but did not understand	0.475[a]	0.4218 to 0.5278
Did not feel at ease but understood	0.480[a]	0.4172 to 0.5437
Hospital setting (reference: community)	0.025	−0.0098 to 0.0595
Cost per year	−0.018[a]	−0.0201 to −0.0167

a　Significant at the <1% level.
Number of observations = 78,144.
Clusters (individuals) = 814.
Log-likelihood = −26,402.
Pseudo R^2 = 0.0774.

TABLE 30　Willingness to pay for DCE attributes

Attribute	Marginal WTP (£)	95% CI
Alternative specific constant (reference: no monitoring)	27.57	18.93 to 35.38
Outcomes		
One OHT person per 10,000 developing glaucoma within 10 years	−0.03	−0.03 to −0.02
One OHT person per 10,000 developing severe glaucoma within 10 years	−0.06	−0.09 to −0.03
One OHT person per 10,000 becoming visually impaired within 10 years	−0.65	−0.90 to −0.44
Unwanted treatment effects (reference: none)		
Some	−15.50	−18.68 to −12.92
Quite a lot	−33.63	−39.15 to −28.54
Severe	−59.36	−68.29 to −51.50
Communication and understanding (reference: did not feel at ease and did not understand)		
Felt at ease and understood	46.75	40.40 to 54.42
Felt at ease but did not understand	25.76	22.25 to 29.93
Did not feel at ease but understood	26.07	22.05 to 31.05
Hospital setting (ref: community)	1.35	−0.53 to 3.20

to glaucoma over the next 10 years [−0.0005/−(−0.01843)]. Thus, if 860 people with OHT per 10,000 avoided progressing to glaucoma within 10 years (as might occur when moving from no monitoring service to monitoring service A), the value would be £25.80 (£0.03×860). Similarly, if 155 people with OHT per 10,000 avoided progressing to severe glaucoma within 10 years the value would be £9.30 (£0.06×155), and if 23 people per 10,000 avoided becoming visually impaired within 10 years the value would be £14.95 (£0.65×23).

As expected, unwanted effects of treatment reduce the value of any monitoring service. For example, WTP would be reduced by £15.50 if respondents experience some unwanted effects of treatment [−0.28573/−(−0.01843)]; £34 if they experience quite a lot of unwanted effects of treatment [−0.61979/−(−0.01843)]; and £59 if they experience severe unwanted effects of treatment [−1.094/−(−0.01843)].

If individuals feel at ease and understood they would value the monitoring programme at an additional £46; experiencing one of these components would increase value by approximately £26.

Table 31 shows the WTP for monitoring programmes A and B. Respondents were willing to pay £122 per annum for programme A, that is, they valued the monitoring programme up to the value of £122. However, programme B was far less desirable and they would have to be compensated by £17 per annum to undertake programme B (as indicated by the negative WTP). Failure to compensate would result in a very low take-up of such a monitoring programme.

Table 32 reports WTP (and 95% CIs) for the age subgroups. The *p*-value for the likelihood ratio chi-squared ($p = 0.000$) indicates that we can reject the null hypothesis that the age subgroup models are no better than the full model. In other words, there is heterogeneity in preferences that can be explained by age. In terms of WTP, those aged ≥ 50 years have a stronger preference to be monitored, with the 50–64 years age group having the strongest preference (as indicated by the higher alternative specific constant). These older age groups also had the strongest preference against unwanted side effects. However, they have the lowest preference for becoming visually impaired, with the preference being lower in the 50–64 years age group than in those aged ≥ 65 years. This may reflect an increasing fear of ill health, and therefore blindness, as you get older (although it does not reach the levels of the younger group, who have the strongest aversion to visual impairment, as expected). The 50–64 years age group also valued more highly the provision of communication and understanding, with preferences being consistently higher than for those who are > 65 years.

Discussion

We conducted a DCE to elicit preferences for alternative glaucoma monitoring programmes for individuals with OHT. We used a mixed methods approach to derive the attributes and levels, employing a project advisory group and focus group. We developed a composite measure of utility that took account of both patient experiences and health outcomes. Pictures were used to explain the impact of glaucoma on eyesight and graphs were used to explain the concept of risk. We pre-piloted and piloted the survey and used parameters generated from our pilot data to develop the statistical design for the main DCE survey. The results from the DCE support the advisory groups and focus groups, with all attributes included in the DCE, other than location, impacting on preferences. The results were theoretically valid, with all coefficients moving in the direction expected. Age was a predictor of preferences, with the 50+ age groups valuing monitoring more highly, as well as avoiding unwanted side effects and communication and understanding. They also had a lower aversion to the 10-year risk of becoming visually impaired (compared with the < 50 years age group).

TABLE 31 Willingness to pay for alternative monitoring programmes

Attribute	Regression coefficient	Marginal WTP (£)	Monitoring programme A		Monitoring programme B	
			Attribute level	WTP (£)	Attribute level	WTP (£)
Alternative specific constant (reference: no monitoring)	0.508[a]	27.57	Yes	27.57	Yes	27.57
10-year risk of						
Conversion to glaucoma of individuals with OHT	−0.001[a]	−0.03	−860[b]	23.29[c]	−190[b]	5.14[c]
Progressing to severe glaucoma for OHT individuals	−0.001[a]	−0.06	−155[b]	9.71[c]	−50[b]	3.13[c]
Becoming visually impaired for OHT individuals	−0.012[a]	−0.65	−23[b]	15.03[c]	−10[b]	6.54[c]
Unwanted treatment effects (reference: none)						
Some	−0.286[a]	−15.50	No	0.00	No	0.00
Quite a lot	−0.620[a]	−33.63	No	0.00	No	0.00
Severe	−1.094[a]	−59.36	No	0.00	Yes	−59.36
Communication and understanding (reference: did not feel at ease and did not understand)						
Felt at ease and understood	0.861[a]	46.75	Yes	46.75	No	0.00
Felt at ease but did not understand	0.475[a]	25.76	No	0.00	No	0.00
Did not feel at ease but understood	0.480[a]	26.07	No	0.00	No	0.00
Hospital setting (reference: community)	0.025[d]	1.35	No	0.00	No	0.00
Cost	−0.018[a]					
Total WTP				122.35		−16.98

a   Significant at the < 1% level.
b   Compared with no monitoring.
c   Figures do not match exactly because of rounding.
d   Significant at the < 10% level.

From a methodological perspective, a novel component of our study was the method of data collection – the use of an internet-based company with a panel of individuals who are willing to respond to surveys. This was in response to the low response rates in previous DCEs, with studies struggling to achieve response rates > 15%. The internet offers a promising alternative to conventional mail surveys, offering novel ways of communicating difficult and unfamiliar concepts (in terms of graphics and sound); guaranteeing response rates; offering relative ease of data collection and reduced costs, speed of response and the tailoring of questions to individuals' responses; ensuring that all questions are complete and therefore eliminating item non-response; and avoiding the need for manual data entry.[258] However, there are a number of issues that arose using this method of data collection. We do not have information on the number of individuals who attempted but did not complete the questionnaire. Also, the approach forces a response even in situations in which an individual finds it difficult to make one. Concern does exist over the generalisability of the results to the general population as well as the effect of such survey methods on the preferences elicited. Although our sample of respondents matches the general population in terms of age and gender mix, future work should explore representativeness with respect to others characteristics such as attitudes. Only two studies, to the authors' knowledge, have compared WTP estimates when using internet and mail surveys. Olsen,[258] in a study concerned with preferences for protecting different types of landscape from road encroachment, found no significant different between the WTP values generated. Similar results were found by Windle and Rolfe[259] when valuing future improvements in the Great Barrier Reef. This is clearly an important avenue for future research within health.

TABLE 32 Willingness to pay for subgroup analysis

| | WTP (£) (95% CI) | | | |
| | | Age | | |
Variable	Full model	< 50 years	Between 50 and 64 years	≥ 65 years
Number of observations	78,144	44,160	18,624	15,360
Clusters (individuals)	814	460	194	160
Alternative specific constant (reference: no monitoring)	28 (19 to 35)	23 (12 to 32)	35 (17 to 51)	38 (10 to 60)
10-year risk of				
Conversion to glaucoma of individuals with OHT	–0.03 (–0.03 to –0.02)	–0.03 (–0.03 to –0.02)	–0.03 (–0.04 to –0.02)	–0.03 (–0.05 to –0.02)
Progressing to severe glaucoma for OHT individuals	–0.06 (–0.09 to –0.03)	–0.07 (–0.1 to –0.03)	–0.02 (–0.08 to 0.04)	–0.09 (–0.17 to –0.02)
Becoming visually impaired for OHT individuals	–0.65 (–0.9 to –0.44)	–0.88 (–1.2 to –0.61)	–0.17 (–0.62 to 0.22)	–0.47 (–1.24 to 0.13)
Unwanted treatment effects (reference: none)				
Some	–16 (–19 to –13)	–13 (–16 to –10)	–16 (–24 to –11)	–24 (–36 to –16)
Quite a lot	–34 (–39 to –29)	–31 (–37 to –25)	–36 (–48 to –26)	–41 (–61 to –27)
Severe	–59 (–68 to –51)	–53 (–63 to –44)	–66 (–87 to –50)	–74 (–109 to –51)
Communication and understanding (reference: did not feel at ease and did not understand)				
Felt at ease and understood	47 (40 to 54)	36 (30 to 45)	62 (47 to 82)	64 (45 to 93)
Felt at ease but did not understand	26 (22 to 30)	22 (18 to 27)	30 (22 to 41)	35 (24 to 51)
Did not feel at ease but understood	26 (22 to 31)	18 (14 to 24)	37 (27 to 50)	40 (27 to 61)
Hospital setting (ref: community)	1 (–1 to 3)	1 (–1 to 3)	–1 (–6 to 4)	5 (–1 to 11)
LR test				
LR χ^2(24)	1846.23			
Probability $> \chi^2$	0.000			

LR, likelihood ratio.

From a policy perspective, although all attributes are important, there is a gradient on how the different stakeholders can influence them. For instance, the risk of developing glaucoma will depend on both the effectiveness of treatment and how the patient complies with treatment (and the latter may depend on other factors such as the unwanted effects of treatment). The effectiveness of a particular treatment will not be actionable by the individual but how much the person complies with treatment will be.

The importance of 'severe unwanted effects of treatment' and 'feeling at ease and understanding the purpose of monitoring' were major drivers of preferences and both are within the decision domain of the health-care professional. It is likely that severe effects of treatment are already considered as changes in medical treatment would occur in current practice when severe side effects are observed. However, the findings from the focus group suggest that community eye services might not have sufficient time to devote to each patient, with a resulting lack of explanation of the purpose of monitoring. Therefore, policy-makers/decision-makers should look at this area as a potential target for improvement.

The setting where monitoring takes place was not a major driver of preferences. This was also reflected in the focus group findings. Individuals appear to be prepared to attend a visit anywhere as long as a skilled testing service is available. Consequently, policy-makers might have a certain degree of freedom when deciding where to set up a monitoring programme as long as the quality of the service is not compromised.

Although avoiding glaucoma and its consequences were highlighted in the advisory panel and focus group as important to patients, when it comes to decision-making by the public, and in particular making choices about how the service could be configured, the marginal WTP value of reducing the frequency of developing glaucoma was lower than those for the patient experience factors. It should be remembered that the β parameter for the frequency attributes indicates the marginal value of a 1 in 10,000 person change in the number of people developing glaucoma/severe glaucoma/visual impairment, whereas for the dummy variable (unwanted treatment effects and communication and understanding) the β parameter indicates the marginal value of a shift from the reference category (i.e. from 'not feeling at ease' to 'feeling at ease and understanding'). This is shown clearly in *Table 30* where the value of a reduction of say 860 in 10,000 people progressing to glaucoma is £25.80. Thus, values are still relatively low.

One possible reason for this finding is that respondents had difficulties understanding risk levels.[11] Although we used a graphical representation to explain risk, problems of understanding may still exit. The small risk levels for the attributes may also have contributed to the relatively lower values generated as it is also known that valuing small changes in risks is even more challenging, and that as a result values generated are often insensitive to the scale of the risk change.[260] Future research is clearly needed.

The results may also relate to the sample. There is debate about whose (society or service users) values should inform the allocation of society's scarce resources. Current guidance for decision-making within the NHS recommends the use of society's preferences.[261] This approach has been previously used in a wide range of contexts.[252] It is, however, possible that sampling a higher-risk group would have led to higher values for changes in the risks of developing glaucoma/severe glaucoma and visual impairment. This is clearly an important area for future research.

Whatever the interpretation of the risk attributes, our results do point to the importance of the process of care in the delivery of care. Although economic evaluations have tended to focus on health outcomes when valuing benefits from alternative interventions, there has been increased debate about the importance of valuing factors beyond such narrow outcome measures. Very few studies have looked at the value of health outcomes and process attributes within one holistic measure.[4] In this study we attempted to do this, valuing health and patient experience attributes within one composite measure. de Bekker-Grob and colleagues,[252] in a review of the application of DCEs in health, noted the absence of DCEs within an economic evaluation decision-making framework. Indeed, QALYs continue to be the preferred valuation method when making recommendations regarding provision of alternative health-care interventions by bodies such as NICE.[261] However, the importance of patient experiences is being increasingly recognised,[262,263] and was clearly demonstrated in this study with 'communication and understanding' of the reasons for testing and an explanation of what the tests involve being a major driver of preferences. In the next chapter we incorporate our broader measure of value into an economic evaluation framework.

Conclusions

The DCE attributes of risk of developing glaucoma, severe glaucoma or visual impairment, unwanted effects of treatment, and communication and understanding, but not location, have a statistically significant impact on individuals' preferences. Respondents also valued positively the fact of being monitored, irrespective of the particular form that this monitoring might take. The main drivers of individuals' preferences were the unwanted effects of treatment and communication by the health-care professional and an understanding of the purpose of monitoring. Age was a predictor of preferences, with those aged ≥ 50 years valuing monitoring and communication and understanding more highly than those < 50 years. Those aged ≥ 50 years also valued unwanted side effects of treatment more highly than those < 50 years.

Relatively low WTP values were obtained for the risk of developing glaucoma, progressing to severe glaucoma and visual impairment. Further research should compare these results with those of a DCE conducted for ocular hypertensive respondents as these individuals might have a better appreciation of what glaucoma is and might understand more fully the ramifications of visual field loss.

From the policy perspective, our study suggests that any changes to a surveillance programme for OHT should take into account the communication between health-care provider and patient and give time for the patient to understand the purpose of monitoring.

Chapter 8

Economic evaluation

Economic evaluation is the comparative analysis of alternative courses of action in terms of both their costs and their benefits.[264] Economic evaluations may take several forms: (1) cost-effectiveness analysis, in which costs are equated with a single natural or clinical measure of effectiveness; (2) cost–utility analysis, in which costs are equated against a composite measure that combines both quality and quantity of life; and (3) cost–benefit analysis, in which costs and benefits are measured in commensurate units, normally money. A further form of economic evaluation also exists, a cost–consequence analysis. Here, costs are equated against several different measures of outcomes. These outcomes may include both health and non-health effects. For this study, as illustrated in the results section in this chapter, we use each of these different approaches to explore the relative efficiency of alternative surveillance approaches.

Systematic review of economic evaluations of ocular hypertension surveillance programmes

We systematically searched for economic evaluations comparing the costs and outcomes of two or more pathways for surveillance of individuals with OHT, including situations in which no active or organised surveillance service is available. The search strategy can be found in *Appendix 1*. To be included, studies had to include costs (regardless of the way that these were estimated) and effects (regardless of the way that these were estimated). We found no economic evaluations that met the inclusion criteria (see *Appendix 10* for details). Two cost comparisons were identified that compared alternative pathways to manage individuals with stable OAG. Both of these were performed within a RCT context, one in the UK[265] and one in the Netherlands.[266] These studies evaluated alternative management schemes for people with OAG and were not directly relevant to policy decisions on surveillance for those with OHT. An abstract from conference proceedings reporting a study by van Gestel and colleagues[30] was retrieved. The abstract reports findings from an economic evaluation in which two follow-up pathways (i.e. treat all individuals with OHT or do not treat them until they convert to OAG) were defined. As this was an abstract it was not included in our systematic review but we used this evidence to inform one of our model pathways. Furthermore, the model reported within the NICE guidelines[1] focuses on alternative treatments [i.e. no treatment or medical treatment with beta-blockers (BBs) or prostaglandin analogues (PGAs)] for individuals with OHT and does not explicitly compare alternative monitoring pathways. Given the paucity of the evidence base, no definite conclusion on the cost-effectiveness of surveillance programmes for individuals with OHT could be drawn. There is a need for an economic evaluation of alternative surveillance programmes. In the following section we report the results of an economic evaluation of five potential pathways for surveillance of adults with OHT.

Economic evaluation of surveillance pathways for individuals with ocular hypertension

The relative efficiency of pathways for surveillance was assessed using a discrete event simulation model. This approach provided more flexibility when modelling the sequence of events involved in surveillance because it does not impose restrictive assumptions about when events might

occur and it allows the incorporation of heterogeneity between individuals in a modelled cohort compared with alternative approaches commonly adopted in economic evaluations (e.g. Markov models). *Figure 62* provides an overview of the process captured within the discrete event simulation model. This basic process is identical for each of the surveillance pathways compared.

The economic evaluation involved several stages, described in the methods section:

- Defining the interventions compared. The method followed to define the care pathways together with a description of the compared pathways are provided in *Developing the surveillance care pathways* and *Description of the surveillance pathways compared and outline of the model. Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway* outlines how these pathways are tailored to reflect differences in the underlying risks of those being monitored; and *The sequence of treatment* describes treatments for OHT and OAG.
- Structuring the discrete event simulation economic model (*Description of the discrete event simulation economic model*) to capture the essential elements of the disease and the care provided as identified in the conceptual models that form the care pathways described in *Developing the surveillance care pathways.*
- Paramaterising the model, specifically:
  – estimation of parameters describing the natural history of OHT and OAG (*Estimation of parameters for the natural history of ocular hypertension and open-angle glaucoma*)
  – estimation of parameters used to modify the natural history of OHT and OAG (*Parameters used to modify natural history of ocular hypertension and open-angle glaucoma*)
  – estimation of parameters used to describe the costs of surveillance and treatment (*Parameters used to describe the costs of surveillance, diagnosis and treatment*)
  – estimation of parameters used to value the health effects and other benefits of surveillance (*Health state utilities and benefit valuations*).
- Calibrating the model (*Model calibration and validation*) to ensure that the model outputs are credible and informative for decision-making.
- Data analysis (*Developing the surveillance care pathways*) describing how estimates of costs, effects and measures of relative efficiency were produced.

The results section presents the model calibration results, cost-effectiveness analysis and cost–utility and cost–benefit analyses. The final section provides a summary of the key finding, the strengths and limitations of the analysis and key uncertainties that remain.

Methods
Developing the surveillance care pathways
The clinical care pathways for surveillance of patients diagnosed with OHT were developed by the Aberdeen project management group. The group included three ophthalmologists (AAB, AT and JB), three health economists (LV, RH and MR), a statistician (JC) and two project research fellows (APB and KB). Over a number of meetings the group reviewed national guidelines and
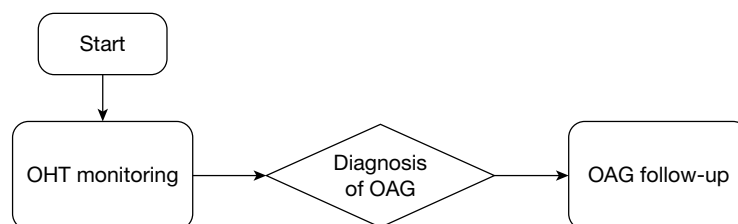

FIGURE 62 Overview of the process captured by the economic model.

used information from a review of economic evaluations of monitoring for OHT (*Systematic review of economic evaluations of ocular hypertension surveillance programmes*), a review of risk prediction tools (see *Chapter 4*), the optimal criteria for monitoring (see *Chapters 5* and *6*) and public preferences for monitoring (see *Chapter 7*) to determine the most appropriate care pathways to be compared in the model.

Five pathways were developed:

1. SOH in a primary care (community) setting
2. SOH in a hospital eye service setting
3. 'treat all' pathway
4. NICE guidelines (intensive)
5. NICE guidelines (conservative).

Public preferences for monitoring (see *Chapter 7*) and the evaluation of optimal monitoring criteria (see *Chapters 5* and *6*) informed a new pathway of care defined as the SOH pathway. This pathway was developed for both a community eye care setting (optometry) and a hospital eye service setting.

In the 'treat all' pathway, all individuals with an IOP > 21 mmHg would be treated with a PGA with advice to have their IOP checked annually at their community optometrist. This pathway was informed by a modelling evaluation by van Gestel and colleagues,[30] which suggests that this pathway is cost-effective in a Dutch context.

The comparator pathways were defined by the NICE guidance.[1] Detailed specification of these pathways can be found in the next section. In brief, the NICE guidelines present a range of recommended monitoring intervals for patients dependent on both the initial risk stratification and treatment recommendation. We developed two pathways for current care from the NICE guidelines: 'NICE intensive', representing monitoring on the more intensive side of the guidelines' recommended monitoring frequency (4-monthly to annually depending on initial risk stratification), or 'NICE conservative', representing the more conservative end of the recommended frequency of monitoring (6-monthly to biennially).

Description of the surveillance pathways compared and outline of the model

The SOH pathways were defined for a primary care (community) setting and a hospital eye service setting. In both of these pathways baseline IOP would be measured twice and a mean value calculated. Also, the frequency of the surveillance visits was taken to be 2 years. This interval was based on the work conducted to estimate the signal-to-noise ratio for measurement of IOP, reported in detail in *Chapter 6*. Other features of the pathways, such as the level of communication and understanding between the health-care professional and the patient or the formula for the estimation of the risk level that would trigger treatment, were informed by professional experts and public preferences for surveillance (see *Chapter 7)* and the comparison of IOP measurement agreement between alternative tonometers (see *Chapter 5*), respectively.

The 'treat all' pathway reflects a minimum level of care of people with OHT (i.e. no explicit or organised surveillance). In this pathway, all individuals with OHT receive treatment and these individuals are advised to seek, and assumed to take up, an annual IOP check-up from a provider based in the community (e.g. community optometrist). If there is a < 15% reduction in IOP from baseline (model entry) individuals are referred to secondary care.

The NICE guidelines[1] provide a broad description of the range of management options but do not provide a sufficiently precise definition of the management that would be provided to an individual with a given set of risk characteristics; such detail is required to construct a model. Nevertheless, using the descriptions provided within the guidelines we developed two pathways for current care: 'NICE intensive' or 'NICE conservative'. These pathways differed in terms of the time intervals between surveillance visits. For instance, with a range of 6–12 months stated within the NICE guidelines, the 'NICE intensive' pathway used a 6-month interval between surveillance visits whereas the 'NICE conservative' pathway used a 12-month interval between visits. For both of these surveillance pathways the location of surveillance was taken to be a hospital setting. Treatment was recommended as detailed in the NICE guidance.

These pathways are summarised in *Table 33*.

Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway

Table 34 gives a detailed description of the risk stratification rules and surveillance and treatment criteria for each pathway. Every simulated individual entering the model has confirmed OHT based on an IOP of > 21mmHg and no ocular comorbidity. The pathways described in the first two columns of *Table 34* are founded on the premise that an individual's risk of converting to OAG is based on a set of definable characteristics. In contrast to this the 'treat all' pathway, described in the third column of *Table 34*, is implicitly based on the premise that individuals, regardless of their underlying characteristics, face the same risk of converting to OAG.

The sequence of treatment

Within the model the initial treatment choice for each individual is defined by the individual's characteristics at the time when surveillance is initiated. Although the initial treatment may vary between individuals, based on a judgement of their risk of conversion to OAG, a common sequence of treatments was defined following discussions within the project team and with expert clinicians. Individuals will vary according to the point at which they enter this sequence (*Figure 63*). An individual with OHT would progress through these treatments if a judgement is made that IOP is not controlled. IOP is considered uncontrolled if the reduction from the IOP level at the start of the model is < 15% of the initial IOP [e.g. if an individual's IOP level at the start of the model is 26 mmHg, then an IOP level > 22 mmHg (26 – 26 × 0.15 = 22.1) after treatment has been initiated would be regarded as uncontrolled]. Individuals who convert to OAG would receive PGAs if they had previously received either no treatment or BBs. However, a proportion of individuals will receive surgery (trabeculectomy) if prior to conversion they had received treatment with PGAs or combination therapy. The sequence of treatments allows for a

TABLE 33 Surveillance pathways compared within the model

Pathway	Risk stratification	Surveillance criteria	Treatment
SOH (primary care)	Risk prediction algorithm (see *Chapter 4*)	NICE guidelines modified according to findings in SOH substudy B (see *Chapters 5* and *6*)	PGAs for medium- or high-risk individuals
SOH (hospital)	Risk prediction algorithm (see *Chapter 4*)	NICE guidelines modified according to findings in SOH substudy B (see *Chapters 5* and *6*)	PGAs for medium- or high-risk individuals
NICE guidelines (intensive)	NICE guidelines (based on age, CCT and IOP level)	NICE guidelines; using *minimum* intervals between surveillance visits	NICE guidelines
NICE guidelines (conservative)	NICE guidelines (based on age, CCT and IOP level)	NICE guidelines; using *maximum* intervals between surveillance visits	NICE guidelines
'Treat all'	IOP > 21 mmHg: no further risk stratification	IOP surveillance only once a year	PGAs for all with IOP > 21 mmHg

TABLE 34 Risk stratification rules and surveillance and treatment decision criteria for each pathway

The NICE pathways (intensive and conservative)	The SOH pathways (community/hospital)	'Treat all' pathway
Risk		
The criteria used to categorise individuals with OHT as low, medium or high risk of conversion to OAG are not explicitly stated	Used the best available risk prediction tool (see *Chapter 4*) to inform the choice of initial treatment of individuals	All individuals with IOP > 21 mmHg judged to be at high risk of conversion to OAG
Guidelines provide clear criteria to inform surveillance and treatment decisions, with decisions on surveillance and treatment made on CCT and level of IOP together with age[1]	Information on age, VCD ratio, IOP, CCT and a measure of visual field loss (PSD) combined using an algorithm based on the pooled OHTS-EGPS model[64] to calculate the 5-year risk of conversion to OAG	No further risk stratification considered
	The 5-year risk of developing OAG was grouped into three categories: low (< 6%), intermediate (6–13%) and high (> 13%)[13]	
Surveillance		
For those with untreated OHT, a full assessment is recommended every 6 or 24 months depending on risk (level of IOP or CCT – see *Table 35* for details)	All those starting treatment with a PGA, and those changing to a new medical treatment, have two consecutive (same visit) IOP measurements within 2 months of starting or changing a treatment	No active monitoring: individuals are advised (and assumed) to attend a community optometrist annually for measurement of IOP
For treated OHT, IOP measurement 2 months after initiating treatment is recommended. Full assessments are every 4, 6 or 12 months depending on risk (see *Table 35* for details)	Individuals are monitored every 2 years in either a secondary or primary care setting for the 'SOH hospital' and 'SOH primary care' pathways respectively	
	For the 'SOH primary care' pathway individuals would only be referred to secondary care if IOP was 'off target' or conversion to OAG being detected	
Treatment decisions		
Details of the criteria under which treatment is initiated are provided in *Table 35*	Individuals with low risk (5-year risk of conversion < 6%) are not treated	All individuals with IOP > 21 mmHg are treated with PGAs
Medical treatment is stopped when individuals reach 60, 65 or 80 years of age if they are taking BBs, PGAs or combination therapy, respectively	Individuals with intermediate or high risk (5-year risk of conversion ≥ 6%) are treated with a PGA	
The decision to stop treatment taken only if IOP remains on target and progression to OAG has not occurred. Finally, the decision about what initial treatment to be given is based upon the age and measures of CCT and IOP as defined in *Table 35*	If IOP off target the sequence of treatments is as outlined in *The sequence of treatment*	If IOP off target (< 15% reduction) from baseline (model entry) then individuals are referred to an ophthalmologist in a secondary-care setting
Care following conversion to OAG		
All surveillance and care once an individual has converted to OAG provided by an ophthalmologist in a secondary care setting	All surveillance and care once an individual has converted to OAG provided by an ophthalmologist in a secondary care setting	All surveillance and care once an individual has converted to OAG provided by an ophthalmologist in a secondary-care setting

PGA, prostaglandin agonist.

maximum of two trabeculectomies, but does not allow those who are not diagnosed with OAG to receive surgery.

A further restriction on use of treatment is that we assume that individuals can progress through the treatment sequence but if they are initiated on a treatment that is part way through the sequence then they cannot subsequently go back to a treatment earlier in the sequence. The possible changes in treatment for those who have uncontrolled OHT are described in *Table 36*.

TABLE 35 Treatment and follow-up frequency rules for individuals with OHT developed from the NICE glaucoma guidelines[1]

Treatment allocation criteria							
CCT	>590 μm		555–590 μm		<555 μm		Any
Untreated IOP (mmHg)	>21 to 25	>25 to 32	>21 to 25	>25 to 32	>21 to 25	>25 to 32	>32
Age (years)	Any	Any	Any	Treat until 60	Treat until 65	Treat until 80	Any
Initial treatment	No treatment	No treatment	No treatment	BB	PGA	PGA	PGA
Surveillance intervals							
NICE intensive pathway							
IOP only	NA	NA	NA	At 2 months	At 2 months	At 2 months	At 2 months
Full assessment[a]	Every 12 months	Every 6 months	Every 6 months	Every 6 months	Every 4 months	Every 4 months	Every 4 months
NICE conservative pathway							
IOP only	NA	NA	NA	At 2 months	At 2 months	At 2 months	At 2 months
Full assessment[a]	Every 24 months	Every 12 months	Every 12 months	Every 12 months	Every 6 months	Every 6 months	Every 6 months

NA, not applicable; PGA, prostaglandin agonist.
a   Full assessment: IOP and assessment of the optic nerve head, and visual fields.

No treatment → BB → PGA → PGA & BB → surgery (trabeculectomy) →
PGA → PGA & BB → second surgery (trabeculectomy) → PGA → PGA & BB

FIGURE 63 Treatment sequence assumed for the model.

TABLE 36 Treatment transitions for those with OHT

From	To			
	No treatment	BB	PGAs	PGAs and BB
No treatment	Yes	Yes	No	No
BB	Yes	Yes	Yes	No
PGAs	Yes	No	Yes	Yes
PGAs and BB	Yes	No	No	Yes

PGA, prostaglandin agonist.

Table 37 describes the possible changes in treatment for those who convert to OAG and *Table 38* outlines the possible sequence of treatments for those with OAG.

Description of the discrete event simulation economic model

As already noted a discrete event simulation model was used to compare the pathways outlined in *Table 33*. On entry each individual has a predefined risk of developing OAG. This risk is based on their characteristics (e.g. age, IOP, etc., as described in *Estimation of parameters for the natural history of ocular hypertension and open-angle glaucoma*) defined above and the risk prediction algorithm reported in detail in *Chapter 4*. The model then compares the consequences for an individual, depending on their risk characteristics, for each of the five surveillance pathways described in *Developing the surveillance care pathways*. As described in *Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway*, these

TABLE 37　Treatment transitions for those observed to convert to OAG from OHT

| From | To | | | |
	BB	PGA	PGA and BB	Surgery (trabeculectomy)
No treatment	No	Yes	No	No
BB	No	Yes	No	No
PGAs	No	No	No	Yes
PGAs and BB	No	No	No	Yes

PGA, prostaglandin agonist.

TABLE 38　Treatment transitions for individuals with OAG

| From | To | | | |
	No treatment	PGA	PGA and BB	Surgery (trabeculectomy)
PGAs	Yes	Yes	Yes	No
PGAs and BB	Yes	No	Yes	Yes
Surgery (trabeculectomy)[a]	Yes[b]	Yes[c]	No	No

PGA, prostaglandin agonist.
a　Individuals can have a second surgery if the first surgery was unsuccessful and they cannot be controlled with medication therapy (e.g. PGAs or PGAs and BB).
b　If surgery successful.
c　If surgery not successful.

surveillance pathways differ in terms of the process of surveillance, but following a correct diagnosis of OAG the progression of disease and the process of care are assumed to be identical. Thus, the model for these elements for the five surveillance pathways compared is also identical.

In discrete event simulation models a process or system is represented as a chronological sequence of events. The situation of an individual being modelled changes as a result of an event happening. For instance, in a model for the surveillance of individuals with OHT, the event can be a visit to the ophthalmologist where a particular set of tests is conducted. The change that might occur as a result of this event would be the diagnosis (or not) of OAG at that time.

Discrete event simulation models usually have five elements:

1. *Entities:* in this evaluation the entities are the individuals modelled within the discrete event simulation.
2. *Characteristics:* a set of characteristics for those individuals, which for this model are age and other risk predictors for OAG. These were defined based on the OHTS-EGPS risk prediction algorithm presented in *Chapter 4*.
3. *Events:* events occur at a particular point in time. The main events in this model are surveillance visits where the levels of defined characteristics are checked and the disease condition assessed. Other events include glaucoma surgery (e.g. trabeculectomy) after which individuals might enjoy a length of time with no progression without the need for further treatment.
4. *Outcomes:* these include mean cost, clinical outcomes (e.g. number of individuals converted to OAG), mean QALYs and mean WTP (as described in *Chapter 7*).

5. *Relationships:* mathematical or logical relationships that link the elements noted above together. An example of relationships in the model would be the way that the *characteristics* of an individual would link to the *outcome* of converting to OAG.

The important feature of a discrete event simulation is that an individual's pathway through the model can change only if an event occurs. The likelihood of these events occurring is governed by the characteristics of the individual (i.e. the entity) and the relationships we define for the disease process and hence pathways of care. As events occur at particular moments in time during the individual's pathway of care, the model includes a 'clock' to register an event.

Figure 64 illustrates how this process is operationalised to model the sequence of events involved in surveillance. This figure is based on the visual representation of the model in Simul8 version 17.0 (Simul8 Corporation, Boston, MA, USA), the software package used to construct the model.

In *Figure 64* we have marked five parts of the sequence of events involved as A, B, C, D and E. These points represent (A) the starting point, described in modelling terminology as the work entry point (WEP); (B) 'queues', where individuals wait until a further component in the model is free; (C) work centres (WC), where an event happens and some action is taken (e.g. a surveillance visit at which IOP measurement or perhaps a full assessment will take place); (D) end points, from which individuals cannot move elsewhere (i.e. death); and (E) the point at which, as a result of the findings of surveillance tests, IOP is judged to be off target or conversion to OAG is observed.

At the start of the model (marked as A in *Figure 64*) and within the WEP, the level for the characteristics (age, IOP, etc.) of each individual are defined, which determines the risk of progression to OAG. The value for each characteristic is based on data on the distribution of these characteristics in the population; we used the data from the Rotterdam study,[34] the details of which are described in *Chapter 3*. As described in *Table 34*, for each surveillance pathway a set of criteria are used to decide whether treatment for OHT is prescribed and, if so, which treatment. This is reflected in *Figure 64* by allocating individuals to no treatment, BBs or PGAs at the point at which they wait to start regular surveillance visits, illustrated by the queues marked as points B in the 'monitoring visit loop'). The queues are an important aspect of a discrete event simulation as they are used to capture the potentially variable length of time that might exist between events, which in the context of *Figure 64* are the surveillance visits.

When individuals enter the monitoring visit loop they go through a succession of surveillance visits (these are marked with a C in *Figure 64*). Before each visit the model re-estimates the level of IOP that the individual has at that time and whether or not they have truly converted to OAG. During each visit, IOP and disease stage are assessed and, based on the findings of the tests, any change in management is made (e.g. remain with current therapy and follow-up; change of therapy if IOP is 'off target'; or referral to OAG follow-up if conversion to OAG is detected). This is marked with an E.

Figure 65 provides a more detailed illustration of this sequence and shows the model schematic for an individual who is currently receiving BBs. If at a surveillance visit IOP is off target or conversion to OAG is detected then the individual is sent to a queue (identified by point F) where they await assignment to the next treatment in the sequence of treatments available to them (see *The sequence of treatment*).

If an individual converts to OAG then subsequent management will follow the NICE guidelines[1] for individuals with diagnosed OAG. The model structure describing the sequence of events following conversion is shown in *Figure 66* and is identical for all individuals who convert to
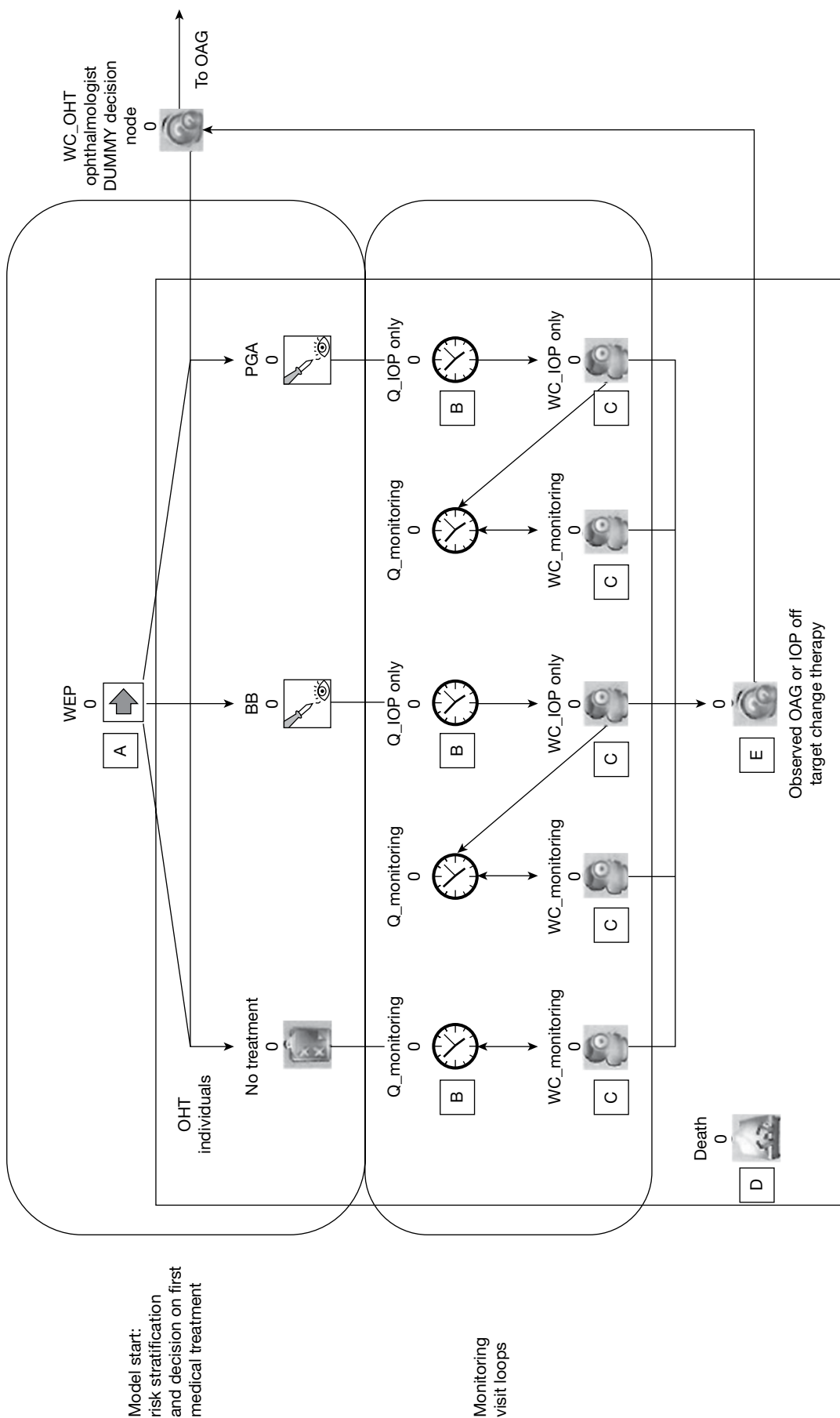
FIGURE 64 Model structure: OHT process. BB, beta-blockers; Q, queues (waiting for further action); WC, work centres; WEP, work entry point.
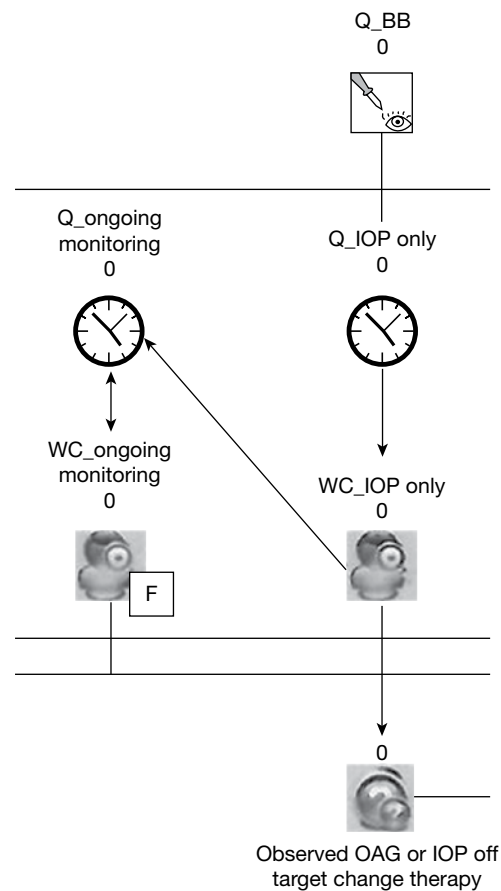
FIGURE 65 Surveillance visit loop. Q represents queues for specific surveillance visits defined as WC at which either IOP measurements only are taken (WC_IOP only) or there is a full assessment (WC_ongoing monitoring).

OAG regardless of the surveillance pathway that they were on. As described in *Table 34* the model assumed that all surveillance and care once an individual has converted to OAG would be provided by an ophthalmologist in a secondary care setting.

The first stage of the treatment sequence (not shown in *Figure 66*) is when the ophthalmologist decides what treatment should be given (see *Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway*). Within the model this occurs in the working centre 'WC_ophthalmologist decision node' (marked as G in *Figure 66*).

Within the model there are three possible therapeutic options: (1) medical management (PGAs only or in combination with BBs); (2) surgery; and (3) PGA only or in combination with BBs prescribed after failed surgery. Once an individual is prescribed a medical treatment, he or she will undergo a programme of surveillance that will continue until the individual exhausts the available medical treatment options (because medical treatments fail to keep IOP on target or the OAG continues to progress despite medical treatment and potentially escalation of medical treatment). This option is represented by the rectangle to the left-hand side of *Figure 66*.

Once medical treatment options have been exhausted then surgery will be considered as the second therapeutic option (represented by the rectangle in the centre of *Figure 66*). Surgery may also be ineffective at keeping IOP on target or preventing OAG progressing. In such a situation, individuals might return to medical therapy (e.g. PGAs only or in combination with BBs), the third therapeutic option. This option is represented by the rectangle on the right-hand side of *Figure 66*.
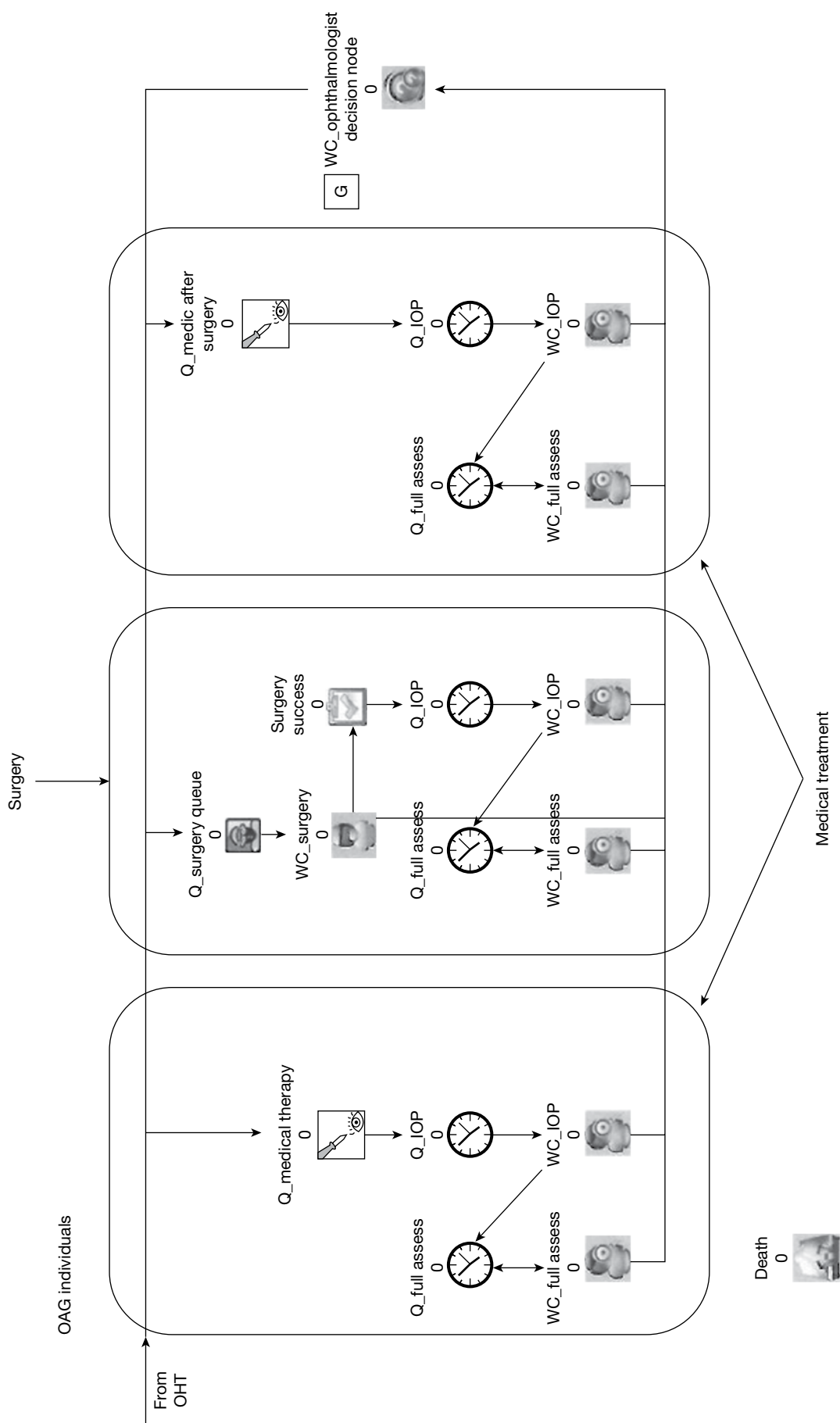
FIGURE 66 Model structure: surveillance and treatment sequence following conversion to OAG. assess, assessment. Q, queues (waiting for further action).

Estimation of parameters for the natural history of ocular hypertension and open-angle glaucoma

In this section we describe how we define several key sets of parameters used to populate the model structure described in *Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway*. The first set of parameters relates to how we simulated the characteristics of the individuals who enter into the model. The second and third sets of parameters relate to assessment of IOP and changes in OAG status. Within the model both IOP and OAG status are assessed as these determine changes in management. The final set of parameters relates to the natural history of OHT and OAG and how treatments may alter the natural history.

Baseline characteristics of the simulated population

As described above, the model starts with the simulation of relevant individual characteristics shown to be predictors of conversion to OAG (e.g. age, IOP, CCT, VCD ratio and a measure of visual field loss, PSD). The characteristics for each individual were drawn from probability distributions for these characteristics obtained from sources that were judged by the study team to be most appropriate for the population of interest. *Table 39* provides the point estimates of the cohort of individuals modelled as well as information used to define the distribution of the characteristic values in the population. The age- and sight-related characteristic values estimated for each individual were drawn from probability distributions fitted to the individual data from the Rotterdam study. Expected mortality rates for each individual were developed from UK interim life tables.[267]

Parameters relating to the modelling of the underlying level of intraocular pressure

An integral part of the model was that at surveillance visits IOP would be assessed. Therefore, an underlying level of IOP was needed. The level of IOP for an individual at any particular time was obtained by using the predictions of a mixed linear regression model, derived from the analysis described in *Chapter 6* and based on data from two RCTs, the Rotterdam study[34] and the Moorfields study,[32] with time as the dependent variable. The value of IOP at any particular time was given by:

$$\text{IOP}_i = (0.26 \times T_{\text{iop}i} + 1)^{1/0.26} \qquad \text{[Equation 15]}$$

TABLE 39 Model population: baseline characteristics

Description	Mean	Type of distribution and its parameters	Source/note
Age (years)	58.1	$N(58.1,109)$	Rotterdam data set[34] (see *Chapter 3*)
CCT (μm)	574.7	$N(574.7,1027)$	Rotterdam data set[34]
IOP (mmHg)	4.96[a]	$N(4.96,0.068)$	Rotterdam data set.[34] Note: normal distribution on transformed (log) IOP – see Equation 15. Mean = 24.19 mmHg
PSD (dB)	1.71	$N(1.708,0.11)$	Rotterdam data set[34]
VCD ratio	0.37	$N(0.37,0.03)$	Rotterdam data set[34]
Mortality	Various		UK interim life tables 2007–9 (males)[267]

a   This is the transformed IOP level.

where IOPi is the IOP (measured in mmHg) for individual i, and $T_{\text{iop}i}$ was given by:

$$T_{\text{iop}i} = \alpha + \beta_t \qquad\qquad \text{[Equation 16]}$$

where $\sim N(4.96, 0.068)$, β followed an empirical distribution (see *Chapter 6*) and t is time, measured in number of days from baseline.

Parameters used to model the conversion to open-angle glaucoma

The risk of conversion to OAG was developed using the risk prediction algorithm based on the pooled OHTS-EGPS model, described in *Chapter 4*, which is available as an online calculator (http://ohts.wustl.edu/risk/calculator.html), which provides an estimate of the 5-year risk of developing OAG. The risk of conversion in any subsequent 5-year period was estimated using the same risk prediction tool by entering into the calculator the level for each of the individual's characteristics at the start of the that 5-year period. Also included in *Table 40* are details of the type of distribution and information used to define the shape of the distribution.

Parameters used to model progression of open-angle glaucoma

As OAG is a chronic condition with a long natural history, the model also considered what would happen to an individual once he or she had converted from OHT to OAG. This required consideration of how OAG would progress over time. OAG was defined in terms of four levels of severity that differed according to the degree of visual impairment (and hence patient-reported health status and quality of life). The levels of severity considered were mild OAG, which individuals were assumed to have immediately following conversion from OHT, moderate OAG, severe OAG and visual impairment. These levels were consistent with those used in the DCE reported in *Chapter 7*. The risks of progression into moderate and severe OAG and eventually visual impairment were obtained from a systematic review conducted by Burr and colleagues.[268] Burr and colleagues estimated the number of years before visual field progression to the next level of OAG severity for untreated OAG. Based on these estimates, the model will delay the time of progression to the next level of severity depending on treatment efficacy and an individual's adherence to treatment. *Table 41* shows the average number of years to progression as well as the distribution used to sample the time to progression for each individual. Also, the lower part of this table reports the visual field-based staging system used.

Parameters used to modify natural history of ocular hypertension and open-angle glaucoma

Within the model there are two sets of parameters that have the effect of modifying the natural history of OHT and OAG. The first set relates to how accurate the measurement of IOP is in determining whether or not there is a clinically important change in IOP. This set of parameters determines whether or not the treatment of an individual should change. The second set of parameters relates to the effectiveness of treatments to control IOP and slow the rate of conversion to and progression of OAG.

TABLE 40 Parameters for risk of conversion to OAG

Description	Mean (years)	Type of distribution and its parameters	Source/note
5-year risk of conversion to OAG		Predictions from regression model on transformed data (see *Chapter 4*)	Calculated in the model based on literature (see *Chapter 4*)
Time to conversion	2.5	Uniform (0,5)	Assumption

TABLE 41 Open-angle glaucoma progression

Description	Mean (years)	Type of distribution and its parameters	Source/note
Time of progression from OAG conversion to moderate OAG	4	$N(4,2)$	Developed from Burr 2007.[268] Untreated individuals. Variance based on assumption
Time of progression from moderate OAG to severe OAG	9	$N(9,2)$	
Time of progression from severe OAG to visual impairment	10	$N(10,2)$	
Visual field-based OAG staging system			
Mild OAG		MD score (dB): −0.01 to −6.00	
Moderate OAG		MD score (dB): −6.01 to −12.00	
Severe OAG		MD score (dB): −12.01 to −20.00	
Visual impairment		MD score (dB): −20.01 to worse	

Parameters describing the precision of the measurement of intraocular pressure and progression of open-angle glaucoma

Within the model it is not assumed that IOP measurement, conversion to OAG or progression between the different severities of OAG will be known with certainty. The data on IOP were based on the Rotterdam and Moorfields data sets. These data sets collected IOP measurements made by ophthalmologists in secondary-care settings using GAT. Measurement of IOP by an ophthalmologist in the secondary-care setting and using GAT is generally considered to be the most reliable and accurate measure, but it is not perfect (see *Chapter 5*). To reflect this within the secondary care-based surveillance visits as well as for referrals to secondary care within the 'SOH primary care' and 'treat all' pathways we used the error term in the regression model reported in *Chapter 6* to estimate the uncertainty surrounding the measurement of IOP by ophthalmologists in a secondary-care setting.

Measurement of IOP by other staff in other settings and using other equipment might be less precise. To capture this further level of uncertainty for the 'SOH primary care' and 'treat all' pathways we used data obtained from the tonometry review reported in detail in *Chapter 5* to generate the uncertainty in the measurement of IOP when measurements were made by tonometers other than GAT.

With respect to the diagnosis of OAG, the assumption was made that when this was made in a secondary-care setting by an ophthalmologist the sensitivity and specificity were equal to 1 (i.e. the ophthalmologist in a secondary-care setting would always make a correct diagnosis). As already noted the model assumed that all surveillance and care once an individual has converted to OAG would be provided by an ophthalmologist in a secondary-care setting. It was therefore assumed that, if the individual progressed to a more severe stage of OAG, this would be identified by the ophthalmologist at the next visit of the individual.

Data on the ability of a non-ophthalmologist in a primary-care setting to detect conversion to OAG (e.g. within the 'SOH primary care' pathway) were taken from Azuara-Blanco and colleagues.[269] Azuara-Blanco and colleagues estimated the sensitivity and specificity of a diagnosis of OAG made from tests conducted by an optometrist (*Table 42*).

Parameters describing the clinical effectiveness of treatments

The model allowed for the effects of treatment on the level of IOP, the risk of conversion to OAG and the speed of progression of OAG once an individual had converted but the rules describing possible sequences of treatments, as described in *The sequence of treatment*, were the same.

TABLE 42 Measurement precision of IOP and accuracy of conversion to OAG

Description	Mean value	Source/note
Ophthalmologist		
IOP measurement error (mmHg)[a]	1	Based on regression model in *Chapter 6*
OAG diagnosis sensitivity	1	Assumption
OAG diagnosis specificity	1	Assumption
Community optometrist/primary-care setting professional		
IOP measurement error (mmHg)	0	Based on NCT (see *Appendix 4*, *Table 61*); sampled from N (0,5.76)
OAG diagnosis sensitivity	0.76	Azuara-Blanco 2007[269]
OAG diagnosis specificity	0.93	Azuara-Blanco 2007[269]

a   IOP (mmHg) will result from applying Equation 15 but using only the original regression model error term ($\varepsilon \sim N(0,0.036)$).

The type of treatment initiated varied by screening pathway, as described in *Description of the surveillance pathways compared and outline of the model*, and treatments for OHT were assumed to work solely by reducing IOP. If a treatment resulted in a $\geq 15\%$ reduction in IOP compared with IOP at the start of the model we assumed that IOP was on target (based on expert opinion and the model calibration) and no change in treatment was warranted. If the percentage reduction in IOP was $< 15\%$ then a change of treatment would occur in accordance with the decision rules set out in *Characterisation of the risk profile of the simulated cohort and surveillance and treatment criteria for each pathway*. To estimate whether treatment maintained IOP on target, the effect of treatment on IOP was entered into the model as percentage reduction in IOP level measured when the individual entered the model.

The percentage reductions in IOP for each treatment were developed from data reported in the NICE guidelines[1] and are shown in *Table 43*. The studies considered by the NICE guidelines had similar mean IOPs (i.e. between 23.1 mmHg and 25 mmHg) as the values estimated for the 10,000 individuals simulated in the model. However, the follow-up in the studies varied from 5 years for a comparison of BBs versus no treatment to 6 months for PGAs versus BBs and for combination therapy versus PGAs only.[1]

As described in *The sequence of treatment*, the possible sequences of treatments given to those who converted to OAG (see *Table 37*) and those for whom OAG had progressed (see *Table 38*) varied according to what previous treatment an individual had received, but the possible sequence of treatments was the same for each surveillance pathway (see *Figure 63*).

Thus, the treatment effect for those who had converted to OAG and those for whom OAG had progressed came from the same source, Maier and colleagues,[237] for all surveillance pathways (*Table 44*). The model assumed that this treatment effect would be enjoyed only by individuals who adhered to the treatment regimen and whose IOP was on target. Data on adherence to medical treatment were sparse. Therefore, based on expert opinion we assumed that individuals receiving treatment while undergoing surveillance would adhere to the treatment regimen only 75% of the time. The exception to this was the 'treat all' community IOP pathway in which everyone is prescribed PGAs and advised to have their IOP measured yearly by a community optometrist. For this pathway a lower rate of adherence (50%) was assumed as it was felt that there would not be the same level of reinforcement given to promote adherence.

Other parameter values required for the model were the proportion of individuals converting to OAG who will go through surgery (e.g. trabeculectomy) as well as the proportion of individuals

TABLE 43 Treatment effect on IOP for those with OHT

Description	Mean IOP reduction (%)	Source/note
BBs vs no treatment	12	Developed based on weighted mean differences in NICE guidelines[1]
PGAs vs BBs	5	Developed based on weighted mean differences in NICE guidelines[1]
PGAs + BBs vs PGAs only	1.5	Developed based on weighted mean differences in NICE guidelines[1]

TABLE 44 Treatment parameter values for those with glaucoma

Description	Mean value	Source/note
Reduction in risk of progression from any medical treatment (HR)		
Individuals with OHT	0.56	Maier 2005[237]
Individuals with OAG	0.65	Maier 2005[237]
Proportion of OAG individuals receiving surgery for OAG	8.50%	Burr 2009[270]
Proportion of successful surgeries	88% at 1 year and 79% at 4 years	Burr 2009.[270] Note: based on expert opinion the failure rates were assumed to be equal to the proportions of surgical patients subsequently in need of argon laser trabeculoplasty

with successful surgery. These data were obtained from a Cochrane review conducted by Burr and colleagues.[270] The surgical failure rate was assumed to be equal to the proportion of patients assessed to be in need of further treatment (e.g. argon laser trabeculoplasty) from the Collaborative Initial Glaucoma Treatment Study.[271]

Parameters used to describe the costs of surveillance, diagnosis and treatment

When surveillance was performed in a secondary-care setting (i.e. the NICE pathways and the SOH hospital eye service pathway, which were defined in *Description of the surveillance pathways compared and outline of the model*), we used the unit cost for an ophthalmology service outpatient visit from the Scotland National Statistics Information Services Division[272] to cost a visit to measure IOP only. This cost was doubled to reflect the cost of a full assessment (i.e. IOP measurement together with visual field measurement and assessment of the optic nerve) conducted in a secondary-care setting.

For the 'SOH primary care' pathway the unit cost for a non-ophthalmologist full assessment conducted in a primary care or community setting was assumed to be that of the NHS sight fee.[273] Likewise, we assumed that the cost of an appointment at which only an IOP measurement was taken would be half of that of a full assessment (this latter value was also used as the cost of the annual measurement of IOP by a community optometrist recommended as part of the 'treat all' pathway).

Table 45 reports the unit costs for full assessment and IOP measurement only visits.

Treatment costs

Medication costs were obtained from the *British National Formulary* (BNF)[274] (*Table 46*). When a BB was used, we assumed one bottle of eye drops per month and took the unit cost to be that of non-proprietary timolol. Because of the variation in the type of PGA used we selected the unit costs for Xalatan® (Pharmacia) and Xalacom® (Pharmacia) to calculate the annual cost for

TABLE 45 Costs of surveillance visits

Description	Mean cost (£)	Source/note
Secondary-care based		
IOP measurement only	90	Information Services Division Scotland 2010 (Table R044X Speciality group costs for consultant – ophthalmology – outpatients)[272]
Full assessment	180	Information Services Division Scotland 2010 (Table R044X Speciality group costs for consultant – ophthalmology – outpatients)[272]
Primary-care based		
IOP measurement only	10.35	Assumption
Full assessment	20.7	Department of Health (General Ophthalmic Services: NHS sight test fee)[273]

PGA and combination therapy. The cost for OAG surgery was obtained from *NHS reference costs 2008–9*[275] (e.g. HRG BZ18Z).

Health state utilities and benefit valuations
Health state utilities
The main objective of the economic evaluation was to inform decision-making in the UK. Current guidelines advocate the use of a quality-adjusted measure of health status utility obtained from the general population. Therefore, data were sought on health state valuation obtained using the EQ-5D valued using UK population tariffs,[276] or something similar (e.g. Short Form questionnaire-6 Dimensions).

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 OAG and visual impairment and these data were used in the model to value the time spent in these health states. The EQ-5D data obtained from 255 OAG individuals from Leeds and Aberdeen as well as members of the International Glaucoma Association were used.[277] As the degree of visual impairment with mild OAG is minimal, it was assumed that the score for those with OHT would be the same as the score for those with mild OAG. This value is comparable to the mean age group population EQ-5D score[276] that would be estimated for the model-simulated cohort age group. *Table 47* shows the utility weights used in the model.

Monetary values for cost–benefit analysis (willingness to pay)
As an alternative to estimating benefits in terms of QALYs derived from the EQ-5D scores, the WTP values obtained from the DCE reported in detail in *Chapter 7* were used. The WTP estimates per unit change of a specific outcome (see *Table 30* in *Chapter 7*) compared with a situation described in the DCE as 'no surveillance' were used as weights within the model.

A 'no surveillance' strategy implies that individuals could be opportunistically diagnosed with OAG and eventually treated. However, within the DCE (see *Chapter 7*) it was assumed that there were no unwanted effects of treatment in a 'no surveillance' strategy. This level of service was not considered a relevant option to formally compare within the economic model because of the absence of any form of surveillance, although it has been used as a baseline against which the total WTP for the other treatments was estimated and facilitated the indirect comparison between the pathways formally compared within the model. It was believed that the cost of opportunistic case finding should be close to that of the 'treat all' pathway as these two approaches assume that a minimum level of health care is provided. Therefore, we assumed a cost for a 'no surveillance' strategy equal to 80% of the cost of the 'treat all' pathway as an approximation of the costs of opportunistic case finding. Hence, in order to complete the

TABLE 46 Costs of medical and surgical treatments

Description	Mean cost (£)	Source/note
BB (£ per year of treatment)	18.72	BNF 2011,[274] based on non-proprietary timolol, 0.5%, 5 ml = £1.56. Assumed one bottle per month
PGAs (£ per year of treatment)	149.76	BNF 2011,[274] based on Xalatan, 2.5 ml = £12.48. Assumed one bottle per month
PGAs + BB (£ per year of treatment)	171.84	BNF 2011,[274] based on Xalacom, 2.5 ml = £14.32. Assumed one bottle per month
Surgery for OAG	1479	*NHS reference costs 2008–9*[275] (HRG BZ18Z)

TABLE 47 European Quality of Life-5 Dimensions utility weights used in the model

Description	Mean score	Source/note
Individuals with OHT	0.8015	Assumed equal to OAG mild individuals
Individuals with mild OAG	0.8015	Burr 2007[277]
Individuals with moderate OAG	0.7471	Burr 2007[277]
Individuals with severe OAG	0.7133	Burr 2007[277]
Visually impaired individuals with OAG	0.535	Burr 2007[277]

cost–benefit analysis results, the costs of each surveillance pathway compared within the model were reduced by a value equal to 80% of the cost of the 'treat all' pathway.

The DCE included seven attributes; six of these had four levels and one had two levels (see *Table 26* for details). Some levels defined in the DCE were merged to facilitate the incorporation of the DCE results into the economic evaluation model. The first situation where this occurred was when two levels within the unwanted treatment effects (i.e. 'some' and 'quite a lot') were collapsed into a single level. The rationale underpinning this was that the data obtained from the literature on the proportion of individuals suffering unwanted effects of treatment did not distinguish between 'some' unwanted treatment effects and 'quite a lot' of unwanted treatment effects. A second situation where this occurred was when two levels within the communication and understanding attribute ('felt at ease but did not understand' and 'did not feel at ease but understood') were collapsed into one level. This decision was made to simplify the analysis and also because there was little evidence to suggest that preferences for these two levels were different (this is illustrated by the overlap in the CIs surrounding the WTP values for these two levels – see *Chapter 7*, *Table 30*). *Table 48* shows the marginal WTP for the considered attributes together with the 95% CIs. The 95% CI data were used to produce distributions around the point estimates of marginal WTP. It was assumed that a normal distribution would be most appropriate to define the variance surrounding the mean value as distributions of this form are assumed by the econometric methods used to analyse the DCE data.

The WTP values in *Table 48* were obtained by asking how much individuals in the DCE survey would be willing to pay per year for a surveillance programme with alternative sets of characteristics. Therefore, these values are WTP amounts per year that the individuals are under surveillance.

To estimate the mean WTP to avoid unwanted effects of treatments for each surveillance pathway, the proportion of the 10,000 cohort who developed unwanted effects of treatment (obtained from the literature) was combined with the WTP to avoid this event. It was assumed

TABLE 48 Marginal WTP for alternative surveillance programme attributes

Attribute	Marginal WTP (£)	95% CI
Alternative specific constant (reference: no surveillance)	28.31	19.61 to 36.35
10-year risk of		
Conversion to OAG of individuals with OHT	−0.03	−0.04 to −0.03
Progressing to severe OAG for OHT individuals	−0.06	−0.09 to −0.04
Becoming visually impaired for OHT individuals	−0.50	−0.74 to −0.29
Unwanted treatment effects (reference: none):		
Some or quite a lot	−24.40	−28.61 to −21.00
Severe	−60.05	−69.19 to −52.17
Communication and understanding (reference: did not feel at ease and did not understand)		
Felt at ease and understood	46.42	39.95 to 53.71
Felt at ease but did not understand *or* did not feel at ease but understood	26.87	23.16 to 31.24
Hospital setting (reference: community)	1.05	−0.90 to 3.10

that the unwanted effects of treatment would last for the period of time between the last and the current surveillance visits.

We used the reference category level ('did not feel at ease and did not understand') for the communication and understanding attribute to define the level of this attribute associated with the NICE guidelines-based pathways. This was based on the assumption that the hospital eye service does not communicate well with patients (because of, for instance, lack of time given that they are working in excess of their capacity). Similarly, we used 'felt at ease but did not understand' *or* 'did not feel at ease but understood' to define the level of communication and understanding associated with the SOH and 'treat all' pathways. Finally, as each surveillance strategy took place in a predefined setting (e.g. the community or in hospital), the mean WTP for each pathway with respect to location of care was either the positive (+£1.05 for surveillance in hospital) or the negative (–£1.05 for surveillance in the community) of the mean value reported in *Table 48* for a unit change in this attribute.

In the DCE reference case no surveillance and no treatment were assumed unless cases were found opportunistically. From the model the estimated number of individuals who converted to OAG in 20 years and the numbers who progressed to severe OAG and visual impairment were obtained. The differences between the model estimates and the numbers of cases in the DCE reference case (e.g. 3200, 360 and 50 for the numbers of individuals who converted to OAG and progressed to severe OAG and visual impairment, respectively) were calculated. These were multiplied by the cumulative WTP for the number of years under surveillance (e.g. assumed to be 10 years on average) to obtain the total benefits resulting from the numbers of avoided cases of OAG, severe OAG and visual impairment. Furthermore, for final net benefit calculations we assumed, as described above, that the average cost for the DCE reference case was equal to 80% of the cost for the 'treat all' pathway. This was deducted from the average cost for each pathway (see *Table 52*).

Model calibration and validation

In the development of any simulation model, care should be taken throughout the process so that credible and eventually useful results are produced. Parts of the development process are the

verification, validation and eventually calibration of the model. Verification of the model refers to the steps taken to ensure that the model of an individual's pathway, as well as all mathematical calculations that govern the journey through the model, are correct. Model validation refers to the actions taken to verify that the model fits with empirical data. If the model predictions do not appear to tally with the available empirical data, calibration is needed (i.e. change of model parameter values in order to correct these discrepancies).

To calibrate the model we first simulated individual characteristics for a population of 1000 people at the point at which they enter into the model. The data produced were then presented to the project management group and clinical experts and interpreted alongside data from the literature to form a judgement as to whether they had face validity. We did this by comparing the model simulated values with those of the Rotterdam data set (see *Table 5*).

Data analysis

The model compared alternative surveillance regimes by simulating for each pathway the cost per individual, clinical outcomes (e.g. number of individuals converted to OAG, number of individuals who progressed to moderate OAG, severe OAG and visual impairment), mean QALYs (e.g. EQ-5D-based QALYs) and mean WTP (e.g. monetary measure of benefits) for 10,000 individuals with OHT. A cohort of this size was chosen to enable the variation that would be expected between individuals to be represented in the estimates of costs and outcomes. Within the base-case analysis we adopted an NHS perspective and discounted costs and QALYs at the recommended 3.5% discount rate. WTP estimates were not discounted in the base-case analysis as these data implicitly included the respondents' rates of time preference as respondents to the DCE were asked to consider trade-offs between short- and long-term outcomes. The model itself estimated costs and outcomes over a 20-year time horizon and all monetary values are expressed in 2009/10 prices.

From the data produced by the model for each surveillance pathway the following comparative results between surveillance pathways were estimated:

- incremental cost per number of OAG cases avoided, and incremental cost per case of each level of severity of OAG avoided
- incremental cost per QALY gained and
- incremental net benefit.

We addressed uncertainty by conducting deterministic (e.g. one-way) sensitivity analyses. The following sensitivity analyses were considered.

Changes to the 5-year risk of conversion

Based on a 13-year follow-up of the participants in the OHTS study,[61] the cumulative proportion of participants developing OAG, categorised into tertiles according to their baseline risk,[64] was < 6%, 6–13% and > 13%, respectively.[13] The benefit of treatment was most evident among participants at a higher baseline risk. We thus defined 6% as the threshold at which to start treatment for the base-case analysis. We increased this threshold to 40% to explore the effects of treating only a very high-risk group.

Within the base case individuals would enter the model if they were diagnosed with OHT (defined as an IOP level > 21 mmHg). In a further sensitivity analysis, individuals would enter the model if their 5-year risk of conversion was equal to or above a stated threshold (e.g. from 6% up to 50%). The intuition behind this was that, for individuals with higher levels of risk, monitoring pathways could become cost-effective compared with a minimal level of care 'treat all' pathway or a no formal monitoring pathway.

Changes to the unit price of prostaglandin analogue

In the near future the patents for the PGAs will expire, which will allow the entry of manufacturers of generic versions of the drug to enter the market. This will drive the price of PGAs down, although at this stage it is unclear by how much. To address this uncertainty we ran the model (1) when the unit price of the PGA was 50% of the value used in the base-case analysis and (2) making the unit price equal to the unit price of a BB used in the base-case analysis.

Changes to the unit price of monitoring visits

Follow-up visits are potentially the main resource use within this model. Further sensitivity analyses were undertaken using the *NHS reference costs* data for 2009–10[275] as a way of exploring alternative values for unit costs for follow-up visits. Unit costs for consultant-led and non-consultant-led follow-up attendance were used for full assessment and IOP measurement only visits, respectively (code 130). Three sensitivity analyses were run using these national average figures. These were:

1. using £73 and £51 for full assessment and IOP measurement only visits, respectively, as opposed to £180 and £90 in the base-case analysis
2. using the lower estimated value from the IQR of £61 for full assessment and £32 for IOP measurement only visits
3. using the upper estimated value from the IQR of £83 for full assessment and £68 for IOP measurement only visits.

It should be noted that the base-case analysis assumption of a longer time needed to perform a full assessment (i.e. twice that of a normal visit) was dropped for these sensitivity analyses. We tested this assumption by conducting an additional one-way sensitivity analysis on the unit cost of IOP measurement only visits. This involved maintaining the unit cost of full assessment at £180 (the same as for the base-case analysis) but varying the cost of IOP measurement only visits within the range of £51–68 per visit.

Scenario analysis

Given the base-case analysis results we actively identified specific situations in which conclusions might change. Therefore, we looked for those variable values that would favour monitoring, particularly monitoring with the 'SOH hospital' pathway, which appeared to have the greatest possibility of being cost-effective compared with a 'treat all' pathway in terms of average costs and QALYs. The values for some variables could eventually favour the 'treat all' pathway much more so than the monitoring pathways. For instance, a very much higher rate of conversion and progression would eventually favour a 'treat all' pathway; more individuals would benefit from the PGAs that are initially prescribed to everyone with this pathway. However, we identified a group of variables that unequivocally favour 'SOH hospital' monitoring. These were adherence with treatment (higher for 'SOH hospital', lower for 'treat all' compared with the base case), higher precision of IOP measurement with GAT, lower precision for non-GAT IOP measurements in the community and lower accuracy for testing for OAG in the community. *Table 49* reports the figures used in this scenario analysis. The analysis is reported for alternative values of risk at which these individuals would initiate treatment within the SOH pathways.

The model was computationally too demanding (e.g. around 10 minutes for 10,000 individual runs). Consistency between 10,000 and 1000 individual runs was high for the base-case cost–utility analysis results. Therefore, all sensitivity analyses were conducted for 1000 ocular hypertensive individuals and reported on cost–utility analysis results only.

TABLE 49 Scenario analysis: extreme plausible values

Variable	Base case	Scenario analysis	Source
Time of progression from OAG conversion to OAG moderate	4	2	Assumption
Time of progression from OAG moderate to OAG severe	9	4.5	Assumption
Time of progression from OAG severe to visual impairment	10	5	Assumption
Test performance			
Community OAG diagnosis sensitivity	0.76	0.32	Azuara-Blanco 2007[269]
Community OAG diagnosis specificity	0.93	0.96	Azuara-Blanco 2007[269]
Proportion of individuals adhering to treatment			
NICE and SOH monitoring pathways	0.75	0.95	
'Treat all' pathway	0.5	0.2	

Further one-way sensitivity analyses

To further explore results from the scenario analysis, further one-way sensitivity analyses were conducted, particularly on the adherence to treatment rates for non-monitoring and monitoring pathways. Starting from the base-case analysis, we conducted a threshold analysis to identify values for the adherence rates which could result in estimates of the incremental cost per QALY that society might find acceptable.

Results

Model calibration

The simulated individual characteristics for a population of 10,000 ocular hypertensive individuals at model entry are described in *Table 50*.

We consulted the project management group and clinical experts about whether or not these data had face validity. It was agreed that, in comparison with available data detailed in *Table 5* (the Rotterdam and Moorfields data sets), these estimates were appropriate; in *Table 5* the mean age ranged from 56 to 64 years, while IOP ranged from 21 to 25 mmHg for those without OAG. Similar assessments can be drawn for CCT, VCD ratio and PSD. Finally, the predicted average 5-year risk of conversion for the (untreated) simulated individual group was 10%.

Estimates of effectiveness and mean cost per pathway

Table 51 shows the discounted cost and effectiveness analysis results. Effectiveness of SOH was expressed in several different ways: the number of individuals who converted to OHT; the number of individuals who progressed to moderate or severe OAG; and the number who progressed to visual impairment. Although not a prespecified form of analysis, the presentation of cost and clinical effect data for each pathway is equivalent to presenting the analysis as a cost–consequence analysis.

Of the 10,000 OHT individuals entered into each surveillance pathway, 2073 individuals were estimated to have converted to OAG over the 20-year time horizon of the model for the 'SOH hospital' pathway, while 2060 and 2079 converted for the 'NICE intensive' and 'NICE conservative' pathways, respectively. For the 'SOH primary care' pathway 2114 individuals converted to OAG and for the 'treat all' pathway 2282 individuals converted.

In terms of progression to moderate OAG, the worst-performing pathway was the 'treat all' pathway (1708 people progressed to moderate OAG), whereas the best performing was the 'NICE intensive' pathway (1507 progressed). A total of 336 individuals in the 'NICE intensive' pathway

TABLE 50 Descriptive statistics for simulated individual characteristics at the start of the model: ocular hypertensive individuals (*n* = 10,000)

Variable	Age (years)	IOP at start (mmHg)	Untreated IOP at year 5 (mmHg)	CCT (µm)	PSD (dB)	VCD ratio	5-year risk of conversion (units)
Mean (SD)	57 (9.8)	24.9 (2.4)	25.1 (5.4)	574.5 (31.7)	1.8 (0.3)	0.4 (0.2)	0.1 (0.1)
Median (IQR)	57.1 (13.2)	24.6 (3.3)	24.6 (6.4)	574.2 (42.3)	1.8 (0.4)	0.4 (0.2)	0.1 (0.1)

TABLE 51 Cost-effectiveness analysis results (discounted)

Pathway	Average total cost (£)	Incremental cost (£)	Number of individuals — Converted to OAG	Number of individuals — Progressed to moderate OAG	Number of individuals — Progressed to severe OAG	ICER (£) — Converted to OAG	ICER (£) — Progressed to moderate OAG	ICER (£) — Progressed to severe OAG
Treat all	3393		2282	1708	468			
SOH hospital	3956	562	2073	1536	343	3	3	4
SOH primary care	4696	740	2114	1545	365	Dominated	Dominated	Dominated
NICE conservative	5087	1131	2079	1528	349	Dominated	141	Dominated
NICE intensive	6862	1776	2060	1507	336	224	85	415

ICER, incremental cost-effectiveness ratio.
Dominated = more costly and less effective than another alternative.

progressed to severe OAG. The corresponding figures were 343 for the 'SOH hospital', 365 for the 'SOH primary care', 349 for the 'NICE conservative' and 468 for the 'treat all' pathways. No individuals progressed to visual impairment under any of the pathways, primarily because of the relatively short (20-year) time horizon of the model and the relatively long lead time it would take an individual to develop disease of this level of severity (on average 9 years for progression from moderate to severe OAG and 10 years for progression from severe OAG to visual impairment respectively). The pathway with the lowest total average cost was the 'treat all' pathway (£3393), and the most costly pathway was the 'NICE intensive' pathway (£6862).

Incremental cost-effectiveness

Table 51 details the incremental cost-effectiveness ratios (ICERs) for the cost–consequences outcomes. As the data presented in this table illustrate, a judgement about which pathway is most cost-effective is dependent on which single measure of effectiveness (e.g. progression, cases of moderate or severe OAG) is believed most important as well as how much society would be willing to pay for that particular extra unit of effectiveness. Overall, the 'SOH primary care' pathway is dominated by the 'SOH hospital' pathway. Moreover, the 'NICE conservative' pathway is dominated if cases of conversion to OAG or progression to severe OAG are used as measures of effectiveness. The decision on which pathway to adopt depends on society's WTP to avoid a case of OAG, moderate OAG or severe OAG. If society were willing to pay > £3 (and < £224) for an additional case of OAG avoided, then the 'SOH hospital' pathway should be adopted. Should avoidance of moderate or severe OAG be the sole measure of effectiveness relevant to a decision-maker then the 'SOH hospital' pathway would be considered cost-effective should society be willing to pay > £3 or > £4 per additional case avoided of moderate and severe OAG, respectively. However, the 'NICE intensive' pathway might be selected if WTP were > £224, > £85 or > £415 per additional case avoided of OAG or moderate or severe OAG, respectively.

Cost–utility analysis

The advantage of cost–utility analysis over cost-effectiveness analysis is that multiple dimensions of clinical effectiveness can be combined into a single unitary measure, the QALY. *Table 52* shows the cost–utility results when both costs and QALYs are discounted at 3.5%. In the cost–utility analysis the costs are identical to those reported in *Table 51*. Given the modest number of individuals out of the cohort of 10,000 OHT individuals who progress to any stage of OAG the differences in QALYs are relatively small. Both NICE pathways as well as the 'SOH primary care' pathway are dominated (are more costly and produce less QALYs) by the 'SOH hospital' pathway. The 'SOH hospital' pathway is more costly and provides more QALYs than the 'treat all' pathway; however, the ICER is well above the usual threshold of WTP for a QALY adopted by NICE (i.e. £30,000).[21]

Cost–benefit analysis

Table 53 shows the results of the cost–benefit analysis. As noted above a hypothetical reference pathway was defined (which in the DCE was called 'no surveillance'). This reference case is relevant as estimates of cost and WTP derived from the DCE are relative to this reference case. For the hypothetical reference case 3200 individuals out of a cohort of 10,000 would develop OAG, while another 360 and 50 out of a cohort of 10,000 would progress to severe OAG and visual impairment, respectively, over a time horizon of 20 years (the time horizon used in the DCE – see *Chapter 7*). Differences between the numbers of cases within the model pathways and the reference case were used when calculating average total benefits.

The SOH pathways provided on average £600 more benefits than the NICE-based pathways. This can be explained by the added value generated by the better communication and understanding assumed within the SOH pathways.

The average total benefits for every pathway except the 'SOH hospital' pathway were less than the average costs. Hence, the net benefits for all of the pathways except for the 'SOH hospital'

TABLE 52 Cost–utility analysis results (discounted)

Pathway	Average total cost (£)	Incremental cost (£)	Average total QALYs	Incremental QALYs	ICER (£)
Treat all	3393		9.7866		
SOH hospital	3956	562	9.7932	0.0066	85,312
SOH primary care	4696	740	9.7920	Dominated	Dominated
NICE conservative	5087	391	9.7923	Dominated	Dominated
NICE intensive	6862	1776	9.7931	Dominated	Dominated

Dominated = more costly and less effective than another alternative.

TABLE 53 Cost–benefit analysis results (discounted)

Pathway	Average total cost (£)	Average total benefits (£)	Average net benefit (£)
Treat all	679	564	−114
SOH hospital	1241	1693	452
SOH primary care	1981	1484	−498
NICE conservative	2372	1051	−1321
NICE intensive	4148	1099	−3049

pathway were negative. This indicates that compared with a 'no surveillance pathway' these surveillance pathways would not be regarded as being worthwhile. The only pathway that has a positive net benefit was the 'SOH hospital' pathway, indicating that this pathway would be considered worthwhile compared with the reference case. The main driver of these results is the very low valuation put on prevention of OAG and subsequent progression coupled with the low rates of OAG observed.

Sensitivity analyses

We ran a number of sensitivity analyses. For all of these analyses the model would report outcomes in clinical (e.g. number of cases of OAG), utility (e.g. QALYs) and monetary (e.g. British pounds) units; however, for the sake of simplicity we focus only on the cost–utility analysis results.

Table 54 reports the results for the one-way sensitivity analysis for the risk threshold at which the decision to initiate treatment would be made within the SOH pathways. This threshold value was 6% for the base-case analysis (i.e. we would treat only if the risk of converting over 5 years was ≥ 6%). *Table 53* shows the results for the cost–utility analysis when the risk threshold that triggered treatment was increased to 50%. 'Treat all' was the pathway with the lowest mean cost when the risk threshold values were ≤ 10%. When the risk threshold was between 10% and 15% the 'SOH hospital' pathway becomes the pathway with the lowest average cost and because it provides greater QALYs it dominates the 'treat all' pathway. However, over the range of risk thresholds considered none of the other pathways had an incremental cost per QALY that approached £30,000.[21]

In the base-case analysis individuals would enter the model if they had OHT (i.e. IOP >21mmHg.). We ran a sensitivity analysis in which individuals would enter the model if they were ocular hypertensive but also if their 5-year risk of conversion was higher than a particular threshold (e.g. from 6% to 50%). This sensitivity analysis showed similar results to those reported in the base-case analysis (data not shown). In these analyses, SOH pathways dominated NICE-based pathways but, although the SOH pathways were more effective than the 'treat all' pathway, they were also more costly and the ICERs were well above the usual threshold considered by NICE.[21]

Sensitivity analysis was also conducted on the unit price of PGAs and combination therapy (data not shown). As expected, as only unit price changed, no effect was observed on QALYs for any of the pathways. However, average costs for all pathways fell but proportionally more in those that used PGAs more intensively (e.g. the 'treat all' pathway). Therefore, none of the more costly pathways was associated with an incremental cost per QALY approaching £30,000.[21]

Further sensitivity analysis was conducted for the unit price of follow-up visits. For the base-case analysis it was assumed that a full assessment visit would take twice the time of a normal visit. This assumption was dropped for the sensitivity analysis and this partially explains the lower unit costs used. When using the national average costs (i.e. unit cost of £73 and £51 for full assessment and IOP measurement only visit, respectively) the 'treat all' pathway was the lowest cost pathway followed by the 'SOH hospital' pathway (*Table 55*). The ICER in this case was £10,857 while all other pathways were dominated by the 'SOH hospital' pathway. Using the lower value from the IQR from the NHS reference costs (i.e. £61 and £32 for full assessment and IOP measurement only visits, respectively) resulted in the 'SOH hospital' pathway dominating all others (data not shown). Finally, when using the upper value from the IQR (i.e. £83 and £68, respectively) the 'treat all' pathway had the lowest average cost followed by the more effective 'SOH hospital'

TABLE 54 Treatment risk threshold: one-way sensitivity analysis

Treatment risk threshold	Pathway	Average total cost (discounted) (£)	Average total QALYs (discounted)	ICER (£)
6%	Treat all	3412	9.7757	
	SOH hospital	3974	9.7828	78,911
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated
8%	Treat all	3412	9.7757	
	SOH hospital	3811	9.7813	71,704
	SOH primary care	4401	9.7808	Dominated
	NICE conservative	5156	9.7813	21,722,687
	NICE intensive	6879	9.7828	1,175,663
10%	Treat all	3412	9.7757	
	SOH hospital	3631	9.7812	39,701
	SOH primary care	4379	9.7807	Dominated
	NICE conservative	5156	9.7813	15,022,678
	NICE intensive	6879	9.7828	1,175,663
15%	SOH hospital	3311	9.7805	
	Treat all	3412	9.7757	Dominated
	SOH primary care	4118	9.7802	Dominated
	NICE conservative	5156	9.7813	2,213,752
	NICE intensive	6879	9.7828	1,175,663
20%	SOH hospital	3183	9.7805	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3956	9.7803	Dominated
	NICE conservative	5156	9.7813	2,445,687
	NICE intensive	6879	9.7828	1,175,663
25%	SOH hospital	3158	9.7793	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3888	9.7795	3,565,397
	NICE conservative	5156	9.7813	677,141
	NICE intensive	6879	9.7828	1,175,663
30%	SOH hospital	3122	9.7789	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3885	9.7799	753,380
	NICE conservative	5156	9.7813	890,088
	NICE intensive	6879	9.7828	1,175,663
35%	SOH hospital	3117	9.7782	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3915	9.7795	606,993
	NICE conservative	5156	9.7813	693,262
	NICE intensive	6879	9.7828	1,175,663
40%	SOH hospital	3063	9.7793	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3988	9.7791	Dominated
	NICE conservative	5156	9.7813	1,045,106
	NICE intensive	6879	9.7828	1,175,663
50%	SOH hospital	3071	9.7787	
	Treat all	3412	9.7757	Dominated
	SOH primary care	3954	9.7797	871,101
	NICE conservative	5156	9.7813	746,284
	NICE intensive	6879	9.7828	1,175,663

pathway (ICER for 'SOH hospital' pathway vs 'treat all' pathway was £40,332 – data not shown). All other pathways were dominated by the 'SOH hospital' pathway.

One-way sensitivity analysis shows that the results were not sensitive to the unit price of full assessment visits but were to the unit price of IOP measurement only visits. *Table 56* presents the cost–utility analysis results when the unit cost for full assessment visits was maintained at £180 as for the base-case analysis, but alternative values were used for the unit cost for hospital-based IOP measurement only visits. The £180 cost for the full assessment visit assumed that the time for a full assessment was twice that needed for a normal ophthalmological follow-up visit. Comparison between *Tables 55* and *56* show that increasing the unit cost from £73 to £180 has a small impact on the incremental cost per QALY between the 'SOH hospital' and the 'treat all' pathways (i.e. it increases from £10,857 to £11,410). However, varying the unit cost for IOP measurement only visits has a larger impact on the results, namely, the incremental cost per QALY between the 'SOH hospital' and the 'treat all' pathways increases to >£30,000 when the unit price for an IOP measurement only visit is between £61 and £62. The incremental cost per QALY reaches £40,834 when the unit cost for a hospital-based IOP measurement only visit is £68.

Table 57 gives details of the scenario analysis results for alternative values of 5-year risk threshold at which individuals initiate treatment within the SOH pathways. As this risk threshold is used only within the SOH pathways, the values for mean cost and mean QALYs for the other pathways do not change. Furthermore, the higher the risk threshold, the lower the mean cost and QALYs associated with the SOH pathways. The NICE-based pathways either are dominated or have an incremental cost per QALY well above the £30,000 threshold.[21]

For the stated variable values (see *Table 49*) and a 6% risk threshold (i.e. base-case level) the 'SOH hospital' pathway has the lowest mean cost followed by the 'treat all' pathway. The next most costly pathways are 'SOH primary care', 'NICE conservative' and 'NICE intensive'. The 'SOH hospital' pathway would give additional QALYs at a slightly lower cost and therefore dominates the 'treat all' pathway. Moreover, at a 10% risk threshold 'SOH primary care' dominates the 'treat all' pathway. However, even at a risk threshold of up to 50% the incremental cost per QALY for the comparison between the 'SOH primary care' and 'SOH hospital' pathways is well above £30,000.[21]

Tables 58 and *59* report the findings from the one-way sensitivity analysis on the adherence rate for the 'treat all' and monitoring pathways, respectively. For the base-case analysis, adherence to treatment rates was assumed to be 50% for the 'treat all' pathway and 75% for the NICE or SOH monitoring pathways. When adherence to treatment decreases for the 'treat all' pathway, the cost-effectiveness of the 'SOH hospital' pathway improves. For adherence rates <40% the incremental cost per QALY for the SOH pathway compared with the 'treat all' approaches £30,000. The incremental cost per QALY falls below £30,000 when the adherence rate drops towards 30%.

Similarly, when the adherence rate to treatment within the monitoring pathways rises, the cost-effectiveness of these pathways improves because the average cost for each falls and the QALYs increase. At an adherence rate to treatment just below 90% the incremental cost per QALY for the 'SOH hospital' pathway compared with the 'treat all' pathway will be approximately £30,000.

In summary, differences in adherence rates between the 'treat all' and the SOH pathways of approximately ≥40% will result in the SOH pathway having an incremental cost per QALY of ≤£30,000.[21]

TABLE 55 Sensitivity analysis on costs: cost–utility analysis results (discounted)

Strategy	Average total cost (£)	Incremental cost (£)	Average total QALYs	Incremental QALYs	ICER (£)
Treat all	2736		9.7757		
SOH hospital	2814	77	9.7828	0.0071	10,857
NICE conservative	3120	306	9.7813	Dominated	Dominated
SOH primary care	3180	60	9.7826	Dominated	Dominated
NICE intensive	3861	680	9.7828	Dominated	Dominated

Unit costs: full assessment visits = £73; IOP only visits = £51.
Dominated = more costly and less effective than another alternative.

TABLE 56 One-way sensitivity analysis on unit cost of IOP measurement only visits: cost–utility analysis results (discounted)

Unit cost of IOP measurement only visits (£)	Strategy	Average total cost (£)	Incremental cost (£)	Average total QALYs	Incremental QALYs	ICER (£)
51	Treat all	3188		9.7757		
	SOH hospital	3270	81	9.7828	0.0071	11,410
	SOH primary care	3856	586	9.7826	Dominated	Dominated
	NICE conservative	4381	526	9.7813	Dominated	Dominated
	NICE intensive	6075	1694	9.7828	Dominated	Dominated
55	Treat all	3211		9.7757		
	SOH hospital	3342	131	9.7828	0.0071	18,333
	SOH primary care	3927	586	9.7826	Dominated	dominated
	NICE conservative	4461	533	9.7813	Dominated	Dominated
	NICE intensive	6158	1697	9.7828	Dominated	Dominated
60	Treat all	3240		9.7757		
	SOH hospital	3432	192	9.7828	0.0071	26,987
	SOH primary care	4017	585	9.7826	Dominated	Dominated
	NICE conservative	4560	543	9.7813	Dominated	Dominated
	NICE intensive	6261	1701	9.7828	Dominated	Dominated
61	Treat all	3246		9.7757		
	SOH hospital	3450	205	9.7828	0.0071	28,718
	SOH primary care	4035	585	9.7826	Dominated	Dominated
	NICE conservative	4580	545	9.7813	Dominated	Dominated
	NICE intensive	6281	1701	9.7828	Dominated	Dominated
62	Treat all	3251		9.7757		
	SOH hospital	3468	217	9.7828	0.0071	30,449
	SOH primary care	4053	584	9.7826	Dominated	Dominated
	NICE conservative	4600	547	9.7813	Dominated	Dominated
	NICE intensive	6302	1702	9.7828	Dominated	Dominated
65	Treat all	3269		9.7757		
	SOH hospital	3523	254	9.7828	0.0071	35,641
	SOH primary care	4107	584	9.7826	Dominated	Dominated
	NICE conservative	4659	553	9.7813	Dominated	Dominated
	NICE intensive	6364	1704	9.7828	Dominated	Dominated
68	Treat all	3286		9.7757		
	SOH hospital	3577	291	9.7828	0.0071	40,834
	SOH primary care	4160	584	9.7826	Dominated	Dominated
	NICE conservative	4719	559	9.7813	Dominated	Dominated
	NICE intensive	6426	1706	9.7828	Dominated	Dominated

TABLE 57 Scenario analysis: extreme plausible values favouring monitoring

Treatment risk threshold	Pathway	Average total cost (discounted) (£)	Average total QALYs (discounted)	ICER (£)
6%	SOH hospital	3556	9.7671	
	Treat all	3642	9.7280	Dominated
	SOH primary care	3774	9.7634	Dominated
	NICE conservative	4691	9.7658	Dominated
	NICE intensive	6506	9.7681	2,844,867
8%	SOH hospital	3390	9.7651	
	Treat all	3642	9.7280	Dominated
	SOH primary care	3704	9.7659	420,454
	NICE conservative	4691	9.7658	Dominated
	NICE intensive	6506	9.7681	1,241,806
10%	SOH hospital	3227	9.7631	
	SOH primary care	3583	9.7649	207,454
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	1,151,063
	NICE intensive	6506	9.7681	794,299
15%	SOH hospital	2924	9.7605	
	SOH primary care	3293	9.7638	114,736
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	678,076
	NICE intensive	6506	9.7681	794,299
20%	SOH hospital	2772	9.7595	
	SOH primary care	3096	9.7637	77,533
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	756,147
	NICE intensive	6506	9.7681	794,299
25%	SOH hospital	2720	9.7578	
	SOH primary care	3010	9.7622	66,109
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	458,465
	NICE intensive	6506	9.7681	794,299
30%	SOH hospital	2712	9.7569	
	SOH primary care	3047	9.7608	84,955
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	329,129
	NICE intensive	6506	9.7681	794,299
35%	SOH hospital	2669	9.7574	
	SOH primary care	3104	9.7612	113,636
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	344,544
	NICE intensive	6506	9.7681	794,299
40%	SOH hospital	2667	9.7568	
	SOH primary care	3071	9.7610	97,413
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	333,038
	NICE intensive	6506	9.7681	794,299
50%	SOH hospital	2662	9.7561	
	SOH primary care	3026	9.7606	81,022
	Treat all	3642	9.7280	Dominated
	NICE conservative	4691	9.7658	319,080
	NICE intensive	6506	9.7681	794,299

TABLE 58 One-way sensitivity analysis on the adherence rate for the non-monitoring pathway ('treat all')

Adherence rate	Pathway	Average total cost (discounted) (£)	Average total QALYs (discounted)	ICER (£)
50%	Treat all	3412	9.7757	
	SOH hospital	3974	9.7828	78,911
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated
40%	Treat all	3526	9.7725	
	SOH hospital	3974	9.7828	43,153
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated
30%	Treat all	3635	9.7699	
	SOH hospital	3974	9.7828	26,334
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated
20%	Treat all	3713	9.7670	
	SOH hospital	3974	9.7828	16,455
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated

TABLE 59 One-way sensitivity analysis on the adherence rate for the monitoring pathways

Adherence rate	Pathway	Average total cost (discounted) (£)	Average total QALYs (discounted)	ICER (£)
75%	Treat all	3412	9.7757	
	SOH hospital	3974	9.7828	78,911
	SOH primary care	4555	9.7826	Dominated
	NICE conservative	5156	9.7813	Dominated
	NICE intensive	6879	9.7828	Dominated
80%	Treat all	3412	9.7760	
	SOH hospital	3886	9.7833	64,144
	SOH primary care	4666	9.7829	Dominated
	NICE conservative	5034	9.7832	Dominated
	NICE intensive	6811	9.7845	2,504,296
90%	Treat all	3412	9.7763	
	SOH hospital	3715	9.7868	28,723
	SOH primary care	4506	9.7865	Dominated
	NICE conservative	4864	9.7863	Dominated
	NICE intensive	6646	9.7879	2,678,852
95%	Treat all	3412	9.7765	
	SOH hospital	3635	9.7890	17,828
	SOH primary care	4429	9.7887	Dominated
	NICE conservative	4783	9.7878	Dominated
	NICE intensive	6559	9.7897	3,739,358

Discussion and summary

Key findings

No previous economic evaluation has reported a comparison of alternative surveillance pathways for surveillance of individuals with OHT from the perspective of the UK NHS. As a consequence, we developed an economic evaluation of alternative surveillance pathways.

The base-case cost-effectiveness analysis showed that the 'SOH hospital' pathway, 2-yearly monitoring to measure IOP and detect the onset of glaucoma (standard clinical assessment), was able to reduce the number of cases converting to OAG as well as avoid extra cases of progression to moderate and severe OAG compared with a 'treat all' pathway. However, the incremental cost per extra QALY (generated by those cases avoided) is far beyond the usual stated threshold of WTP for a QALY.[21] For the cost–benefit analysis the 'SOH hospital' pathway was the only pathway relative to 'no surveillance' that had a positive net benefit. The main drivers of these results appeared to be the very low valuations put on avoiding disease, the low number of events occurring and the relatively high costs of surveillance.

Strengths and limitations

Because of a lack of existing information on relative efficiency relevant to the NHS, we developed an economic evaluation of alternative pathways that was based on a discrete event simulation model. This approach enabled the introduction of further disease and event complexities than those allowed by alternative modelling approaches such as decision tree models or Markov models. We incorporated into the economic model data obtained from rigorous systematic reviews[268] as well as information retrieved and produced within this project, namely, we used the best available glaucoma risk prediction tool identified from a systematic review of alternative risk prediction tools (see *Chapter 4*) to generate the risk of conversion to OAG for OHT individuals, and data to inform the optimal frequency of surveillance came from the analysis of the signal-to-noise ratio reported in *Chapter 6*.

A novel feature of this analysis was the conduct and comparison of different methods of economic evaluation. Overall, the results of the cost–utility and cost–benefit analyses are similar, with surveillance pathways that provide more benefits (QALYs or WTP) not being considered worth the extra cost. The main difference of note was that the 'SOH hospital' pathway is unlikely to be considered worthwhile when efficiency is measured in terms of incremental cost per QALY but may be when efficiency is measured in terms of net benefits. This is mainly because of the value that society places on being monitored for a particular condition as well as the value placed on obtaining better communication and understanding during the monitoring visits. It is a matter for judgement whether the cost per case of unit of clinical effectiveness avoided is worthwhile, but the data on clinical effects does allow policy-makers to question whether the methods used to weight different clinical or natural measures into QALYs or WTP are reasonable.

The results were sensitive to the rate of adherence to treatment assumed in the model. We conducted extensive searches for relevant data but could find little robust evidence. As a consequence, the values used within the model were based on the limited available evidence[278–280] and expert opinion. Should the rates of adherence to medical treatments differ between monitoring and non-monitoring pathways then it is possible that the incremental cost per QALY of a surveillance pathway ('SOH hospital') may fall to a value that society might find acceptable.[261] Further research in this area is thus warranted.

The results were also sensitive to the unit price of hospital-based IOP measurement only follow-up visits with an ICER for the 'SOH hospital' pathway compared with the 'treat all'

pathway of around £30,000 at a unit price of just over £61. This can be explained by the relatively high number of IOP measurement only visits within the monitoring pathways. In this economic model there are no restrictions to the number of subsequent times that the IOP can be measured as 'off target', and therefore the number of times that individuals could – depending on the pathway– return for an IOP check at 2 or 4 months. This could have led to an overestimation of the number of IOP visits in the model and a slight underestimation of the cost-effectiveness of the monitoring pathways. This could partially explain the sensitivity of the results to the unit cost of hospital-based IOP measurement only visits. In reality, clinicians will eventually find the adequate drop combination to control IOP. However, caution should be used when interpreting the results of this sensitivity analysis as the lower hospital-based visit unit costs might not fully account for the opportunity cost of the use of those resources (e.g. take into account the potential lack of capacity in hospital-based eye services).

The value of the different methods of surveillance in terms of their ability to prevent conversion to or progression of OAG is modest. This is because OAG is a chronic disease in which the majority of affected individuals progress slowly to stages at which vision is substantially affected. This slow progression together with the relatively short time horizon (20 years) considered in the base-case analysis resulted in no one progressing into the most severe stages of disease. Increasing the time horizon would allow many of those who had converted to OAG to progress to more severe disease (as well as allowing more people to develop OAG).

Coupled with this, the available data suggested that there was no difference in quality of life between individuals with OHT and those converted to OAG because in its early stages OAG is asymptomatic. As a consequence of this, the difference in QALYs between surveillance pathways is, on average, very small. Similarly, in the cost–benefit analysis, the mean WTP to avoid a single case of OAG was very modest (£0.03) and given the predicted low risk of developing OAG the net benefits of surveillance are, as a consequence, very small.

The incorporation of a DCE into an economic evaluation model is novel and there have been relatively few examples where this has been attempted in the past. Previous work[281] attempted to incorporate an existing DCE into an economic model comparing alternative methods of surgery for inguinal hernia but as the DCE was not designed for incorporation into the economic model its incorporation placed restrictions on the model structure. For example, the DCE imposed a time horizon on the economic model that differed between attributes and was also insufficient to fully capture all of the differences in clinical effectiveness. A further study[282] developed a DCE and economic model in conjunction but the model developed was a simple decision-analytic model. This modelling approach was appropriate to the context but is rarely appropriate to capture the complexity of decision problems more usually faced. In this study, the DCE has been explicitly designed to facilitate its incorporation into a far more complex form of model, a discrete event simulation, which to the best of our knowledge has never before been attempted. To do this represented a major challenge, and it was surprising how little respondents valued avoiding an extra case of OAG, severe OAG and visual impairment, especially given comments during initial piloting about how important avoiding OAG and its consequences were. One reason for this is that the respondents to the DCE were members of the general public and not those with OAG or OHT. It is possible that had the results of the DCE been restricted to these groups the valuations would have differed and hence the results of the cost–benefit analysis would have differed. Further research will be needed to explore this issue in more detail.

Care should be taken when interpreting the results of the cost–benefit analysis. The WTP values generated in *Chapter 7* were derived from a DCE in which individuals were asked to choose between alternative arrangements compared to those as a result of the health-care status quo for a 10-year time horizon. We tried to adjust for this when calculating final net benefit figures for the alternative model pathways. The 10-year time horizon for the DCE was chosen before the

results of the discrete event simulation were available (and hence before predictions on rates of conversion and progression were estimated). This was due to the necessity of conducting the different elements of the project concurrently rather than consecutively.

NICE-based pathways were developed in this study as a way of representing current practice within the economic evaluation model. It has been argued that there is a lack of capacity within hospital eye services in the UK. We retrieved figures from the NHS Scotland Information Services Division to cost ophthalmology visits in the model. This might be an underestimation of the true 'opportunity' costs of these visits. The opportunity cost in this case would be given by the benefits foregone as a result of other conditions or diseases not being attended to at the hospital level. If there is a lack of capacity in UK hospital eye services then hospital-based pathways will not be feasible and only the 'treat all' and 'SOH primary care' pathways would be viable.

We used the risk prediction tool from *Chapter 4* to model the underlying condition of the sampled individuals in the model. We used this same tool to assess the risk of conversion within the SOH pathways. This helps explain the better performance of the SOH pathways compared with the NICE-based pathways. However, this is the best available information we have to assess OHT individuals at risk of developing OAG.

Finally, because the discrete event simulation model was computationally demanding we could not run probabilistic sensitivity analyses and therefore we did not obtain the value of information for further research. The current analysis failed to fully capture the parameter uncertainty in our economic model and, as such, its results should be interpreted with caution.

Uncertainties

As with any economic evaluation there are a number of uncertainties. These can be split into two broad categories: structural and parameter uncertainties.[283,284] The first category relates to uncertainties about the particular care pathway that should be modelled or about the type of mathematical functions that govern the relationship between different elements of the model. In the second category, uncertainties are related to what particular parameter values should be within the economic model. Outlined below are the key structural and parameter uncertainties.

Structural uncertainties

■ Model time horizon. The base-case analysis took a 20-year time horizon. This period of time might not be long enough to capture the whole effect of surveillance pathways compared with no surveillance. However, we chose this particular time period as we believed that extrapolating 5-year effectiveness data (e.g. 5-year risk of conversion to OAG) to even longer time horizons would be questionable. Had we chosen a longer time horizon some cases of visual impairment would have occurred and hence differences in QALYs would have potentially been captured. However, given the time it takes to develop visual impairment, discounting would tend to reduce the impact of differences in the number of cases of visual impairment.

■ Time to event function. In the absence of other information we opted to use a uniform distribution to model time to conversion, and progression to moderate OAG, severe OAG and visual impairment. This approach means that the number of events occurring in each unit of time is the same. An alternative would have been to use Kaplan–Maier curves derived from the data from the identified RCTs, which would have provided a cumulative conversion function for the 20-year period. However, such a function would not have provided predictions of the risk of conversion that were fully consistent with those provided by the risk prediction tool.

■ Pathways included in the model. We did not formally include a 'do nothing' alternative, although a simplified version of one was required for analytical purposes in the cost–benefit analysis. When developing the pathways to be compared it was believed that a 'do nothing'

pathway would not be relevant as the policy currently is to monitor individuals with OHT in the UK.

Parameter uncertainty

- We ran one-way sensitivity analysis as well as scenario analysis. As often stated, this might not fully capture the joint parameter uncertainty and its effect on results.[284] Probabilistic sensitivity analysis is often used to address this issue; however, our model proved to be computationally too demanding and probabilistic sensitivity analysis was not feasible. As a result, we could not produce cost-effectiveness acceptability curves or value of information analysis. This is a limitation of the present analysis.
- We assumed 50% and 75% adherence rates for non-monitoring ('treat all') and monitoring (NICE or SOH) pathways, respectively. We based these figures on limited available evidence and expert opinion. These rates might range between 20% and 95%, depending on which data are judged relevant.[278] The results of the cost–utility analysis were shown to be sensitive to variations in the rate of adherence between surveillance pathways and further research on this would reduce the uncertainty surrounding the results.

Conclusions

We conducted cost-effectiveness, cost–utility and cost–benefit analyses of monitoring pathways for individuals with OHT using a discrete event simulation model. We incorporated into this economic model data obtained from rigorous systematic reviews[268] as well as information retrieved and produced within this project [e.g. best available glaucoma risk prediction tool (see *Chapter 4*) and data to inform the optimal frequency of surveillance (see *Chapter 6*)].

Overall, the results of the cost–utility and cost–benefit analyses are similar, with interventions that provide more benefits (QALYs or WTP) not being considered worth the extra cost. The main difference of note was that the 'SOH hospital' pathway is unlikely to be considered worthwhile when efficiency is measured in terms of incremental cost per QALY but may be when efficiency is measured in terms of net benefits. This is mainly because of the value that society places on being monitored for a particular condition as well as the value placed on the patient experience of the testing process during the monitoring visits, which is captured in the WTP estimates but not by the QALYs. The cost–utility analysis results were particularly sensitive to the rate of adherence to medical treatment assumed for the 'treat all' and the other surveillance pathways modelled as well as the unit price of IOP measurement only hospital visits. The 'SOH hospital' pathway has an incremental cost per QALY of < £30,000 compared with the 'treat all' pathway when the cost of return visits to hospital to measure IOP was around £60 or less. It is unlikely that either of the NICE strategies is cost-effective compared with the less intensive monitoring pathways developed in this study. The feasibility of a 'treat all' pathway and its acceptability by patients and providers have not been explored.

Any policy recommendations about the placing of a surveillance programme at a hospital level should take into account capacity in hospital eye services in the UK. If a hospital-based surveillance programme is not feasible then the modelling suggests that basing the programme in the community is the second best option.

Further research is required to provide new primary data to update the economic model. In particular, robust data are required on adherence to ocular hypotensive treatment and the impact of treatment on health status. The feasibility of alternative low-cost monitoring pathways should be explored. New technologies suitable for self-monitoring are required.

Chapter 9
Discussion

Summary of main findings

Glaucoma is a chronic progressive optic neuropathy leading to impaired vision and blindness if inadequately treated. OAG is the most common form. A raised IOP is the only modifiable risk factor. OHT is defined as IOP of > 21 mmHg and the absence of clinical signs of glaucoma. Although it is important to treat early disease to reduce the costs and disability associated with glaucoma, only a minority of those with OHT develop glaucoma and, of these, not all are at risk of visual impairment in their lifetime. The challenge is therefore to identify the groups most at risk of developing significant glaucoma, and to determine monitoring criteria for those individuals who are most likely to benefit from surveillance and treatment. For monitoring to be worthwhile it would have to lead to early identification of the need for treatment or change of treatment, or development of glaucoma, and thus reduce sight loss while not causing unacceptable harms associated with monitoring and treatment. Furthermore, it must represent a good use of scarce health-care resources and also be acceptable to patients.

To guide management of patients with OHT there is a need for accurate and reliable risk assessment that would help predict conversion from OHT to glaucoma. We identified three published models predicting the risk of developing OAG for individuals with OHT.[42,62–64] These models were derived using data from two large multicentre RCTs evaluating the efficacy of ocular hypotensive medication, the OHTS[42] and the EGPS.[14] The OHTS-EGPS means model[64] was the most robust and estimated the 5-year risk of conversion to glaucoma based on age, IOP, CCT, VCD ratio and PSD. The model uses the means of the right and left eyes of an individual to calculate eye-specific predictors and all variables can be routinely collected in clinical practice. The discriminatory ability of the model was good in the four populations included in our study, with c-indexes between 0.69 and 0.83. However, in calibration analyses, the OHTS-EGPS model generally overestimated the risk of OAG in all cohorts, although for the Rotterdam cohort the calibration slope was close to 1 (1.09, 95% CI 0.72 to 1.46), the ideal value when there is complete agreement between predicted and observed risks. A calculator to estimate the 5-year risk of developing OAG, based on the pooled OHTS–EGPS means model, is available online at http://ohts.wustl.edu/risk/calculator.html. This can be used in conjunction with clinical assessment to inform patients of their glaucoma risk and to guide treatment decisions.

Once OHT is diagnosed, monitoring includes measuring IOP by tonometry (as treatment decisions may need to be changed according to the level of IOP) and tests to detect glaucoma. These include tests of visual function, commonly SAP, and an evaluation of structural changes in the optic nerve by clinical examination, photography and/or automated estimates of RNFL or optic disc neuroretinal rim loss. The clinically important changes in the test, the signal, need to be considered alongside the noise, which includes biological variability (over short periods) and measurement error (including interobserver and device variability), in successive measurements.

Several tonometers are available for the measurement of IOP. To estimate the measurement error between different tonometers we undertook a systematic review of published directly comparative studies. We compared the agreement of IOP readings between tonometers available in clinical practice and GAT, the commonly accepted reference standard device for measuring

IOP. We identified 102 studies comparing eight candidate tonometers with GAT. The studies were generally poorly reported. Sizeable inter- and intraobserver variability was observed for all tonometers, including GAT, casting doubt on the validity of GAT as the default standard.

We explored the signal-to-noise relationship of serial measures of IOP and visual fields in a secondary analysis of IPD from a RCT including data on 300 participants with OHT over a 4-year follow-up. The results were validated in a second independent trial data set. In terms of IOP, the average change in IOP over time for the whole group was close to zero (a decrease of < 1 mmHg in a 3-year period), but when it changes it can change quickly: > 5 mmHg at 3 years in about 25% of participants. For most individuals, any true change in underlying IOP ('signal') was smaller than the estimated 'noise'. The 'noise' was lower in those with high baseline IOP (≥ 26 mmHg). Observed changes of ≤ 3 mmHg could be explained as 'noise'. Assuming independence of repeated measures, the mean of two baseline IOP readings increased the signal-to-noise ratio such that true change in IOP could be detected at 2 years. With three baseline measures, true change could be detectable between 1 and 2 years. In those with lower baseline IOP (< 26 mmHg), a true change in IOP would be unlikely to happen within 3 years. Age and observer did not affect the findings. A similar exercise was carried out regarding visual field tests. Serial data on visual fields, measured by SAP, were available in only one data set, and these were limited to sufficient data to explore only the variability of one index of visual field loss, the MD. The variability of the MD of repeated visual fields was explored, estimating the mean and variance of change at every visit. The MD fluctuated, both increasing and decreasing, with minimal signal detected over 4 years.

To estimate the efficiency of alternative surveillance regimes, we undertook an economic evaluation, which was informed by public preferences for monitoring ascertained by a survey of the public ($n = 814$) using a DCE design. The findings of the survey showed a general preference for being monitored if an individual had OHT. Individuals were willing to pay £28 per year for a monitoring service, everything else being equal. Health outcome (risk of conversion to OAG, progression to severe glaucoma or visual impairment) and patient experience (communication and understanding between patient and health-care professional and side effects of treatment), other than location of service (hospital or community optometrist), were significant predictors of preferences at the 1% level. Marginal valuations of the risk of developing glaucoma and sight loss over 10 years were statistically significant in the expected direction but very small. Side effects of treatment reduced the value of any service, with disbenefits increasing as side effects became more severe. Good communication with the health professional and understanding of the process of testing were important predictors of the value of alternative services. Preferences varied according to age, with those aged > 50 years being less concerned with the risk of sight loss (compared with those aged < 50), but more concerned about medication side effects and the importance of good communication and understanding of the process.

In the economic modelling component of the evaluation we used cost–consequence, cost–utility and cost–benefit analyses to compare the efficiency of five alternative surveillance pathways. Two were based on NICE guidelines, with the monitoring interval depending on initial risk stratification: 'NICE intensive' – 4-monthly to annual monitoring – and 'NICE conservative' – 6-monthly to biennial monitoring; two further pathways were developed based on the findings of this study, differing in location ('SOH hospital' and 'SOH primary care') and with monitoring biennially; and a final 'treat all' pathway, identified in the literature, involved treatment with a PGA if IOP was > 21 mmHg, with IOP measured annually in community optometry and referral to a hospital eye service if response to treatment was inadequate (< 15% IOP reduction).

'Treat all' was the least costly pathway and 'NICE intensive' the most costly. The 'SOH hospital' pathway reduced the number of cases of conversion to glaucoma and progression compared

with the 'treat all' pathway and provided more QALYs but the ICER was considerably > £30,000. The 'NICE intensive' pathway also avoided conversion to glaucoma and progression, but NICE-based pathways are dominated (more costly and less effective) by the 'SOH hospital' pathway. In the cost–benefit analysis, compared with a 'no monitoring option', 'SOH hospital' was the only pathway with a net benefit.

The results were sensitive to the risk threshold for initiating treatment. If treatment was initiated when the 5-year risk of developing OAG was > 10% (e.g. a 60 year old with ocular parameters based on means of measurements in both eyes of IOP of 27 mmHg, CCT of 560 μm, VCD ratio of 0.4 and PSD of 1.4 dB has a 10.3% likelihood of developing OAG in at least one eye), the 'SOH hospital' pathway was less costly and more effective than the 'treat all' pathway. Differences in treatment adherence rates between the 'treat all' and SOH pathways of approximately ≥ 40% led to the SOH pathways having an ICER of ≤ £30,000.

The results were also sensitive to the unit price of hospital-based IOP measurement only follow-up visits, with an ICER for the 'SOH hospital' pathway compared with the 'treat all' pathway of around £30,000 at a unit price of just over £61. This can be explained by the relatively high number of IOP measurement only visits within the monitoring strategies. This could have led to an overestimation of the number of IOP visits in the model and a slight underestimation of the cost-effectiveness of the monitoring strategies. This could partially explain the sensitivity of the results to the unit cost of hospital-based IOP measurement only visits. Caution should be taken when interpreting the results of this sensitivity analysis as the lower hospital-based visit unit costs might not fully account for the opportunity cost of the use of those resources (e.g. take into account the potential lack of capacity in hospital-based eye services). Patient self-measurement as part of a point of care testing programme to monitor responses to initiation or changes in therapy could be considered but this option is limited by the reliability of the currently available tonometers.

In every sensitivity analysis conducted, NICE-based pathways were the most costly and were either dominated or had an ICER well above £30,000 per QALY.

Strengths and limitations

The strength of this study is the multidisciplinary approach including robust systematic reviews of the literature, analysis of existing IPD sets and consultation with patients, clinical experts and the public to inform the individual and ocular characteristics of a population with OHT and the surveillance pathways for comparison within an economic modelling evaluation framework. In addition to cost-effectiveness and cost–utility analyses, we report a cost–benefit analysis using the monetary values derived from a survey of public preferences for alternative monitoring programmes. Thus, the economic evaluation considered patient experience factors as well as health outcomes when valuing alternative monitoring programmes.

There are several limitations to the study and interpretation of the findings should be carried out with caution. The main limitation is the limited available data, both in the published literature and in primary individual-level data sets, on which our analyses were based.

The most robust risk prediction model identified from the literature was the OHTS-EGPS model. However, this was developed in a selected population within the context of a RCT and is not necessarily generalisable to the population of patients with OHT in clinical practice. Important prognostic factors may not have been included. Similarly, our validation data sets were selected populations, although we made an effort to avoid this problem by including both a trial-based

and an observational setting. The definitions of glaucoma varied across data sets, and the observational data sets did not include all patients with OHT. Thus, the findings may not be generalisable. However, despite these limitations in the available data the OHTS-EGPS model generally discriminated well in estimating the 5-year risk of conversion to OAG. The model was not well calibrated in all populations. Further validation, especially using a larger sample including non-Caucasians, and recalibration in other patient subgroups (e.g. those with a family history of glaucoma) are required.

Ideally, the optimal interval for repeat testing to detect the onset of glaucoma should be informed by the variability in the tests (noise) and the number of tests required to detect a true change (signal). We aimed to explore the variability of IOP measurements and clinical measures of glaucoma within a signal-to-noise framework. Data were available on repeated measures of IOP, but no data were available to estimate the variability in measures of structural damage to the optic nerve and only limited data were available on repeated visual field examinations. Serial data on visual fields, measured by SAP, were available only in the Moorfields data set and these were sufficient to explore only the variability of one index of visual field loss, the MD. The MD represents the overall reduction in sensitivity, averaged across all test point locations in the visual field, compared with age-matched healthy control subjects.[285] The MD, although widely used to monitor glaucoma,[286] is not a useful measure for monitoring OHT as repeated measures of MD measured only noise over a 4-year period. There is no agreed summary index of visual field loss that constitutes definite glaucoma. The glaucoma hemifield test (GHT) has been incorporated into the commonly used standard automated perimeter, the Humphrey perimeter, to diagnose glaucoma,[287] and classifies the visual field as 'within normal limits', 'outside normal limits', 'borderline', 'general loss of sensitivity' and 'abnormally high sensitivity'. Its performance to detect conversion from OHT to glaucoma is uncertain.[288] The variability of the GHT in a signal-to-noise framework has not be explored. The Visual Field Index is a new measure[289] that quantifies the total visual field loss as a percent rather than in decibels, but its performance in a context of monitoring OHT to detect early glaucoma has not been established.[290]

We established the measurement variability in IOP based on a rigorous systematic review and on a secondary analysis of serial measures of IOP by GAT from the Moorfields data set. A number of factors, such as CCT, are thought to influence IOP measurement, but our systematic review was not able to identify their impact. Primarily this was because of poor reporting of the studies not having the IPD to tease out any relationships for important subgroups. Similarly, we were unable to estimate any interobserver variability. In the IPD analysis of the Moorfields data set, variability in IOP measurement did not differ between observers.

Estimates of the variability between tonometers and the signal-to-noise relationship in repeated measures of IOP were incorporated into the economic evaluation. We found that IOP level remains fairly constant in subjects with OHT. True changes in IOP, within 2 years, were very unlikely. Thus, we used this 2-year interval to define the care pathways for the 'SOH hospital' and 'SOH primary care' strategies in the economic evaluation. Although IOP is not a clinically valid measure to detect glaucoma it is the measure on which treatment decisions are made and the variability in IOP is important when monitoring response to treatment. Further research is required to estimate the variability of visual field testing to inform the optimal monitoring interval for those with OHT, and to explore the role of automated imaging technologies in detecting conversion to glaucoma.

We developed the surveillance pathways based on input from clinical experts, methodologists, the literature and findings from other study objectives. These included the best risk prediction model, optimal frequency of testing and the survey of public preferences (the DCE). The NICE-based pathways were considered to be current best monitoring practice. Ideally, the new care

pathways in the economic model should have been explored in the development and conduct of the DCE. This was not possible in the timeline of the project. A 'treat all' pathway was included based on emerging findings from the literature and appears to be a more efficient strategy. However, there are major uncertainties around the acceptability of a 'treat all' strategy to patients or heath-care providers and these were not explored in this study.

Across all pathways it was not possible to disentangle quality of life decrements as a result of side effects of treatment. However, our quality of life data (EQ-5D scores) were retrieved from a sample of treated participants, and a proportion of these had side effects. Therefore, the average quality of life scores used did incorporate quality of life decrements as a result of treatment side effects.

In the cost–utility analysis we did not compare surveillance with a 'do nothing' alternative as this was not an acceptable option given current NHS policy. The modelling took a 20-year time horizon, which might not be sufficient to capture the long-term benefits of surveillance. The economic model used (discrete event simulation approach) proved to be computationally demanding. Because of this we could not undertake a probabilistic sensitivity analysis. Therefore, the sensitivity analyses conducted may not fully capture the uncertainty surrounding parameter estimates. However, we believe that our deterministic sensitivity analysis reflects the key uncertainties in the analysis and managed to identify the main drivers of the economic model results (e.g. adherence rate to treatment and side effects of treatment and the costs of visits to secondary care to measure IOP following treatment changes when the IOP was 'off target'). Although patient views were consulted when developing the DCE, the results were based on public preferences, which could differ from those of patients.

Chapter 10

Conclusions

Implications for health care

The best available risk prediction model to estimate the 5-year risk of developing glaucoma is based on age and ocular predictors, namely IOP, CCT, VCD ratio and PSD. This risk prediction model is publicly available online at http://ohts.wustl.edu/risk/calculator.html and can be used to provide more accurate risk assessment in conjunction with clinical assessment. The model was developed from the characteristics of participants enrolled in RCTs and thus may not include all important predictors in the general population of patients with OHT.

An algorithm for the measurement of IOP using the average of repeat measurements would reduce measurement 'noise'. Our findings support the clinical importance of establishing a true baseline IOP before initiating monitoring or treatment. Using the average of two measures at baseline at a single visit increases the signal-to-noise ratio such that one would expect to detect a true change, of the order of 2 mmHg, at 2 years. The data sets analysed did not have repeated measures on the same visit, thus the recommendation of reducing noise is based on an assumption in the model that the measurement errors across repeated readings are not correlated (i.e. all errors moving in the same direction). Repeated measures at the same visit will not incorporate diurnal or short-term biological variation. Diurnal variation can be addressed by standardising the time of measurement. Short-term biological variation could be addressed by averaging the IOP over repeat visits but this has to be balanced against the opportunity costs of an additional visit. IOP measurement using the NCT or Perkins tonometer is in close agreement with measurement using GAT in > 50% of measurements. However, the evidence suggests that GAT may not be the most appropriate reference standard. The same tonometer should be used when comparing IOP measurements in an individual. The optimal frequency of measurement of the visual field or the optic nerve head/RNFL to detect the onset of glaucoma remains uncertain.

For confirmed OHT there is no clear benefit from intensive monitoring. Our findings, based on data derived from participants in two relatively small RCTs, which may not be representative of the general population of patients with OHT, suggest that 2-yearly IOP monitoring of untreated or stable treated OHT is the recommended interval to detect true change of the order of 2 mmHg. The average of repeated measures of IOP at each visit reduces measurement noise. Our findings, based on an economic evaluation, suggest that if the costs of repeat visits to measure IOP following treatment initiation or change are minimised (< £60 per visit), then the surveillance (SOH) pathway developed in this study, based in the hospital, with 2-yearly monitoring to detect glaucoma, may be a cost-effective option compared with a 'treat all' policy with annual monitoring of IOP only, without clinical tests for glaucoma. Our findings may not fully account for the opportunity costs of the use of those resources (e.g. may not take into account the potential lack of capacity in hospital-based eye services). The feasibility of alternative low-cost monitoring pathways should be explored. Patient self-measurement as part of a point of care testing programme to monitor responses to initiation or changes in therapy could be considered, but this option is limited by the reliability of the currently available tonometers. Service reconfigurations should consider the patient experience, ensuring adequate time to explain the purpose of monitoring, and avoid treatment side effects.

If services have adopted NICE guidance on monitoring OHT, then our findings suggest that efficiency may improve by moving to a less intensive approach. If NICE guidance has not been adopted, the findings of this study could be used to inform the development of an alternative hospital- or community-based service.

Implications for research

Further research is recommended to provide new primary data to inform on the optimal organisation of a surveillance programme for those identified at risk of glaucoma as a result of OHT:

- A prospective cohort study including newly diagnosed patients with OHT with long-term outcomes. The purpose of the study will be to update and validate the OHT-EGPS risk prediction model and refine the monitoring and treatment criteria developed in this study in a larger sample representative of the general population of patients with OHT. Before commencing the study, standardisation of a summary measure of automated perimetry (and/or RNFL analysis), with consensus on what constitutes conversion to glaucoma, is required. A comparison of alternative tonometers, and factors influencing IOP measurements, could be nested within the study. Costs and patient preferences for alternative surveillance and treatment programmes should be explored. The data could be used to update the economic model developed in this study.
- The development of new technologies to meet the needs of the NHS and patients. This includes further development of tonometers suitable for self-monitoring.

Acknowledgements

We would like to thank:

- the study authors who we contacted who provided additional details of their studies
- Lara Kemp for secretarial support, Fiona Stewart for referencing the final report and Richard Wormald for valuable comments on early drafts
- Paul Tappenden for support with the economic evaluation modelling, Barbara Eberth for support with regression models for the DCE analysis and Verity Watson for general advice on the DCE
- The Rotterdam Ophthalmic Institute for access to datasets from the 'Rotterdam' trial
- Karen Armstrong-Owen (specialist nurse, Eye Department, Queens Medical Centre, Nottingham), David Holgate and Stuart Dobson (Medical Informatics, Queen Margaret Hospital, Dunfermline, Fife) and Ananth Viswanathan (consultant ophthalmologist, Moorfields Eye Hospital, London) for providing and collating data for the Nottingham, Dunfermline and Moorfields data sets, respectively
- Takehiro Yamashita (research fellow, Moorfields Eye Hospital, London) for grading the optic disc photographs in the Rotterdam data set
- Kirsty McCormack for support in setting up the study and data abstraction for the systematic review of tonometers
- Ann Hoste on behalf of the Belgian Glaucoma Society for permission to use their pictorial representation of glaucoma
- Alia Ali for clinical support in the focus groups, and for supporting the data abstraction and analysis of the systematic review of tonometers
- members of the advisory panel – Aachal Kotecha (academic optometrist, London), Gordon Porteous (specialist ophthalmic nurse, Aberdeen), Anthony King (consultant ophthalmologist, Nottingham), Graeme Shand (community optometrist with a special interest in glaucoma, Angus) and Joan Black (Eye Services Manager, Queens Medical Centre, Nottingham) – for their contribution to the development of the attributes included in the DCE
- the patients who participated in the focus group discussion and the advisory panel for their invaluable contribution to the project.

Contribution of authors

Jennifer Burr led the writing of *Chapters 1–3* and *9–10*. Paola Botello-Pinzon was the research fellow/project manager and led the collation of the data sets and co-ordinated and contributed to the systematic review of risk prediction, agreement between tonometers and the DCE. Jennifer Burr and Katie Banister commented on and edited the final report. Cynthia Fraser developed and executed the searches and supervised all information retrieval for the study. Jonathan Cook provided methodological oversight for the whole project.

Yemisi Takwoingi and Jonathan Deeks led the work on risk prediction and wrote the associated chapter (*Chapter 4*). Jennifer Burr, Paola Botello-Pinzon and Augusto Azuara-Blanco independently reviewed all titles and abstracts in terms of their eligibility for inclusion.

References

1. National Institute for Health and Clinical Excellence. *CG85 glaucoma: diagnosis and management of chronic open angle glaucoma and ocular hypertension*. London: NICE; 2009. URL: http://guidance.nice.org.uk/CG85/Guidance/pdf/English (accessed February 2011).

2. Bamashmus MA, Matlhaga B, Dutton GN. Causes of blindness and visual impairment in the West of Scotland. *Eye* 2004;**18**:257–61.

3. Bunce C, Wormald R. Leading causes of certification for blindness and partial sight in England & Wales. *BMC Public Health* 2006;**6**:58.

4. Burr JM, Kilonzo M, Vale L. Developing a preference-based glaucoma utility index using a discrete choice experiment. *Optom Vis Sci* 2007;**84**:797–808.

5. 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.

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

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

8. 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.

9. 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.

10. Leske MC. The epidemiology of open-angle glaucoma: a review. *Am J Epidemiol* 1983;**118**:166–91.

11. Varma R, Ying-Lai M, Francis BA, Nguyen BBT, Deneen J, Wilson MR, *et al.* Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. *Ophthalmology* 2004;**111**:1439–48.

12. Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, *et al.* The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:701–13.

13. Kass MA, Gordon MO, Gao F, Heuer DK, Higginbotham EJ, Johnson CA, *et al.* Delaying treatment of ocular hypertension: the Ocular Hypertension Treatment Study. *Arch Ophthalmol* 2010;**128**:276–87.

14. Miglior S. Results of the European Glaucoma Prevention Study. *Ophthalmology* 2005;**112**:366–75.

15. Vass C, Hirn C, Sycha T, Findl O, Bauer P, Schmetterer L. Medical interventions for primary open angle glaucoma and ocular hypertension. *Cochrane Database Syst Rev* 2007; Issue 4, Art. No. CD003167.

16. Office for National Statistics. *Census 2001: population report United Kingdom*. London: Office for National Statistics; 2002. URL: www.statistics.gov.uk/census2001/pop2001/united_kingdom.asp (accessed March 2011).

17. National Institute for Health and Clinical Excellence. *Commissioning a service for people at risk of developing glaucoma*. London: NICE; 2010. URL: www.nice.org.uk/usingguidance/ commissioningguides/glaucoma/commissioningservicepeopledevelopingglaucoma.jsp (accessed March 2011).

18. College of Optometrists. *Examining the patient at risk of glaucoma*. London: College of Optometrists; 2011. URL: www.college-optometrists.org/download.cfm/docid/D15DFBC3-5146-432D-BEFE9681ED555393 (accessed May 2011).

19. NHS Scotland. *Ophthalmology patient pathways*; 2008. URL: www.pathways.scot.nhs.uk/ ophthalmology.htm (accessed March 2011).

20. The National Health Service (General Ophthalmic Services and General Dental Services) (Scotland) Amendment Regulations 2010. Edinburgh: Scottish Statutory Instruments; 2010. URL: www.legislation.gov.uk/ssi/2010/378/contents/made (accessed March 2011).

21. National Institute for Health and Clinical Excellence. *CG85 glaucoma: diagnosis and management of chronic open angle glaucoma and ocular hypertension. Costing report: implementing NICE guidance*. London: NICE; 2009. URL: www.nice.org.uk/nicemedia/pdf/ CG85CostReport.pdf (accessed March 2011).

22. European Glaucoma Society. *Terminology and guidelines for glaucoma*, 3rd edn. Savona, Italy: European Glaucoma Society; 2008. URL: www.eugs.org/eng/EGS_guidelines.asp (accessed May 2011).

23. American Academy of Ophthalmology Glaucoma Panel. *Preferred practice pattern guidelines: primary open-angle glaucoma suspect*. San Francisco: American Academy of Ophthalmology; 2010. URL: http://one.aao.org/asset.axd?id=a860f57a-0e6a-4c4f-b0f7-1a42e05073ff (accessed May 2011).

24. Australian Government National Health and Medical Research Council. *NHMRC guidelines for the screening, prognosis, diagnosis, management and prevention of glaucoma 2010*. Canberra: Australian Government National Health and Medical Research Council; 2010. URL: www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/cp113_glaucoma_ nov_2010.pdf (accessed May 2011).

25. Chihara E. Assessment of true intraocular pressure: the gap between theory and practical data. *Surv Ophthalmol* 2008;**53**:203–18.

26. Damji KF, Muni RH, Munger RM. Influence of corneal variables on accuracy of intraocular pressure measurement. *J Glaucoma* 2003;**12**:69–80.

27. Liang SYW, Lee GA, Shields D. Self-tonometry in glaucoma management – past, present and future. *Surv Ophthalmol* 2009;**54**:450–62.

28. 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.

29. Kymes SM, Kass MA, Anderson DR, Miller JP, Gordon MO, Ocular Hypertension Treatment Study Group. Management of ocular hypertension: a cost-effectiveness approach from the Ocular Hypertension Treatment Study. *Am J Ophthalmol* 2006;**141**:997–1008.

30. van Gestel A, Severens H, Webers C, Beckers H, Jansonius N, Schouten J. A novel approach to establish at which IOP to start treatment for ocular hypertension. World Glaucoma Congress, Boston, 2009.

31. van Gestel A, Severens JL, Webers CA, Beckers HJ, Jansonius NM, Schouten JS. Modeling complex treatment strategies: construction and validation of a discrete event simulation model for glaucoma. *Value Health* 2010;**13**:358–67.

32. Kamal D, Garway-Heath D, Ruben S, O'Sullivan F, Bunce C, Viswanathan A, *et al.* Results of the betaxolol versus placebo treatment trial in ocular hypertension. *Graefes Arch Clin Exp Ophthalmol* 2003;**241**:196–203.

33. Gaasterland DE, Ederer F, Sullivan EK, Caprioli J, Cyrlin MN. Advanced glaucoma intervention study: 2. Visual field test scoring and reliability. *Ophthalmology* 1994;**101**:1445–55.

34. Reus NJ, Colen TP, Lemij HG. The prevalence of glaucomatous defects with short-wavelength automated perimetry in patients with elevated intraocular pressures. *J Glaucoma* 2005;**14**:26–9.

35. Dielemans I, de Jong PT, Stolk R, Vingerling JR, Grobbee DE, Hofman A. Primary open-angle glaucoma, intraocular pressure, and diabetes mellitus in the general elderly population. The Rotterdam Study. *Ophthalmology* 1996;**103**:1271–5.

36. Grodum K, Heijl A, Bengtsson B. Risk of glaucoma in ocular hypertension with and without pseudoexfoliation. *Ophthalmology* 2005;**112**:386–90.

37. Le A, Mukesh BN, McCarty CA, Taylor HR. Risk factors associated with the incidence of open-angle glaucoma: the Visual Impairment Project. *Invest Ophthalmol Vis Sci* 2003;**44**:3783–9.

38. Leske MC. Risk factors for incident open-angle glaucoma. The Barbados Eye Studies. *Ophthalmology* 2008;**115**:85–93.

39. Wolfs RC, Klaver CC, Ramrattan RS, Van Duijn CM, Hofman A, De Jong LA. Genetic risk of primary open-angle glaucoma: population based familial aggregation study. *Arch Ophthalmol* 1998;**116**:1640–5.

40. Epstein DL, Krug JH Jr, Hertzmark E, Remis LL, Edelstein DJ. A long-term clinical trial of timolol therapy versus no treatment in the management of glaucoma suspects. *Ophthalmology* 1989;**96**:1460–7.

41. European Glaucoma Prevention Study Group, Miglior S, Pfeiffer N, Torri V, Zeyen T, Cunha-Vaz J, *et al.* Predictive factors for open-angle glaucoma among patients with ocular hypertension in the European Glaucoma Prevention Study. *Ophthalmology* 2007;**114**:3–9.

42. Gordon MO. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:714–20.

43. Kass MA, Gordon MO, Hoff MR, Parkinson JM, Kolker AE, Hart WM Jr, *et al.* Topical timolol administration reduces the incidence of glaucomatous damage in ocular hypertensive individuals. A randomized, double-masked, long-term clinical trial. *Arch Ophthalmol* 1989;**107**:1590–8.

44. Schulzer M, Drance SM, Douglas GR. A comparison of treated and untreated glaucoma suspects. *Ophthalmology* 1991;**98**:301–7.

45. Coleman AL. Risk factors for glaucoma onset and progression. *Surv Ophthalmol* 2008;**53**:S3–10.

46. British Cardiac Society, British Hypertension Society, Diabetes UK, HEART UK, Primary Care Cardiovascular Society, Stroke Association. JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005;**91**(Suppl. 5):v1–52.

47. Cooper A, O'Flynn N, Guideline Development Group. Risk assessment and lipid modification for primary and secondary prevention of cardiovascular disease: summary of NICE guidance. *BMJ* 2008;**336**:1246–8.

48. Fechtner RD. Evolving global risk assessment of ocular hypertension to glaucoma. *Curr Opin Ophthalmol* 2007;**18**:104–9.

49. Girkin CA, Kannel WB, Friedman DS, Weinreb RN. Glaucoma risk factor assessment and prevention: lessons from coronary heart disease. *Am J Ophthalmol* 2004;**138**(Suppl. 3):S11–18.

50. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, *et al.* External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;**56**:826–32.

51. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;**19**:453–73.

52. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;**277**:488–94.

53. Coleman AL. Glaucoma. *Lancet* 1999;**354**:1803–10.

54. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;**48**:1503–10.

55. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;**49**:1373–9.

56. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–94.

57. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley; 1987.

58. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87.

59. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York, NY: Springer; 2009.

60. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001;**21**:45–56.

61. Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol* 1999;**117**:573–83.

62. Coleman AL. Baseline risk factors for the development of primary open-angle glaucoma in the Ocular Hypertension Treatment Study. *Am J Ophthalmol* 2004;**138**:684–5.

63. Medeiros FA, Weinreb RN, Sample PA, Gomi CF, Bowd C, Crowston JG, *et al.* Validation of a predictive model to estimate the risk of conversion from ocular hypertension to glaucoma. *Arch Ophthalmol* 2005;**123**:1351–60.

64. Ocular Hypertension Treatment Study Group, European Glaucoma Prevention Study Group, Gordon MO, Torri V, Miglior S, Beiser JA, *et al.* Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology* 2007;**114**:10–19.

65. Ocular Hypertension Treatment Study Group and the European Glaucoma Prevention Study Group. The accuracy and clinical application of predictive models for primary open-angle glaucoma in ocular hypertensive individuals. *Ophthalmology* 2008;**115**:2030–6.

66. Gordon MO, Beiser JA, Kass MA, Ocular Hypertension Treatment Study Group. Is a history of diabetes mellitus protective against developing primary open-angle glaucoma? *Arch Ophthalmol* 2008;**126**:280–1.

67. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. *J Natl Cancer Inst* 2006;**98**:637–40.

68. Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. *Circulation* 2007;**122**:682–9.

69. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001;**286**:180–7.

70. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001;**93**:358–66.

71. Brandt JD, Beiser JA, Kass MA, Gordon MO. Central corneal thickness in the Ocular Hypertension Treatment Study (OHTS). *Ophthalmology* 2001;**108**:1779–88.

72. Foster PJ, Baasanhu J, Alsbirk PH, Munkhbayar D, Uranchimeg D, Johnson GJ. Central corneal thickness and intraocular pressure in a Mongolian population. *Ophthalmology* 1998;**105**:969–73.

73. Doughty MJ, Zaman ML. Human corneal thickness and its impact on intraocular pressure measures: a review and meta-analysis approach. *Surv Ophthalmol* 2000;**44**:367–408.

74. Kniestedt C, Lin S, Choe J, Bostrom A, Nee M, Stamper RL. Clinical comparison of contour and applanation tonometry and their relationship to pachymetry. *Arch Ophthalmol* 2005;**123**:1532–7.

75. Tonnu PA, Ho T, Newson T, El Sheikh A, Sharma K, White E, *et al.* The influence of central corneal thickness and age on intraocular pressure measured by pneumotonometry, non-contact tonometry, the Tono-Pen XL, and Goldmann applanation tonometry. *Br J Ophthalmol* 2005;**89**:851–4.

76. Dimasi DP, Burdon KP, Craig JE. The genetics of central corneal thickness. *Br J Ophthalmol* 2010;**94**:971–6.

77. European Glaucoma Prevention Study Group, Pfeiffer N, Torri V, Miglior S, Zeyen T, Adamsons I, *et al.* Central corneal thickness in the European Glaucoma Prevention Study. *Ophthalmology* 2007;**114**:454–9.

78. Leske MC. Predictors of long-term progression in the Early Manifest Glaucoma Trial. *Ophthalmology* 2007;**114**:1965–72.

79. Aghaian E, Choe JE, Lin S, Stamper RL. Central corneal thickness of Caucasians, Chinese, Hispanics, Filipinos, African Americans, and Japanese in a glaucoma clinic. *Ophthalmology* 2004;**111**:2211–19.

80. Bhatt N. A 10-year follow up of ocular hypertensive patients within the Bolton Corneal Thickness Study. Can measured factors predict prognostic outcomes? *Cont Lens Anterior Eye* 2008;**31**:147–53.

81. Fansi AA, Papamatheakis DG, Harasymowycz PJ. Racial variability of glaucoma risk factors between African Caribbeans and Caucasians in a Canadian urban screening population. *Can J Ophthalmol* 2009;**44**:576–81.

82. La Rosa FA, Gross RL, Orengo-Nania S. Central corneal thickness of Caucasians and African Americans in glaucomatous and nonglaucomatous populations. *Arch Ophthalmol* 2001;**119**:23–7.

83. Nemesure B, Wu SY, Hennis A, Leske MC, Barbados Eye Study Group. Corneal thickness and intraocular pressure in the Barbados eye studies. *Arch Ophthalmol* 2003;**121**:240–4.

84. Shimmyo M, Ross AJ, Moy A, Mostafavi R. Intraocular pressure, Goldmann applanation tension, corneal thickness, and corneal curvature in Caucasians, Asians, Hispanics, and African Americans. *Am J Ophthalmol* 2003;**136**:603–13.

85. Torres RJ, Jones E, Edmunds B, Becker T, Cioffi GA, Mansberger SL. Central corneal thickness in Northwestern American Indians/Alaskan Natives and comparison with White and African-American persons. *Am J Ophthalmol* 2008;**146**:747–51.

86. Friedman DS, Wilson MR, Liebmann JM, Fechtner RD, Weinreb RN. An evidence-based assessment of risk factors for the progression of ocular hypertension and glaucoma. *Am J Ophthalmol* 2004;**138**:31.

87. Landers JA, Hewitt AW, Dimasi DP, Charlesworth JC, Straga T, Mills RA, *et al.* Heritability of central corneal thickness in nuclear families. *Invest Ophthalmol Vis Sci* 2009;**50**:4087–90.

88. Toh T, Liew SH, MacKinnon JR, Hewitt AW, Poulsen JL, Spector TD, *et al.* Central corneal thickness is highly heritable: the twin eye studies. *Invest Ophthalmol Vis Sci* 2005;**46**:3718–22.

89. Zheng Y, Ge J, Huang G, Zhang J, Liu B, Hur YM, *et al.* Heritability of central corneal thickness in Chinese: the Guangzhou Twin Eye Study. *Invest Ophthalmol Vis Sci* 2008;**49**:4303–7.

90. Murdoch IE, Morris SS, Cousens SN. People and eyes: statistical approaches in ophthalmology. *Br J Ophthalmol* 1998;**82**:971–3.

91. Miglior S, Torri V, Zeyen T, Pfeiffer N, Vaz JC, Adamsons I, *et al.* Intercurrent factors associated with the development of open-angle glaucoma in the European Glaucoma Prevention Study. *Am J Ophthalmol* 2007;**144**:266–75.

92. Ekstrom C. Risk factors for incident open-angle glaucoma: a population-based 20-year follow-up study [published online ahead of print 9 July 2010]. *Acta Ophthalmol* 2010. DOI:10.1111/j.1755–3768.2010.01943.x.

93. Budenz DL, Anderson DR, Feuer WJ, Beiser JA, Schiffman J, Parrish RK, *et al.* Detection and prognostic significance of optic disc hemorrhages during the Ocular Hypertension Treatment Study. *Ophthalmology* 2006;**113**:2137–43.

94. Singh K, Shrivastava A. Intraocular pressure fluctuations: how much do they matter? *Curr Opin Ophthalmol* 2009;**20**:84–7.

95. Medeiros FA, Weinreb RN, Zangwill LM, Alencar LM, Sample PA, Vasile C, *et al.* Long-term intraocular pressure fluctuations and risk of conversion from ocular hypertension to glaucoma. *Ophthalmology* 2008;**115**:934–40.

96. Bengtsson B, Heijl A. Diurnal IOP fluctuation: not an independent risk factor for glaucomatous visual field loss in high-risk ocular hypertension. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:513–18.

97. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 2010;**172**:971–80.

98. Yan AT, Jong P, Yan RT, Tan M, Fitchett D, Chow CM, *et al.* Clinical trial-derived risk model may not generalize to real-world patients with acute coronary syndrome. *Am Heart J* 2004;**148**:1020–7.

99. Bhorade AM, Gordon MO, Wilson B, Weinreb RN, Kass MA, Ocular Hypertension Treatment Study Group. Variability of intraocular pressure measurements in observation participants in the ocular hypertension treatment study. *Ophthalmology* 2009;**116**:717–24.

100. Racette L, Wilson MR, Zangwill LM, Weinreb RN, Sample PA. Primary open-angle glaucoma in blacks: a review. *Surv Ophthalmol* 2003;**48**:295–313.

101. Mansberger SL. The probability of glaucoma from ocular hypertension determined by ophthalmologists in comparison to a risk calculator. *J Glaucoma* 2006;**15**:426–31.

102. Boland MV, Quigley HA, Lehmann HP. The impact of risk calculation on treatment recommendations made by glaucoma specialists in cases of ocular hypertension. *J Glaucoma* 2008;**17**:631–8.

103. Ehlers N, Hansen FK, Aasved H. Biometric correlations of corneal thickness. *Acta Ophthalmol* 1975;**53**:652–9.

104. Kaufmann C, Bachmann LM, Thiel MA. Comparison of dynamic contour tonometry with Goldmann applanation tonometry. *Invest Ophthalmol Vis Sci* 2004;**45**:3118–21.

105. Theofylaktopoulos I, Diestelhorst M, Krieglstein GK. Self-tonometry with the Ocuton S versus Goldmann tonometry. *Graefes Arch Clin Exp Ophthalmol* 1999;**237**:720–4.

106. Fernandes P, Diaz-Rey JA, Queiros A, Gonzalez-Meijome JM, Jorge J. Comparison of the ICare rebound tonometer with the Goldmann tonometer in a normal population. *Ophthalmic Physiol Opt* 2005;**25**:436–40.

107. Nakamura M, Darhad U, Tatsumi Y, Fujioka M, Kusuhara A, Maeda H, *et al.* Agreement of rebound tonometer in measuring intraocular pressure with three types of applanation tonometers. *Am J Ophthalmol* 2006;**142**:332–4.

108. Icare. *Icare tonometer: IOP measuring excellence*. Helsinki, Finland: Icare; 2010. URL: http://www.icaretonometer.com/uploads/brochures/ONE/Icare_ONE_tonometer_brochure_EN_low_res.pdf (accessed October 2010).

109. Luce DA. Determining in vivo biomechanical properties of the cornea with an ocular response analyzer. *J Cataract Refract Surg* 2005;**31**:156–62.

110. Kotecha A, Elsheikh A, Roberts CR, Zhu H, Garway-Heath DF. Corneal thickness- and age-related biomechanical properties of the cornea measured with the ocular response analyzer. *Invest Ophthalmol Vis Sci* 2006;**47**:5337–47.

111. Medeiros FA, Weinreb RN. Evaluation of the influence of corneal biomechanical properties on intraocular pressure measurements using the ocular response analyzer. *J Glaucoma* 2006;**15**:364–70.

112. Reichert. *Corneal biomechanics and accurate IOP in one easy-to-use instrument: the ocular response analyzer from Reichert*. Depew, NY: Reichert; 2009. URL: www.ocularresponseanalyzer.com/index.html (accessed October 2010).

113. Fresco BB. A new tonometer – the pressure phosphene tonometer: clinical comparison with Goldman tonometry. *Ophthalmology* 1998;**105**:2123–6.

114. Feltgen N, Leifert D, Funk J. Correlation between central corneal thickness, applanation tonometry, and direct intracameral IOP readings. *Br J Ophthalmol* 2001;**85**:85–7.

115. Abbasoglu OE, Bowman RW, Cavanagh HD, McCulley JP. Reliability of intraocular pressure measurements after myopic excimer photorefractive keratectomy. *Ophthalmology* 1998;**105**:2193–6.

116. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;**3**:25.

117. Craig JV, Lancaster GA, Williamson PR, Smyth RL. Temperature measured at the axilla compared with rectum in children and young people: systematic review. *BMJ* 2000;**320**:1174–8.

118. Williamson PR, Lancaster GA, Craig JV, Smyth RL. Meta-analysis of method comparison studies. *Stat Med* 2002;**21**:2013–25.

119. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Cont Clin Trial* 1986;**7**:177–88.

120. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;**8**:135–60.

121. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**327**:307–10.

122. Lafaut AS, Van Malderen L, Zeyen T. Is pulse synchronized pneumotonometry more reproducible than routine pneumotonometry and more in agreement with Goldmann applanation tonometry? *Eur J Ophthalmol* 2007;**17**:178–82.

123. Doyle A, Lachkar Y. Comparison of dynamic contour tonometry with Goldman applanation tonometry over a wide range of central corneal thickness. *J Glaucoma* 2005;**14**:288–92.

124. Rai S, Moster MR, Kesen M, Fontanarosa J, Spaeth GL, Steinmann WC, *et al.* Level of disagreement between Proview phosphene tonometer and Goldmann applanation tonometer intraocular pressure readings. *J Glaucoma* 2005;**14**:120–3.

125. Tai MC, Chen PL, Wu JN, Lu DW. Clinical evaluation of the intraocular pressure in patients with glaucoma or ocular hypertension by a self-assessable tonometer. *J Ocul Pharmacol Ther* 2005;**21**:55–61.

126. Sacu S, Vass C, Schemper M, Rainer G. Self-tonometry with the Ocuton S: evaluation of accuracy in glaucoma patients. *Acta Ophthalmol Scand* 2004;**82**:405–9.

127. Alvarez TL, Gollance SA, Thomas GA, Greene RJ, Marchetto PM, Moore EJ, *et al.* The Proview phosphene tonometer fails to measure ocular pressure accurately in clinical practice. *Ophthalmology* 2004;**111**:1077–85.

128. Chauhan BC, Henson DB. Clinical evaluation of the Non-Contact Tonometer Mark II. *Am J Optom Physiol Opt* 1988;**65**:751–6.

129. Kotecha A, White A, Schlottmann PG, Garway-Heath DF. Intraocular pressure measurement precision with the Goldmann applanation, dynamic contour, and ocular response analyzer tonometers. *Ophthalmology* 2010;**117**:730–7.

130. Morita T, Shoji N, Kamiya K, Hagishima M, Fujimara F, Shimizu K. Intraocular pressure measured by dynamic contour tonometer and ocular response analyzer in normal tension glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2010;**248**:73–7.

131. Punjabi OS, Ho HKV, Kniestedt C, Bostrom AG, Stamper RL, Lin SC. Intraocular pressure and ocular pulse amplitude comparisons in different types of glaucoma using dynamic contour tonometry. *Curr Eye Res* 2006;**31**:851–62.

132. Eichenbaum KD, Mezej M, Eichenbaum JW. Comparing dynamic contour tonometry to Goldmann and hand-held tonometry in normal, ocular hypertension, and glaucoma populations. *Ann Ophthalmol* 2007;**39**:41–9.

133. Gunvant P, Lievens CW, Newman JM III, Gerstner MD, Chang F, Haine CL. Evaluation of some factors affecting the agreement between the Proview eye pressure monitor and the Goldmann applanation tonometer measurements. *Clin Exp Optom* 2007;**90**:290–5.

134. Medeiros FA, Sample PA, Weinreb RN. Comparison of dynamic contour tonometry and Goldmann applanation tonometry in African American subjects. *Ophthalmology* 2007;**114**:658–65.

135. Pepose JS, Feigenbaum SK, Qazi MA, Sanderson JP, Roberts CJ. Changes in corneal biomechanics and intraocular pressure following LASIK using static, dynamic, and noncontact tonometry. *Am J Ophthalmol* 2007;**143**:39–47.

136. Morledge-Hampton SJ, Kwon RO, Krishna R, Debry PW, Willoughby TL. Comparison of Proview phosphene tonometry with Goldmann applanation tonometry. *Can J Ophthalmol* 2006;**41**:722–6.

137. Brigatti L, Maguluri S. Reproducibility of self-measured intraocular pressure with the phosphene tonometer in patients with ocular hypertension and early to advanced glaucoma. *J Glaucoma* 2005;**14**:36–9.

138. Li J, Herndon LW, Asrani SG, Stinnett S, Allingham RR. Clinical comparison of the Proview eye pressure monitor with the Goldmann applanation tonometer and the Tonopen. *Arch Ophthalmol* 2004;**122**:1117–21.

139. Popovich KS, Shields MB. A comparison of intraocular pressure measurements with the XPERT noncontact tonometer and Goldmann applanation tonometry. *J Glaucoma* 1997;**6**:44–6.

140. Kooner KS, Cooksey JC, Barron JB, Zimmerman TJ, Gupte RK, Wall JL. Tonometry comparison: Goldmann versus Tono-Pen. *Ann Ophthalmol* 1992;**24**:29–36.

141. Realini T, Weinreb RN, Hobbs G. Correlation of intraocular pressure measured with Goldmann and dynamic contour tonometry in normal and glaucomatous eyes. *J Glaucoma* 2009;**18**:119–23.

142. Sullivan-Mee M, Gerhardt G, Halverson KD, Qualls C. Repeatability and reproducibility for intraocular pressure measurement by dynamic contour, ocular response analyzer, and Goldmann applanation tonometry. *J Glaucoma* 2009;**18**:666–73.

143. Lanza M, Borrelli M, De Bernardo M, Filosa ML, Rosa N. Corneal parameters and difference between Goldmann applanation tonometry and dynamic contour tonometry in normal eyes. *J Glaucoma* 2008;**17**:460–4.

144. Salvetat ML, Zeppieri M, Tosoni C, Brusini P. Comparisons between Pascal dynamic contour tonometry, the TonoPen, and Goldmann applanation tonometry in patients with glaucoma. *Acta Ophthalmol Scand* 2007;**85**:272–9.

145. Brusini P, Salvetat ML, Zeppieri M, Tosoni C, Parisi L. Comparison of ICare tonometer with Goldmann applanation tonometer in glaucoma patients. *J Glaucoma* 2006;**15**:213–17.

146. Marchini G, Babighian S, Specchia L, Perfetti S. Evaluation of the new Ocuton S tonometer. *Acta Ophthalmol Scand* 2002;**80**:167–71.

147. Boles CB, Brogliatti B, Tonetto C, Renis E. The Pulsair-Keeler non-contact tonometer in self-tonometry: preliminary results. *Int Ophthalmol* 1992;**16**:295–7.

148. Ceruti P, Morbio R, Marraffa M, Marchini G. Comparison of Goldmann applanation tonometry and dynamic contour tonometry in healthy and glaucomatous eyes. *Eye* 2009;**23**:262–9.

149. Regine F. Validity and limitations of the Nidek NT-4000 non-contact tonometer: a clinical study. *Ophthalmic Physiol Opt* 2006;**26**:33–9.

150. Roszkowska AM, De Grazia L, Cirone M, Ferreri G. Comparison of Goldmann applanation tonometry and dynamic contour tonometry in the measurement of intraocular pressure in eyes with different corneal thicknesses. *Ophthalmologica* 2009;**223**:244–9.

151. Pelit A, Altan-Yaycioglu R, Pelit A, Akova YA. Effect of corneal thickness on intraocular pressure measurements with the Pascal dynamic contour, Canon TX-10 non-contact and Goldmann applanation tonometers in healthy subjects. *Clin Exp Optom* 2009;**92**:14–18.

152. Tonnu PA, Ho T, Sharma K, White E, Bunce C, Garway-Heath DF. A comparison of four methods of tonometry: method agreement and interobserver variability. *Br J Ophthalmol* 2005;**89**:847–50.

153. Bafa M, Lambrinakis I, Dayan M, Birch M. Clinical comparison of the measurement of the IOP with the ocular blood flow tonometer, the tonopen XL and the Goldmann applanation tonometer. *Acta Ophthalmol Scand* 2001;**79**:15–18.

154. Davies LN, Bartlett H, Mallen EA, Wolffsohn JS. Clinical evaluation of rebound tonometer. *Acta Ophthalmol Scand* 2006;**84**:206–9.

155. Kotecha A, White ET, Shewry JM, Garway-Heath DF. The relative effects of corneal thickness and age on Goldmann applanation tonometry and dynamic contour tonometry. *Br J Ophthalmol* 2005;**89**:1572–5.

156. Mackie SW, Jay JL, Ackerley R, Walsh G. *Clinical comparison of the Keeler Pulsair* 2000, American Optical MkII and Goldmann applanation tonometers. *Ophthalmic Physiol Opt* 1996;**16**:171–7.

157. Carrim ZI, Lavy TE. Goldmann tonometry versus the Tono-Pen XL for intraocular pressure measurement: an evaluation of the potential impact on clinical decision making in glaucoma. *Ophthalmic Physiol Opt* 2009;**29**:648–51.

158. Troost A, Yun SH, Specht K, Krummenauer F, Schwenn O. Transpalpebral tonometry: reliability and comparison with Goldmann applanation tonometry and palpation in healthy volunteers. *Br J Ophthalmol* 2005;**89**:280–3.

159. Herdener S, Hafizovic D, Pache M, Lautebach S, Funk J. Is the PASCAL-Tonometer suitable for measuring intraocular pressure in clinical routine? Long- and short-term reproducibility of dynamic contour tonometry. *Eur J Ophthalmol* 2008;**18**:39–43.

160. Ruokonen PC, Schwenteck T, Draeger J. Evaluation of the impedance tonometers TGDc-01 and iCare according to the international ocular tonometer standards ISO 8612. *Graefes Arch Clin Exp Ophthalmol* 2007;**245**:1259–65.

161. Barleon L, Hoffmann EM, Berres M, Pfeiffer N, Grus FH. Comparison of dynamic contour tonometry and Goldmann applanation tonometry in glaucoma patients and healthy subjects. *Am J Ophthalmol* 2006;**142**:583–90.

162. Troost A, Specht K, Krummenauer F, Yun SH, Schwenn O. Deviations between transpalpebral tonometry using TGDc-01 and Goldmann applanation tonometry depending on the IOP level. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:853–8.

163. Pache M, Wilmsmeyer S, Lautebach S, Funk J. Dynamic contour tonometry versus Goldmann applanation tonometry: a comparative study. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:763–7.

164. Meyer MW, Gockeln R, Hoy L, Meyer A, Erb C. Comparison of intraocular pressure measurements with the digital tonometer TGDc-01 'PRA' and the Goldmann applanation tonometer. *Ophthalmic Res* 2004;**36**:250–4.

165. Abraham LM, Epasinghe NC, Selva D, Casson R. Comparison of the ICare rebound tonometer with the Goldmann applanation tonometer by experienced and inexperienced tonometrists. *Eye* 2008;**22**:503–6.

166. Jackson C, Bullock J, Pitt M, Keogh J, Glasson W, Hirst L. Screening for glaucoma in a Brisbane general practice – the role of tonometry. *Aust N Z J Ophthalmol* 1995;**23**:173–8.

167. Pakrou N, Gray T, Mills R, Landers J, Craig J. Clinical comparison of the Icare tonometer and Goldmann applanation tonometry. *J Glaucoma* 2008;**17**:43–7.

168. Herse P, Hans A, Hall J, Langejans J, Markoulli M. The Proview Eye Pressure Monitor: influence of clinical factors on accuracy and agreement with the Goldmann tonometer. *Ophthalmic Physiol Opt* 2005;**25**:416–20.

169. Horowitz GS, Byles J, Lee J, D'Este C. Comparison of the Tono-Pen and Goldmann tonometer for measuring intraocular pressure in patients with glaucoma. *Clin Experiment Ophthalmol* 2004;**32**:584–9.

170. Wells JM. Clinical comparison of the Ocuton self-tonometer with Goldmann tonometry. *Clin Experiment Ophthalmol* 2003;**31**:17–22.

171. Liu L, Lei C, Li X, Dong J. Measurement of intraocular pressure after LASIK by dynamic contour tonometry. *J Huazhong Univ Sci Technol Med Sci* 2006;**26**:372–3.

172. Lam DS, Leung DY, Chiu TY, Fan DS, Cheung EY, Wong TY, *et al.* Pressure phosphene self-tonometry: a comparison with Goldmann tonometry in glaucoma patients. *Invest Ophthalmol Vis Sci* 2004;**45**:3131–6.

173. Cho P, Lui T. Comparison of the performance of the Nidek NT-2000 noncontact tonometer with the Keeler Pulsair 2000 and the Goldmann applanation tonometer. *Optom Vis Sci* 1997;**74**:51–8.

174. Lam A, Chen D, Chiu R, Chui WS. Comparison of IOP measurements between ORA and GAT in normal Chinese. *Optom Vis Sci* 2007;**84**:909–14.

175. Lam AK, Chan R, Lam CH. The validity of a new noncontact tonometer and its comparison with the Goldmann tonometer. *Optom Vis Sci* 2004;**81**:601–5.

176. Cheng ACK, Leung DYL, Cheung EYY, Fan DSP, Law RWK, Lam DSC. Intraocular pressure measurement in patients with previous LASIK surgery using pressure phosphene tonometer. *Clin Exp Ophthalmol* 2005;**33**:153–7.

177. Naruse S, Mori K, Kinoshita S. Evaluation of the pressure phosphene tonometer as a self-tonometer. *Ophthalmic Physiol Opt* 2005;**25**:421–8.

178. Naruse S, Mori K, Kojo M, Hieda O, Kinoshita S. Evaluation of intraocular pressure change after laser in situ keratomileusis using the pressure phosphene tonometer. *J Cataract Refract Surg* 2004;**30**:390–7.

179. Yaoeda K. Measurement of intraocular pressure using the NT-4000: a new non-contact tonometer equipped with pulse synchronous measurement function. *J Glaucoma* 2005;**14**:201–5.

180. Murase H, Sawada A, Mochizuki K, Yamamoto T. Effects of corneal thickness on intraocular pressure measured with three different tonometers. *Jpn J Ophthalmol* 2009;**53**:1–6.

181. Pourjavan S, Boelle PY, Detry-Morel M, De Potter P. Physiological diurnal variability and characteristics of the ocular pulse amplitude (OPA) with the dynamic contour tonometer (DCT-Pascal). *Int Ophthalmol* 2007;**27**:357–60.

182. Van de Velde T, Zeyen T. Reliability of the Nidek NT-1000 non contact tonometer. *Bull Soc Belge Ophtalmol* 1995;**255**:19–22.

183. Ehongo A, De Maertelaer V, Pourjavan S. Effect of topical corneal anaesthesia on ocular response analyzer parameters: pilot study. *Int Ophthalmol* 2009;**29**:325–8.

184. Vandewalle E, Vandenbroeck S, Stalmans I, Zeyen T. Comparison of ICare, dynamic contour tonometer, and ocular response analyzer with Goldmann applanation tonometer in patients with glaucoma. *Eur J Ophthalmol* 2009;**19**:783–9.

185. Kynigopoulos M, Schlote T, Kotecha A, Tzamalis A, Pajic B, Haefliger I. Repeatability of intraocular pressure and corneal biomechanical properties measurements by the ocular response analyser. *Klin Monatsbl Augenheilkd* 2008;**225**:357–60.

186. Duba I, Wirthlin AC. Dynamic contour tonometry for post-LASIK intraocular pressure measurements. *Klin Monatsbl Augenheilkd* 2004;**221**:347–50.

187. Kaufmann C, Bachmann LM, Thiel MA. Intraocular pressure measurements using dynamic contour tonometry after laser in situ keratomileusis. *Invest Ophthalmol Vis Sci* 2003;**44**:3790–4.

188. Schipper I, Senn P, Oyo-Szerenyi K, Peter R. Central and peripheral pressure measurements with the Goldmann tonometer and Tono-Pen after photorefractive keratectomy for myopia. *J Cataract Refract Surg* 2000;**26**:929–33.

189. Martinez-de-la-Casa JM, Garcia-Feijoo J, Vico E, Fernandez-Vidal A, Benitez del Castillo JM, Wasfi M, *et al.* Effect of corneal thickness on dynamic contour, rebound, and Goldmann tonometry. *Ophthalmology* 2006;**113**:2156–62.

190. Martinez-de-la-Casa JM, Garcia-Feijoo J, Castillo A, Garcia-Sanchez J. Reproducibility and clinical evaluation of rebound tonometry. *Invest Ophthalmol Vis Sci* 2005;**46**:4578–80.

191. Garcia Resua C, Giraldez Fernandez MJ, Cervino EA, Gonzalez PJ, Yebra-Pimentel E. Clinical evaluation of the new TGDc-01 'PRA' palpebral tonometer: comparison with contact and non-contact tonometry. *Optom Vis Sci* 2005;**82**:143–50.

192. Milla ED. Poor agreement between Goldmann and Pascal tonometry in eyes with extreme pachymetry. *Eye* 2009;**23**:536–42.

193. AlMubrad TM, Ogbuehi KC. The effect of repeated applanation on subsequent IOP measurements. *Clin Exp Optom* 2008;**91**:524–9.

194. Ogbuehi KC, AlMubrad TM. Accuracy and reliability of the Keeler Pulsair EasyEye non-contact tonometer. *Optom Vis Sci* 2008;**85**:61–6.

195. Ogbuehi KC. Assessment of the accuracy and reliability of the Topcon CT80 non-contact tonometer. *Clin Exp Optom* 2006;**89**:310–14.

196. Gupta V, Sony P, Agarwal HC, Sihota R, Sharma A. Inter-instrument agreement and influence of central corneal thickness on measurements with Goldmann, pneumotonometer and noncontact tonometer in glaucomatous eyes. *Indian J Ophthalmol* 2006;**54**:261–5.

197. Bandyopadhyay M, Raychaudhuri A, Lahiri SK, Schwartz EC, Myatt M, Johnson GJ. Comparison of Goldmann applanation tonometry with the Tonopen for measuring

intraocular pressure in a population-based glaucoma survey in rural West Bengal. *Ophthalmic Epidemiol* 2002;**9**:215–24.

198. Baskaran M. Comparison of Proview phosphene tonometer with the Goldmann applanation tonometer in myopic and non-myopic eyes. *Asian J Ophthalmol* 2006;**8**:57–61.

199. Jorge J, Gonzalez-Meijome JM, Diaz-Rey JA, Almeida JB, Ribeiro P, Parafita MA. Clinical performance of non-contact tonometry by Reichert AT550 in glaucomatous patients. *Ophthalmic Physiol Opt* 2003;**23**:503–6.

200. Jorge J, Diaz-Rey JA, Gonzalez-Meijome JM, Almeida JB, Parafita MA. Clinical performance of the Reichert AT550: a new non-contact tonometer. *Ophthalmic Physiol Opt* 2002;**22**:560–4.

201. Papastergiou GI, Kozobolis V, Siganos DS. Assessment of the pascal dynamic contour tonometer in measuring intraocular pressure in keratoconic eyes. *J Glaucoma* 2008;**17**:484–8.

202. Halkiadakis I, Patsea E, Chatzimichali K, Skouriotis S, Chalkidou S, Amariotakis G, *et al.* Comparison of dynamic contour tonometry with Goldmann applanation tonometry in glaucoma practice. *Acta Ophthalmol* 2009;**87**:323–8.

203. Detorakis E. Applanation tonometry versus dynamic contour tonometry in eyes treated with latanoprost. *J Glaucoma* 2010;**19**:194–8.

204. Ko YC, Liu CJ, Hsu WM. Varying effects of corneal thickness on intraocular pressure measurements with different tonometers. *Eye* 2005;**19**:327–32.

205. Shemesh G, Man O, Michaeli A, Varssano D, Lazar M. Pressure phosphene tonometry versus Goldmann applanation tonometry for measuring intraocular pressure before and after LASIK. *J Refract Surg* 2007;**23**:405–9.

206. Garzozi HJ, Chung HS, Lang Y, Kagemann L, Harris A. Intraocular pressure and photorefractive keratectomy: a comparison of three different tonometers. *Cornea* 2001;**20**:33–6.

207. van der Jagt LH, Jansonius NM. Three portable tonometers, the TGDc-01, the ICARE and the Tonopen XL, compared with each other and with Goldmann applanation tonometry. *Ophthalmic Physiol Opt* 2005;**25**:429–35.

208. Rietveld E, van den Bremer DA, Volker-Dieben HJ. Clinical evaluation of the pressure phosphene tonometer in patients with glaucoma. *Br J Ophthalmol* 2005;**89**:537–9.

209. Rehnman JB, Martin L. Comparison of rebound and applanation tonometry in the management of patients treated for glaucoma or ocular hypertension. *Ophthalmic Physiol Opt* 2008;**28**:382–6.

210. Johannesson G, Hallberg P, Eklund A, Linden C. Pascal, ICare and Goldmann applanation tonometry – a comparative study. *Acta Ophthalmol* 2008;**86**:614–21.

211. Sahin A, Niyaz L, Yildirim N. Comparison of the rebound tonometer with the Goldmann applanation tonometer in glaucoma patients. *Clin Experiment Ophthalmol* 2007;**35**:335–9.

212. Oncel B, Dinc UA, Orge F, Yalvac BI. Comparison of IOP measurement by ocular response analyzer, dynamic contour, Goldmann applanation, and noncontact tonometry. *Eur J Ophthalmol* 2009;**19**:936–41.

213. Vasconcelos De Moraes CG, Castro Reis AS, Sano ME, Barreira AK, Vessani RM, Susanna R. Intraocular pressure profile during the modified diurnal tension curve using Goldman applanation tonometry and dynamic contour tonometry. *J Ocul Biol Dis Infor* 2009;**2**:29–32.

214. Kirwan C, O'Keefe M. Measurement of intraocular pressure in LASIK and LASEK patients using the Reichert Ocular Response Analyzer and Goldmann applanation tonometry. *J Refract Surg* 2008;**24**:366–70.

215. Ku JY, Danesh-Meyer HV, Craig JP, Gamble GD, McGhee CN. Comparison of intraocular pressure measured by Pascal dynamic contour tonometry and Goldmann applanation tonometry. *Eye* 2006;**20**:191–8.

216. Christoffersen T, Fors T, Ringberg U, Holtedahl K. Tonometry in the general practice setting (I): Tono-Pen compared to Goldman applanation tonometry. *Acta Ophthalmol* 1993;**71**:103–8.

217. Hansen MK. Clinical comparison of the XPERT non-contact tonometer and the conventional Goldmann applanation tonometer. *Acta Ophthalmol Scand* 1995;**73**:176–80.

218. Fogagnolo P. Test-retest variability of intraocular pressure and ocular pulse amplitude for dynamic contour tonometry: a multicenter study. *Br J Ophthalmol* 2010;**94**:419–23.

219. Hallberg P, Eklund A, Bäcklund T, Lindén C. Clinical evaluation of applanation resonance tonometry: a comparison with Goldmann applanation tonometry. *J Glaucoma* 2007;**16**:88–93.

220. Wingert TA, Bassi CJ, McAlister WH, Galanis JC. Clinical evaluation of five portable tonometers. *J Am Optomet Assoc* 1995;**66**:670–4.

221. Troost R, Vogel A, Beck S, Schwenn O, Grus F, Pfeiffer N. Clinical comparison of two intraocular pressure measurement methods: SmartLens dynamic observing tonography versus Goldmann. *Graefes Arch Clin Exp Ophthalmol* 2001;**239**:889–92.

222. Bayraktar S, Bayraktar Z. Central corneal thickness and intraocular pressure relationship in eyes with and without previous LASIK: comparison of Goldmann applanation tonometer with pneumatonometer. *Eur J Ophthalmol* 2005;**15**:81–8.

223. Kass MA. Standardizing the measurement of intraocular pressure for clinical research: guidelines from the Eye Care Technology Forum. *Ophthalmology* 1996;**103**:183–5.

224. Goldmann H. Un nouveau tometre d'applanation. *Bull Soc Belge Ophtalmol* 1955;**67**:474–8.

225. Altman DG, Bland JM. Measurement in medicine – the analysis of method comparison studies. *Statistician* 1983;**32**:307–17.

226. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, *et al.* Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;**64**:96–106.

227. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;**149**:889–97.

228. Glasziou PA, Irwig L. *Evidence-based medical monitoring: from principles to practice*. Oxford: Blackwell Publishing; 2008.

229. Takahashi O, Glasziou PP, Perera R, Shimbo T, Suwa J, Hiramatsu S, *et al.* Lipid re-screening: what is the best measure and interval? *Heart* 2010;**96**:448–52.

230. Buclin T, Telenti A, Perera R, Csajka C, Furrer H, Aronson JK, *et al.* Development and validation of decision rules to guide frequency of monitoring CD4 cell count in HIV-1 infection before starting antiretroviral therapy. *PLoS One* 2011;**6**:e18578.

231. Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. *BMJ* 2005;**330**:644–8.

232. Glasziou PP, Irwig L, Heritier S, Simes RJ, Tonkin A; LIPID Study Investigators. Monitoring cholesterol levels: measurement error or true change? *Ann Intern Med* 2008;**148**:656–61.

233. Stevens RJ, Oke J, Perera R. Statistical models for the control phase of clinical monitoring. *Stat Methods Med Res* 2010;**19**:394–414.

234. Louviere J, Hensher DA, Swait JD. *Stated choice methods: analysis and application.* Cambridge: Cambridge University Press; 2000.

235. Ryan M, Gerard K, Amaya-Amaya M. *Using discrete choice experiments to value health and health care.* Dordrecht: Springer; 2008.

236. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. *J Health Serv Res Policy* 2007;**12**:25–30.

237. 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. *BMJ* 2005;**331**:134–6.

238. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In Bryman A, Burgess RG, editors. *Analysing qualitative data.* London: Routledge; 1994. pp. 173–94.

239. Donaldson C, Mason H, Shackley P. Contingent valuation in health care. In Jones AM, editor. *The Elgar companion to health economics.* Cheltenham: Edward Elgar; 2006. pp. 392–404.

240. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy* 2003;**2**:55–64.

241. Hernandez RA, Burr JM, Vale LD. Economic evaluation of screening for open-angle glaucoma. *Int J Technol Assess Health Care* 2008;**24**:203–11.

242. Lloyd AJ. The extent of patients' understanding of the risk of treatments. *Qual Health Care* 2001;**10**:i14–18.

243. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc* 2006;**13**:608–18.

244. Kuhfeld WF. *Experimental design: efficiency, coding, and choice design.* Cary, NC: SAS Institute; 2010. URL: http://support.sas.com/techsup/technote/mr2010c.pdf (accessed February 2011).

245. Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental design with marketing research applications. Cary, NC: SAS Institute; 2010. URL: http://support.sas.com/techsup/technote/mr2010d.pdf (accessed February 2011).

246. San Miguel F, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Econ* 2005;**14**:307–22.

247. Sen A. Internal consistency of choice. *Econometrica* 1993;**61**:495–521.

248. Bech M, Kjaer T, Lauridsen J. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Econ* 2011;**20**:273–86.

249. DeShazo JR, Fermo G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *J Environ Econ Manage* 2002;**44**:123–43.

250. Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing discrete choice experiments: do optimal designs come at a price? *J Consum Res* 2008;**35**:360–75.

251. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics* 2008;**26**:661–7.

252. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ* 2012;**21**:145–72.

253. Bliemer MCJ, Rose JM, Hensher DA. Efficient stated choice experiments for estimating nested logit models. *Transp Res Part B Methodol* 2009;**43**:19–35.

254. Rose JM, Bliemer MCJ, Hensher DA, Collins AT. Designing efficient stated choice experiments in the presence of reference alternatives. *Transp Res Part B Methodol* 2008;**42**:395–406.

255. Rose JM, Bliemer MCJ. Stated preference experimental design strategies. In Hensher DA, Button KJ, editors. *Handbook of transport modelling*. Oxford: Elsevier; 2008. pp. 151–80.

256. Kuhfeld WF. *Marketing research methods in SAS: experimental design, choice, conjoint, and graphical techniques*. Cary, NC: SAS Institute; 2005. URL: http://support.sas.com/techsup/technote/mr2010.pdf (accessed February 2011).

257. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ* 2007;**16**:827–40.

258. Olsen S. Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environ Resource Econ* 2009;**44**:591–610.

259. Windle J, Rolfe J. Comparing responses from web and paper-based collection modes in a choice modelling experiment. 2010. 54th Annual Australian Agricultural and Resource Economics Society Conference, Adelaide, February 2010.

260. Hammitt JK, Graham JD. Willingness to pay for health protection: inadequate sensitivity to probability? *J Risk Uncertain* 1999;**18**:33–62.

261. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal*. London: NICE; 2008. URL: www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed February 2011).

262. Department of Health. 'Now I feel tall': what a patient-led NHS feels like. London: Department of Health; 2005. URL: www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4124476.pdf (accessed February 2011).

263. Coulter A. What do patients and the public want from primary care? *BMJ* 2005;**331**:1199–201.

264. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*, 3rd edn. Oxford: Oxford University Press; 2005.

265. Coast J, Spencer IC, Smith L, Spry PGD. Comparing costs of monitoring glaucoma patients: hospital ophthalmologists versus community optometrists. *J Health Serv Res Policy* 1997;**2**:19–25.

266. Holtzer-Goor KM, Van Sprundel E, Lemij HG, Plochg T, Klazinga NS, Koopmanschap MA. Cost-effectiveness of monitoring glaucoma patients in shared care: an economic evaluation alongside a randomized controlled trial. *BMC Health Serv Res* 2010;**10**:312.

267. Government Actuary's Department. *Interim life tables 2007–09*. London: Government Actuary's Department; 2010. URL: http://www.ons.gov.uk/ons/rel/lifetables/interim_life_tables/index.html/ (accessed September 2010).

268. 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).

269. 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.

270. Burr J, Azuara-Blanco A, Avenell A. Medical versus surgical interventions for open angle glaucoma. *Cochrane Database Syst Rev* 2009; Issue 1, Art. No. CD004399.

271. Lichter PR, Musch DC, Gillespie BW, Guire KE, Janz NK, Wren PA, *et al.* Interim clinical outcomes in the Collaborative Initial Glaucoma Treatment Study comparing initial treatment randomized to medications or surgery. *Ophthalmology* 2001;**108**:1943–53.

272. NHS Information Services Division Scotland. *NHS costs book 2010: specialty costs.* Edinburgh: NHS Information Services Division Scotland; 2010. URL: www.isdscotland.org/isd/costs-book-detailed-tables.jsp?pContentID=3601&p_applic=CCC&p_service=Content.show& (accessed May 2011).

273. Department of Health. *General ophthalmic services: increases to NHS sight test fee 1 April 2010.* Communication from Department of Health to PCT chief executives and NHS trust chief executives dated 25 March 2010. URL: www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_114691.pdf (accessed May 2011).

274. British Medical Association and Royal Pharmaceutical Society of Great Britain. British national formulary. No. 61, March 2011. London: BMA and RPS; 2011. URL: http://bnf.org/bnf/index.htm (accessed March 2011).

275. Department of Health. *NHS reference costs 2009–2010.* London: Department of Health; 2010. URL: www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_111591 (accessed May 2011).

276. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998;**316**:736–41.

277. Burr J, Kilonzo M, Vale L, Ryan M. Developing a preference based glaucoma utility index using a discrete choice experiment. *Optom Vis Sci* 2007;**84**:797–808.

278. Olthoff CM, Schouten JS, van de Borne BW, Webers CA. Noncompliance with ocular hypotensive treatment in patients with glaucoma or ocular hypertension: an evidence-based review. *Ophthalmology* 2005;**112**:953–61.

279. Schwartz GF, Quigley HA. Adherence and persistence with glaucoma therapy. *Surv Ophthalmol* 2008;**53**(Suppl. 1):68.

280. Friedman D, Jonas J, Greenfield D, Goldberg I. Drug delivery. In Weinreb RN, Liebmann JM, editors. *Medical treatment of glaucoma.* Amsterdam: Kugler; 2010.

281. Vale LD. *The progressive economic evaluation of a surgical technique: a case study of surgical repair of inguinal hernia.* PhD thesis. Aberdeen: University of Aberdeen; 2005.

282. Reiger D. *Discrete choice experiments informing cost benefit analysis: a Bayesian approach with an application to genetic testing.* PhD thesis. Aberdeen: University of Aberdeen; 2008.

283. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health* 2009;**12**:739–49.

284. Briggs A. Handling uncertainty in economic evaluation and presenting the results. In Drummon M, McGuire A, editors. *Economic evaluation in health care: merging theory with practice.* Oxford: Oxford University Press; 2001.

285. Heijl A, Lindgren G, Olsson J, Asman P. Visual field interpretation with empiric probability maps. *Arch Ophthalmol* 1989;**107**:204–8.

286. Chauhan BC, Garway-Heath DF, Goni FJ, Rossetti L, Bengtsson B, Viswanathan AC, *et al.* Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol* 2008;**92**:569–73.

287. Asman P, Heijl A. Glaucoma Hemifield Test. Automated visual field evaluation. *Arch Ophthalmol* 1992;**110**:812–19.

288. Artes PH, Chauhan BC, Keltner JL, Cello KE, Johnson CA, Anderson DR, *et al.* Longitudinal and cross-sectional analyses of visual field progression in participants of the Ocular Hypertension Treatment Study. *Arch Ophthalmol* 2010;**128**:1528–32.

289. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol* 2008;**145**:343–53.

290. Giraud JM, Fenolland JR, May F, Hammam O, Sadat AM, Boumezrag AB, *et al.* [Analysis of a new visual field index, the VFI, in ocular hypertension and glaucoma] [French]. *J Fr Ophtalmol* 2010;**33**:2–9.

Appendix 1

Search strategies

Risk prediction (*Chapter 4*)

MEDLINE (1987 to January week 3 2011), MEDLINE In-Process & Other Non-Indexed Citations (28 January 2011) and EMBASE (1987 to week 4 2011)

Ovid Multifile Search – URL: https://shibboleth.ovid.com/

1. ocular hypertension/use mesz
2. intraocular hypertension/use emez
3. glaucoma, open-angle/use mesz
4. open-angle glaucoma/use emez
5. *glaucoma/
6. (glaucoma or poag or iop).ti.
7. (ocular adj3 (hypertension or pressure)).ti.
8. (intraocular adj3 (hypertension or pressure)).ti.
9. or/1–8
10. prediction/use emez
11. prognosis/
12. risk assessment/
13. predictor variable/use emez
14. disease progression/use mesz
15. disease course/use emez
16. (predict$or prognos?s or prognostic or progress$or risk).ti.
17. or/10–16
18. 9 and 17
19. model$.hw,tw.
20. algorithms/
21. algorithm$.tw.
22. tool?.hw,tw.
23. toolkit?.tw.
24. calculator?.tw.
25. or/19–24
26. 18 and 25
27. animal/or nonhuman/
28. human/
29. 27 not 28
30. 26 not 29
31. limit 30 to yr="1987 – 2011"
32. remove duplicates from 31

Science Citation Index (1987 to 27 January 2011), BIOSIS (1987 to 26 January 2011)

Web of Knowledge – URL: http://wok.mimas.ac.uk/

#1   TS=glaucoma

#2   TS=(ocular SAME hypertension)
#3   TS=(ocular SAME pressure)
#4   TS=(intraocular SAME hypertension)
#5   TS=(intraocular SAME pressure)
#6   #1 OR #2 OR #3 OR #4 OR #5 AND Document Type=(Article)
#7   #6 AND TI=(predict* or prognos* or progress* or risk) AND Document Type=(Article)
#8   #7 AND TS=(model* or algorithm* or tool* or calculat*) AND Document Type=(Article)

Database of Abstracts of Reviews of Effects and Health Technology Assessment database (January 2011)

NIHR Centre for Reviews and Dissemination – URL: www.crd.york.ac.uk/crdweb/

#1   MeSH Ocular Hypertension EXPLODE
#2   glaucoma
#3   poag OR iop
#4   "ocular pressure"
#5   "intraocular pressure"
#6   "intraocular hypertension"
#7   #1 OR #2 OR #3 OR #4 OR #5 OR #6
#8   MeSH Prognosis
#9   MeSH Disease Progression
#10 MeSH Risk Assessment
#11 predict* OR prognos* OR progress*
#12 #8 OR #9 OR #10 OR #11
#13 #7 AND #12
#14 model* OR algorithm* OR tool* OR calculat*
#15 #13 AND #14

Journal websites consulted (accessed 31 January 2011)

(model* OR tool*) and (predict* or progress*) and (glaucoma or ocular)

American Journal of Ophthalmology – URL: www.sciencedirect.com/science/journal/00029394
Archives of Ophthalmology – URL: http://archopht.ama-assn.org/
British Journal of Ophthalmology – URL: http://bjo.bmj.com/
Eye – URL: www.nature.com/eye/index.html
Graefe's Archive for Clinical and Experimental Ophthalmology – URL: www.springerlink.com/content/101559/
Investigative Ophthalmology and Visual Science – URL: www.iovs.org/
*Journal of Glaucom*a – URL: www.glaucomajournal.com
Ophthalmology – URL: www.sciencedirect.com/science/journal/0161642

Professional organisations consulted (accessed January 2011)

American Academy of Ophthalmology – URL: www.aao.org/
American Glaucoma Society – URL: www.glaucomaweb.org/
European Glaucoma Society – URL: www.eugs.org/eng/default.asp
Glaucoma Research Foundation – URL: www.glaucoma.org/
National Eye Institute – URL: www.nei.nih.gov/
Royal College of Ophthalmologists – URL: www.rcophth.ac.uk/

Systematic review of tonometers (*Chapter 5*)

MEDLINE (1987 to February week 2 2010), MEDLINE In-Process & Other Non-Indexed Citations (8 February 2010) and EMBASE (1987 to week 5 2010)

Ovid Multifile Search – URL: https://shibboleth.ovid.com/

1. Tonometry, Ocular/use mesz
2. Oculoplethysmography/use emez
3. Ocular Hypertension/di use mesz
4. Intraocular hypertension/di use emez
5. (ocular adj5 (tonometer$or tonometry or tonopen$)).tw.
6. ocular response analys$.tw.
7. ocular hypertension/or intraocular hypertension/or intraocular pressure/or exp glaucoma/
8. 7 and (tonometer$or tonometry or tonopen$).tw.
9. (oculoplethysmograph$or pneumotonomet$).tw.
10. or/1–6,8–9
11. "reproducibility of results"/use mesz
12. reproducibility/use emez
13. observer variation/
14. (variation or variability).tw.
15. (reliable or reliability).tw.
16. (reproducible or reproducibility).tw.
17. (agree? or agreement).tw.
18. or/11–17
19. 10 and 18
20. exp clinical trial/
21. randomized controlled trial.pt.
22. controlled clinical trial.pt.
23. randomization/use emez
24. randomi?ed.ab.
25. randomly.ab.
26. trial.ab.
27. groups.ab.
28. comparative study/use mesz
29. major clinical study/use emez
30. controlled study/use emez
31. clinical trial/use emez
32. (prospective$or retrospective$).tw. use mesz
33. (compare$or compara$).tw.
34. or/20–33
35. 19 and 34
36. humans/
37. 35 and 36
38. limit 37 to (english language and yr="1987 – 2010")

Science Citation Index (1987 to 8 February 2010) and BIOSIS (1987 to 8 February 2010)

Web of Knowledge – URL: http://wok.mimas.ac.uk/

#1　TS=((ocular OR intraocular) SAME (tonometer* or tonometry or tonopen*)) AND Language=(English) AND Document Type=(Article)

#2　TS=(oculoplethysmograph* or pneumotonomet*) AND Language=(English) AND Document Type=(Article)

#3　#1 or #2 AND Language=(English) AND Document Type=(Article)

#4　#3 AND TS =(variation or variability) AND Language=(English) AND Document Type=(Article)

#5　#3 AND TS=(reproducible or reproducibility) AND Language=(English) AND Document Type=(Article)

#6　#3 AND TS=agree* AND Language=(English) AND Document Type=(Article)

#7　#3 AND TS=(reliable or reliability) AND Language=(English) AND Document Type=(Article)

#8　#7 OR #6 OR #5 OR #4

#9　#8 AND TS=(random* OR trial* OR control* OR compar* OR prospective OR retrospective) AND Language=(English) AND Document Type=(Article)

The Cochrane Library (Issue 1 2010)

URL: www.thecochranelibrary.com/

#1　MeSH descriptor Tonometry, Ocular, this term only

#2　MeSH descriptor Ocular Hypertension, this term only with qualifier:DI

#3　MeSH descriptor Intraocular Pressure, this term only

#4　MeSH descriptor Ocular Hypertension, this term only

#5　MeSH descriptor Glaucoma explode all trees

#6　(#3 OR #4 OR #5)

#7　(tonometer* or tonometry or tonopen*)

#8　(#6 AND #7)

#9　(ocular NEAR (tonometer* or tonometry or tonopen*))

#10 (#1 OR #2 OR #8 OR #9)

#11 MeSH descriptor Reproducibility of Results, this term only

#12 MeSH descriptor Observer Variation explode all trees

#13 (variation OR variability) or (reliable OR reliability) or (reproduce OR reproducibility) or (agree*)

#14 (#11 OR #12 OR 13)

#15 (#10 AND #14)

Database of Abstracts of Reviews of Effects and Health Technology Assessment database (February 2010)

NIHR Centre for Reviews and Dissemination – URL: www.crd.york.ac.uk/crdweb/

#1　Tonometry, Ocular/

#2　Ocular Hypertension/

#3　Intraocular Pressure/

#4　Exp Glaucoma/

#5　#2 OR #3 OR #4

#6　#1 AND #5

World Health Organization International Clinical Trials Registry Platform (February 2010)

URL: www.who.int/ictrp/en/

Tonomet* AND ocular

ClinicalTrials.gov (February 2010)

URL: http://clinicaltrials.gov/ct/gui/c/r

Ocular AND (tomometer or tonometry)

Current Controlled Trials (February 2010)

URL: www.controlled-trials.com/

tonomet% AND ocular

Journal websites consulted (accessed 8 February 2010)

American Journal of Ophthalmology – URL: www.sciencedirect.com/science/journal/00029394
Archives of Ophthalmology – URL: http://archopht.ama-assn.org/
British Journal of Ophthalmology – URL: http://bjo.bmj.com/
Eye – URL: www.nature.com/eye/index.html
Graefe's Archive for Clinical and Experimental Ophthalmology – URL: www.springerlink.com/content/101559/
Investigative Ophthalmology and Visual Science – URL: www.iovs.org/
Journal of Glaucoma – URL: www.glaucomajournal.com
Ophthalmology – URL: www.sciencedirect.com/science/journal/0161642

Systematic review of economic evaluations (*Chapter 8*)

MEDLINE (1950 to October week 4 2010), MEDLINE In-Process & Other Non-Indexed Citations (25 October 2010) and EMBASE (1980 to week 41 2010)

Ovid Multifile search – URL: http://gateway.ovid.com

1. intraocular hypertension/use emez
2. exp ocular hypertension/use prmz
3. tonometry, ocular/use prmz
4. oculoplethysmography/use emez
5. iop.tw.
6. (ocular adj3 (pressure or hypertension)).tw.
7. (intraocular adj3 (hypertension or pressure)).tw.
8. ((applanation or contact or noncontact or dynamic or contour or rebound) adj1 tonomet$).tw.
9. or/1–8
10. disease surveillance/use emez
11. periodic medical examination/use emez
12. exp population surveillance/use prmz
13. (surveillance or monitor$).tw.
14. or/10–13
15. exp «costs and cost analysis»/use prmz
16. exp economic evaluation/use emez
17. economics/
18. health economics/use emez
19. exp economics,hospital/use prmz
20. exp economics,medical/use prmz
21. economics,pharmaceutical/use prmz
22. exp budgets/
23. exp models, economic/use prmz

24. exp decision theory/
25. ec.fs. use prmz
26. monte carlo method/
27. markov chains/
28. exp technology assessment, biomedical/
29. cost$.ti.
30. (cost$adj2 (effective$or utilit$or benefit$or minimis$)).ab.
31. economics model$.tw.
32. (economic$or pharmacoeconomic$).tw.
33. (price or prices or pricing).tw.
34. (financial or finance or finances or financed).tw.
35. (value adj2 (money or monetary)).tw.
36. markov$.tw.
37. monte carlo.tw.
38. (decision$adj2 (tree? or analy$or model$)).tw.
39. quality of life/
40. quality adjusted life year/
41. «Value of Life»/use prmz
42. exp health status indicators/use prmz
43. health status/use emez
44. sickness impact profile/use prmz
45. disability evaluation/use prmz
46. disability/use emez
47. activities of daily living/use prmz
48. exp daily life activity/use emez
49. cost utility analysis/use emez
50. rating scale/
51. questionnaires/
52. (quality adj1 life).tw.
53. quality adjusted life.tw.
54. disability adjusted life.tw.
55. (qaly? or qald? or qale? or qtime? or daly?).tw.
56. (euroqol or euro qol or eq5d or eq 5d).tw.
57. (hql or hqol or h qol or hrqol or hr qol).tw.
58. (hye or hyes).tw.
59. health$year$equivalent$.tw.
60. (hui or hui1 or hui2 or hui3).tw.
61. (health adj3 (utilit$or disutili$)).tw.
62. (health adj3 (state or status)).tw.
63. (sf36 or sf 36 or short form 36 or shortform 36).tw
64. (sf6 or sf 6 or short form 6 or shortform 6).tw.
65. (sf12 or sf 12 or short form 12 or shortform 12).tw.
66. (sf16 or sf 16 or short form 16 or shortform 16).tw.
67. (sf20 or sf 20 or short form 20 or shortform 20).tw.
68. willingness to pay.tw.
69. standard gamble.tw.
70. trade off.tw.
71. iop publishing.ab.
72. ocular hypertension/
73. or/15–70
74. 9 and 14 and 73

75. 74 not 71
76. 72 or 75
77. remove duplicates from 76

Health Management Information Consortium (HMIC) (September 2010)

Ovid gateway – URL: http://gateway.ovid.com

1. intraocular hypertension.tw.
2. iop.tw.
3. (ocular adj3 (pressure or hypertension)).tw.
4. tonometry.tw.
5. ((applanation or contact or noncontact or dynamic or contour or rebound) adj1 tonomet$).tw.
6. or/1–5

NHS Economic Evaluations Database (NEED) and Health Technology Assessment database (October 2010)

Centre for Reviews and Dissemination – URL: www.york.ac.uk/inst/crd/

#1   MeSH Ocular Hypertension EXPLODE 1
#2   ocular AND (hypertension OR pressure)
#3   iop
#4   (intraocular AND (hypertension OR pressure))
#5   (applanation OR contact OR noncontact OR dynamic OR contour OR rebound) AND tonomet*
#6   MeSH Population Surveillance EXPLODE 1 2 3 4
#7   surveillance OR monitor*
#8   #1 or #2 or #3 or #4
#9   #5 or #6 or #7
#10 #8 and #9

Appendix 2

List of excluded full-text studies

Risk prediction (*Chapter 4*)

Reference	Reason for exclusion
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	Includes not only patients with ocular OHT
Bengtsson B, Heijl A. Diurnal IOP fluctuation: not an independent risk factor for glaucomatous visual field loss in high-risk ocular hypertension. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:513–18	Conducted before 1987
Bhatt N. A 10-year follow up of ocular hypertensive patients within the Bolton Corneal Thickness Study. Can measured factors predict prognostic outcomes? *Cont Lens Anterior Eye* 2008;**31**:147–53	Retrospective case note review
Cohen SL, Lee PP, Herndon LW, Challa P, Overbury O, Allingham RR. Using the arteriolar Pressure Attenuation Index to predict ocular hypertension progression to open-angle glaucoma. *Arch Ophthalmol* 2003;**121**:33–8	Retrospective case note review
de Voogd S, Ikram MK, Wolfs RC, Jansonius NM, Witteman JC, Hofman AI, *et al*. Is diabetes mellitus a risk factor for open-angle glaucoma? The Rotterdam Study. *Ophthalmology* 2006;**113**:1827–31	Includes not only patients with OHT
de Voogd S, Wolfs RC, Jansonius NM, Witteman JC, Hofman A, de Jong PT. Atherosclerosis, C-reactive protein, and risk for open-angle glaucoma: the Rotterdam Study. *Invest Ophthalmol Vis Sci* 2006;**47**:3772–6	Includes not only patients with OHT
Diaz Aleman, VT, Fernandez-Baca Vaca G, Lozano Lopez V, Garcia Somalo M, Perera Sanz D, Gonzalez De La Rosa M. [Nomogram for ocular hypertension progression risk based on the ocular hypertension treatment study]. [Spanish]. *Arch Soc Esp Oftalmol* 2005;**80**:151–4	Assessment of model performance in any data set (derivation or validation) not available
Doshi V. Sociodemographic, family history, and lifestyle risk factors for open-angle glaucoma and ocular hypertension. The Los Angeles Latino Eye Study. *Ophthalmology* 2008;**115**:639–47	Includes not only patients with OHT
Ekstrom C, Alm A. Pseudoexfoliation as a risk factor for prevalent open-angle glaucoma. *Acta Ophthalmol* 2008;**86**:741–6	Conducted before 1987
European Glaucoma Prevention Study (EGPS) Group, Miglior S, Pfeiffer N, Torri V, Zeyen T, Cunha-Vaz J. Predictive factors for open-angle glaucoma among patients with ocular hypertension in the European Glaucoma Prevention Study. *Ophthalmology* 2007;**114**:3–9	No prediction equation
Gordon MO. Ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma. *Evid Based Eye Care* 2003;**4**:16–17	Not risk prediction model development or validation
Griffin BA, Elliott MN, Coleman AL, Cheng EM. Incorporating mortality risk into estimates of 5-year glaucoma risk. *Am J Ophthalmol* 2009;**148**:925–31	Not risk prediction model development or validation
Heeg GP, Jansonius NM. The groningen longitudinal glaucoma study III. The predictive value of frequency-doubling perimetry and GDx nerve fibre analyser test results for the development of glaucomatous visual field loss. *Eye* 2009;**23**:1647–52	Not risk prediction model development or validation
Hulsman CA, Houwing-Duistermaat JJ, Van Duijn CM, Wolfs R, Borger PH, Hofman A, *et al*. Family score as an indicator of genetic risk of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:1726–31	Includes not only patients with OHT
Jonas JB, Martus P, Horn FK, Jünemann A, Korth M, Budde WM. Predictive factors of the optic nerve head for development or progression of glaucomatous visual field loss. *Invest Ophthalmol Vis Sci* 2004;**45**:2613–18	Includes not only patients with OHT
Kang JH, Pasquale LR, Willett WC, Rosner BA, Egan KM, Faberowski N, *et al*. Dietary fat consumption and primary open-angle glaucoma. *Am J Clin Nutr* 2004;**79**:755–64	Includes not only patients with OHT
Kang JH, Willett WC, Rosner BA, Hankinson SE, Pasquale LR. Prospective study of alcohol consumption and the risk of primary open-angle glaucoma. *Ophthalmic Epidemiol* 2007;**14**:141–7	Includes not only patients with OHT
Kang JH, Willett WC, Rosner BA, Hankinson SE, Pasquale LR. Caffeine consumption and the risk of primary open-angle glaucoma: a prospective cohort study. *Invest Ophthalmol Vis Sci* 2008;**49**:1924–31	Includes not only patients with OHT
Lalezary M, Medeiros FA, Weinreb RN, Bowd C, Sample PA, Tavares IM, *et al*. Baseline optical coherence tomography predicts the development of glaucomatous change in glaucoma suspects. *Am J Ophthalmol* 2006;**142**:576–82	Includes not only patients with OHT

Reference	Reason for exclusion
Lee AJ, Rochtchina E, Wang JJ, Healey PR, Mitchell P. Open-angle glaucoma and systemic thyroid disease in an older population: the Blue Mountains Eye Study. *Eye* 2004;**18**:600–8	Includes not only patients with OHT
Leske MC, Wu SY, Honkanen R, Nemesure B, Schachat A, Hyman L, *et al*. Nine-year incidence of open-angle glaucoma in the Barbados Eye Studies. *Ophthalmology* 2007;**114**:1058–64	Includes not only patients with OHT
Leske MC, Wu SY, Hennis A, Honkanen R, Nemesure B, BESs Study Group. Risk factors for incident open-angle glaucoma. The Barbados Eye Studies. *Ophthalmology* 2008;**115**:85–93	Includes not only patients with OHT
Levine RA, McManus P, Margol V. Asymmetries and visual field summaries as predictors of glaucoma in the Ocular Hypertension Treatment Study. *Invest Ophthalmol Vis Sci* 2006;**47**:3896–903	No prediction equation
Mansberger SL. A risk calculator to determine the probability of glaucoma. *J Glaucoma* 2004;**13**:345–7	Not risk prediction model development or validation
Medeiros FA. Frequency doubling technology perimetry abnormalities as predictors of glaucomatous visual field loss. *Am J Ophthalmol* 2004;**137**:863–71	Not risk prediction model development or validation
Medeiros FA, Weinreb RN. Estimating the risk of developing glaucoma. *Open Ophthalmol J* 2009;**3**:50–3	Not risk prediction model development or validation
Medeiros FA, Sample PA, Weinreb RN. Agreement between stereophotographic and confocal scanning laser ophthalmoscopy measurements of cup/disc ratio: effect on a predictive model for glaucoma development. *J Glaucoma* 2007;**16**:209–14	Not risk prediction model development or validation
Medeiros FA, Weinreb RN, Zangwill LM, Alencar LM, Sample PA, Vasile C, *et al*. Long-term intraocular pressure fluctuations and risk of conversion from ocular hypertension to glaucoma. *Ophthalmology* 2008;**115**:934–40	No prediction equation
Medeiros FA, Alencar LM, Zangwill LM, Bowd C, Vizzeri G, Sample PA, *et al*. Detection of progressive retinal nerve fiber layer loss in glaucoma using scanning laser polarimetry with variable corneal compensation. *Invest Ophthalmol Vis Sci* 2009;**50**:1675–81	Not risk prediction model development or validation
Medeiros FA, Alencar LM, Zangwill LM, Sample PA, Weinreb RN. The relationship between intraocular pressure and progressive retinal nerve fiber layer loss in glaucoma. *Am J Ophthalmol* 2009;**116**:1125–33	Not risk prediction model development or validation
Medeiros FA, Alencar LM, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Prediction of functional loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol* 2009;**127**:1250–6	Includes not only patients with OHT
Miglior S, Torri V, Zeyen T, Pfeiffer N, Vaz JC, Adamsons I, *et al*. Intercurrent factors associated with the development of open-angle glaucoma in the European glaucoma prevention study. *Am J Ophthalmol* 2007;**144**:266–75	No prediction equation
Nemesure B, Wu SY, Hennis A, Leske MC, Barbados Eye Studies Group. Factors related to the 4-year risk of high intraocular pressure: the Barbados Eye Studies. *Arch Ophthalmol* 2003;**121**:856–62	Includes not only patients with OHT
Nordmann JP. [High blood pressure-low ocular pressure: the parallels. The ophthalmologist's point of view] [French]. *J Fr Ophtalmol* 2007;**30**:S14–7	Not risk prediction model development or validation
Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, Hankinson SE. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology* 2006;**113**:1081–6	Includes not only patients with OHT
Pasquale LR, Rosner BA, Hankinson SE, Kang JH. Attributes of female reproductive aging and their relation to primary open-angle glaucoma: a prospective study. *J Glaucoma* 2007;**16**:598–605	Includes not only patients with OHT
Peeters AM, Webers CAB, Prins MHM, Hendrikse FM, Schouten JSAG. The clinical impact of 2 different strategies for initiating therapy in patients with ocular hypertension. *J Glaucoma* 2011;**20**:30–6	Not risk prediction model development or validation
Philippin H. Ten-year results: detection of long-term progressive optic disc changes with confocal laser tomography. *Graefe Arch Clin Exp Ophthalmol* 2006;**244**:460–4	Includes not only patients with OHT
Polo Llorens V, Pablo Julvez LE, Pinilla Lozano I, Larrosa Poves JM, Ruiz MO, Honrubia Lopez FM. [Short-wavelength automated perimetry (SWAP) in patients with suspected glaucoma (II): correlation with a probabilistic multifactorial model of risk for developing glaucomatous damage]. [Spanish]. *Arch Soc Esp Oftalmol* 2000;**75**:97–102	Not risk prediction model development or validation
Polo V, Abecia E, Pablo LE, Pinilla I, Larrosa JM, Honrubia FM. Functional and structural measurements in a multifactorial glaucoma risk model. *Acta Ophthalmol Scand* 2001;**79**:10–14	Not risk prediction model development or validation
Rasker MTE, van den Enden A, Bakker D, Hoyng PFJ. Rate of visual field loss in progressive glaucoma. *Arch Ophthalmol* 2000;**118**:481–8	Not risk prediction model development or validation
Strouthidis NGM, Gardiner SKP, Owen VMF, Zuniga CM, Garway-Heath DFM. Predicting progression to glaucoma in ocular hypertensive patients. *J Glaucoma* 2010;**19**:304–9	No IOP in the model
Tezel G, Kolker AE, Kass MA, Wax MB, Gordon M, Siegmund KD. Parapapillary chorioretinal atrophy in patients with ocular hypertension: I. An evaluation as a predictive factor for the development of glaucomatous damage. *Arch Ophthalmol* 1997;**115**:1503–8	No prediction equation

Reference	Reason for exclusion
Walland M. Use of the Medmont Automated Perimeter with the Scoring Tool for Assessing Risk (STAR) II glaucoma risk calculator. *Clin Exp Ophthalmol* 2008;**36**:899–900	Not risk prediction model development or validation
Weinreb RN, Medeiros F. Risk assessment for glaucoma. *Open Ophthalmol J* 2009;**3**:30–1	Not risk prediction model development or validation
Weinreb RN, Friedman DS, Fechtner RD, Cioffi GA, Coleman AL, Girkin CA. Risk assessment in the management of patients with ocular hypertension. *Am J Ophthalmol* 2004;**138**:458–67	Not risk prediction model development or validation
Weinreb RN, Zangwill LM, Jain S, Becerra LM, Dirkes K, Piltz-Seymour JR, *et al.* Predicting the onset of glaucoma: the confocal scanning laser ophthalmoscopy ancillary study to the Ocular Hypertension Treatment Study. *Ophthalmology* 2010;**117**:1674–83	Not risk prediction model development or validation
Zangwill LM, Weinreb RN, Beiser JA, Berry CC, Cioffi GA, Coleman AL, *et al.* Baseline topographic optic disc measurements are associated with the development of primary open-angle glaucoma: the Confocal Scanning Laser Ophthalmoscopy Ancillary Study to the Ocular Hypertension Treatment Study. *Arch Ophthalmol* 2005;**123**:1188–97	No prediction equation
Zenker HJ. Use of multivariate analysis in evaluating the individual risk of glaucomatous visual field damage. *Chibret Int J Ophthalmol* 1989;**6**:38–44	Includes not only patients with OHT

Agreement and reliability of tonometers (*Chapter 5*)

Study details	Technology	No. of participants	Reasons for exclusion
Losch A, Scheuerle A, Rupp V, Auffarth G, Becker M. Transpalpebral measurement of intraocular pressure using the TGDc-01 tonometer versus standard Goldmann applanation tonometry. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:313–16	Transpalpebral	109	Age range uncertain
Grolman B, Myers KJ, Lalle P. How reliable is the Goldmann tonometer as a standard? *J Am Optomet Assoc* 1990;**61**:857–62	NCT	620 eyes	Age range uncertain
Babalola OE, Kehinde AV, Iloegbunam AC, Akinbinu T, Moghalu C, Onuoha I. A comparison of the Goldmann applanation and non-contact (Keeler Pulsair EasyEye) tonometers and the effect of central corneal thickness in indigenous African eyes. *Ophthal Physiol Optics* 2009;**29**:182–8	NCT	88	Age – included those < 16 years
Barreto J Jr, Babic M, Vessani RM, Susanna R Jr. Dynamic contour tonometry and Goldman applanation tonometry in eyes with keratoconus. *Clinics* 2006;**61**:511–14	DCT	12	Age – included those < 16 years
Brencher HL, Kohl P, Reinke AR, Yolton RL. Clinical comparison of air-puff and Goldmann tonometers. *J Am Optomet Assoc* 1991;**62**:395–402	NCT	227	Age – included those < 16 years
Danesh-Meyer HV, Niederer R, Gaskin BJ, Gamble G. Comparison of the Proview pressure phosphene tonometer performed by the patient and examiner with the Goldmann applanation tonometer. *Clin Exp Ophthalmol* 2004;**32**:29–32	PPT	96	Age – included those < 16 years
Iliev ME, Goldblum D, Katsoulis K, Amstutz C, Frueh B. Comparison of rebound tonometry with Goldmann applanation tonometry and correlation with central corneal thickness. *Br J Ophthalmol* 2006;**90**:833–5	RT	55	Age – included those < 16 years
Kao SF, Lichter PR, Bergstrom TJ, Rowe S, Musch DC. Clinical comparison of the Oculab Tono-Pen to the Goldmann applanation tonometer. *Ophthalmology* 1987;**94**:1541–4	TonoPen	103	Age – included those < 16 years
Kretz G, Demailly P. X-PERT NCT advanced logic tonometer valuation. *Int Ophthalmol* 1992;**16**:287–90	NCT	60	Age – included those < 16 years
Lawson-Kopp W, DeJong A, Yudcovitch L, Williams S, Kohl P, Yolton RL. Clinical evaluation of the Keeler Pulsair 3000 non-contact tonometer. *Optometry* 2002;**73**:81–90	NCT	113	Age – included those < 16 years
Levy Y, Zadok D, Glovinsky Y, Krakowski D, Nemet P. Tono-Pen versus Goldmann tonometry after excimer laser photorefractive keratectomy. *J Cataract Refract Surg* 1999;**25**:486–91	TonoPen	18	Age – included those < 16 years
Moseley MJ, Thompson JR, Deutsch J, Misson GP, Naylor G, Tan-Yee A, *et al.* Comparison of the Keeler Pulsair 2000 non-contact tonometer with Goldmann applanation. *Eye* 1993;**7**:127–30	NCT	160	Age – included those < 16 years

Study details	Technology	No. of participants	Reasons for exclusion
Munkwitz S, Elkarmouty A, Hoffmann EM, Pfeiffer N, Thieme H. Comparison of the iCare rebound tonometer and the Goldmann applanation tonometer over a wide IOP range. *Graefes Arch Clin Exp Ophthalmol* 2008;**246**:875–9	RT	75	Age – included those < 16 years
Pearce CD, Kohl P, Yolton RL. Clinical evaluation of the Keeler PULSAIR 2000 tonometer. *J Am Optomet Assoc* 1992;**63**:106–10	NCT	190	Age – included those < 16 years
Poostchi A, Mitchell R, Nicholas S, Purdie G, Wells A. The iCare rebound tonometer: comparisons with Goldmann tonometry, and influence of central corneal thickness. *Clin Exp Ophthalmol* 2009;**37**:687–91	RT	100	Age – included those < 16 years
Recep OF, Hasiripi H, Cagil N, Sarikatipoglu H. Relation between corneal thickness and intraocular pressure measurement by noncontact and applanation tonometry. *J Cataract Refract Surg* 2001;**27**:1787–91	NCT	60	Age – included those < 16 years
Sandner D, Bohm A, Kostov S, Pillunat L. Measurement of the intraocular pressure with the 'transpalpebral tonometer' TGDc-01 in comparison with applanation tonometry. *Graefes Arch Clin Exp Ophthalmol* 2005;**243**:563–9	Transpalpebral	103	Age – included those < 16 years
Toker MI, Vural A, Erdogan H, Topalkara A, Arici MK. Central corneal thickness and Diaton transpalpebral tonometry. *Graefes Arch Clin Exp Ophthalmol* 2008;**246**:881–9	Transpalpebral	81	Age – included those < 16 years
Yan QC, He YH, Bai QH, Di Y, Wang XL. Comparison of the intraocular pressure value measured with Goldmann applanation tonometer and non-contact tonometer. *Int J Ophthalmol* 2006;**6**:537–9	NCT	265	Age – included those < 16 years
Kniestedt C, Lin S, Choe J, Bostrom A, Nee M, Stamper RL. Clinical comparison of contour and applanation tonometry and their relationship to pachymetry. *Arch Ophthalmol* 2005;**123**:1532–7	DCT	258	Age – included those < 16 years
Kniestedt C, Lin S, Choe J, Nee M, Bostrom A, Sturmer J, *et al.* Correlation between intraocular pressure, central corneal thickness, stage of glaucoma, and demographic patient data: prospective analysis of biophysical parameters in tertiary glaucoma practice populations. *J Glaucoma* 2006;**15**:91–7	DCT	406	Age – included those < 16 years
Paranhos A Jr, Paranhos FR, Prata JA Jr, Omi CA, Mello PA, Shields MB. Influence of keratometric readings on comparative intraocular pressure measurements with Goldmann, Tono-Pen, and noncontact tonometers. *J Glaucoma* 2000;**9**:219–23	TonoPen, NCT	202	Age – included those < 16 years
Bonomi L, Baravelli S, Cobbe C, Tomazzoli L. Evaluation of Keeler Pulsair non-contact tonometry: reliability and reproducibility. *Graefes Arch Clin Exp Ophthalmol* 1991;**229**:210–12	NCT	414 eyes	Age – not reported
Denis P, Nordmann JP, Bertin V, Gayraud JM, Laroche L, Saraux H. Evaluation of the Tono-Pen 2 and the X-Pert noncontact tonometers in cataract surgery. *Ophthalmologica* 1993;**207**:155–61	NCT	45	Age – not reported
Draeger J, Rumberger E, Dauper J, Deutsch C. Microprocessor controlled tonometry. *Eye* 1989;**3**:738–42	NCT, TonoPen	99	Age – not reported
Fresco BB. A new tonometer – the pressure phosphene tonometer: clinical comparison with Goldman tonometry. *Ophthalmology* 1998;**105**:2123–6	PPT	100	Age – not reported
Geyer O, Mayron Y, Loewenstein A, Neudorfer M, Rothkoff L, Lazar M. Tono-Pen tonometry in normal and in post-keratoplasty eyes. *Br J Ophthalmol* 1992;**76**:538–40	TonoPen	82	Age – not reported
Hollo G, Follmann P, Pap G. A clinical evaluation of XPERT NCT (Reichert) for glaucoma screening by optometrists. *Int Ophthalmol* 1992;**16**:291–3	NCT	300 eyes	Age – not reported
Iester M, Mermoud A, Achache F, Roy S. New Tonopen XL: comparison with the Goldmann tonometer. *Eye* 2001;**15**:1–8	TonoPen	104	Age – not reported
Koopmans SA, Kooijman AC, van Rij G, Eisses J, de Groot-Woltjer J. Clinical evaluation of two non-contact tonometers. *Doc Ophthalmol* 1991;**78**: 259–63	NCT	33	Age – not reported
Lam AK, Lam CH, Chan R. The validity of a digital eyelid tonometer (TGDc-01) and its comparison with Goldmann applanation tonometry – a pilot study. *Ophthalmol Physiol Optic* 2005;**25**:205–10	Transpalpebral	40	Age – not reported
Midelfart A, Wigers A. Clinical comparison of the ProTon and Tono-Pen tonometers with the Goldmann applanation tonometer. *Br J Ophthalmol* 1994;**78**:895–8	TonoPen	53	Age – not reported

Study details	Technology	No. of participants	Reasons for exclusion
Mok KH, Wong CS, Lee VW. Tono-Pen tonometer and corneal thickness. *Eye* 1999;**13**:35–7	TonoPen	48	Age – not reported
Parker VA, Herrtage J, Sarkies NJ. Clinical comparison of the Keeler Pulsair 3000 with Goldmann applanation tonometry. *Br J Ophthalmol* 2001;**85**:1303–4	NCT	150 eyes	Age – not reported
Schneider E, Grehn F. Intraocular pressure measurement – comparison of dynamic contour tonometry and Goldmann applanation tonometry. *J Glaucoma* 2006;**15**:2–6	DCT	100	Age – not reported
Theofylaktopoulos I, Diestelhorst M, Krieglstein GK. Self-tonometry with the Ocuton S versus Goldmann tonometry. *Graefes Arch Clin Exp Ophthalmol* 1999;**237**:720–4	Ocuton S	100	Age – not reported
Verdoorn C, Deutman AF. Clinical evaluation of the Topcon CT10 tonometer. *Int Ophthalmol* 1988;**12**:223–5	NCT	115	Age – not reported
Frenkel RE, Hong YJ, Shin DH. Comparison of the Tono-Pen to the Goldmann applanation tonometer. *Arch Ophthalmol* 1988;**106**:750–3	TonoPen	142	Age – not reported
Minckler DS, Baerveldt G, Heuer DK, Quillen-Thomas B, Walonker AF, Weiner J. Clinical evaluation of the Oculab Tono-Pen. *Am J Ophthalmol* 1987;**104**:168–73	TonoPen	130	Age – not reported
Wingert TA, Bassi CJ, McAlister WH, Galanis JC. Clinical evaluation of five portable tonometers. *J Am Optom Assoc* 1995;**66**:670–4	TonoPen, Perkins, NCT, Schiotz	31	Age – not reported
Armstrong TA. Evaluation of the Tono-Pen and the Pulsair tonometers. *Am J Ophthalmol* 1990;**109**:716–20	TonoPen, NCT	130	Age – not reported
Bhan A, Browning AC, Shah S, Hamilton R, Dave D, Dua HS. Effect of corneal thickness on intraocular pressure measurements with the pneumotonometer, Goldmann applanation tonometer, and Tono-Pen. *Invest Ophthalmol Vis Sci* 2002;**43**:1389–92	Ocular blood flow, TonoPen	94	Age – not reported
Boothe WA, Lee DA, Panek WC, Pettit TH. The Tono-Pen. A manometric and clinical study. *Arch Ophthalmol* 1988;**106**:1214–17	TonoPen, pneumotonometer	50	Age – not reported
Siganos DS, Papastergiou GI, Moedas C. Assessment of the Pascal dynamic contour tonometer in monitoring intraocular pressure in unoperated eyes and eyes after LASIK. *J Cataract Refract Surg* 2004;**30**:746–51	DCT, NCT	60	Age – not reported
Farrar SM, Miller KN, Shields MB, Stoup CM. An evaluation of the Tono-Pen for the measurement of diurnal intraocular pressure. *Am J Ophthalmol* 1989;**107**:411–16	TonoPen	30	Age – not reported
Baddon AC, Osborne SF, Quah SA, Batterbury M, Wong D. Comparison of Luneau SA disposable and Goldmann applanation tonometer readings. *Eye* 2007;**21**:789–92	Disposable prism	79	Comparison of disposable prims
Bhatnagar A, Gupta AK. Disposable devices for measuring intraocular pressure: a clinical study to assess their accuracy. *Eye* 2005;**19**:752–4	Disposable prism	40	Comparison of disposable prims
Desai SP, Sivakumar S, Fryers PT. Evaluation of a disposable prism for applanation tonometry. *Eye* 2001;**15**:3–82	Disposable prism	100	Comparison of disposable prims
Goel S, Chua C, Dong B, Butcher M, Ahfat F, Hindi SK, *et al*. Comparison between standard Goldmann applanation prism and disposable applanation prism in tonometry. *Eye* 2004;**18**:175–8	Disposable prism	42	Comparison of disposable prims
Kim P, Lertsumitkul S, Clark M, Gardner L, Macken P. Accuracy of the Tonosafe disposable tonometer head compared to the Goldmann tonometer alone. *Clin Exp Ophthalmol* 2004;**32**:364–7	Disposable prism	69	Comparison of disposable prims
Lotti R, Frau B, Cerruti S, Trillo C, Traverso CE. Reliability of applanation tonometry readings obtained with a disposable latex cap. *Ophthalmologica* 1999;**213**:277–80	Disposable prisms	80	Comparison of disposable prisms
Maino AP, Uddin HJ, Tullo AB. A comparison of clinical performance between disposable and Goldmann tonometers. *Eye* 2006;**20**:574–8	Disposable prisms	153	Comparison of disposable prisms
Maldonado MJ, Rodriguez-Galietero A, Cano-Parra J, Menezo JL, Diaz-Llopis M. Goldmann applanation tonometry using sterile disposable silicone tonometer shields. *Ophthalmology* 1996;**103**:815–21	Disposable prisms	60	Comparison of disposable prisms
Salvi SM, Sivakumar S, Sidiki SS. Use of disposable prism tonometry in routine clinical practice. *Eye* 2005;**19**:743–6	Disposable prisms	200	Comparison of disposable prisms

Study details	Technology	No. of participants	Reasons for exclusion
Lim L, Ng TP, Tan DT. Accurate intraocular pressure measurement in contact lens wearers with normal pressures. *CLAO J* 1997;**23**:130–3	TonoPen	20	Contact lens wearers
Choi WJ, Kim JW, Tchah H, Jin YH, Kim YJ. Non-contact tonometry: an ideal method for mass screening. *Korean J Ophthalmol* 1990;**4**:30–3	NCT	76	Corneal disease
Hamilton KE, Pye DC, Kao L, Pham N, Tran AQ. The effect of corneal edema on dynamic contour and Goldmann tonometry. *Optom Vis Sci* 2008;**85**:451–6	DCT	30	Corneal disease
Chew GS, Sanderson GF, Molteno AC. The pressure phosphene tonometer – a clinical evaluation. *Eye* 2005;**19**:683–5	PPT	100	Corneal disease
Lisle C, Ehlers N. A clinical comparison of the Xpert non-contact tonometer with the Goldmann applanation tonometer after penetrating keratoplasty. *Acta Ophthalmol Scand* 2000;**78**:211–15	NCT	42	Corneal surgery
Alfaro DV, Tran VT. A clinical comparison of the Oculab Tono-Pen with the Goldmann applanation tonometer in eyes filled with silicone oil. *Retina* 1991;**11**:219–20	TonoPen	21	Eyes filled with silicone
Stodtmeister R, Kron M, Gaus W. IOP measurement and central corneal thickness. *Br J Ophthalmol* 2002;**86**:120–1	NA	NA	Letter
Leung DY, Lam DS. Clinical comparison of the Proview eye pressure monitor with the Goldmann applanation tonometer and the TonoPen. *Arch Ophthalmol* 2005;**123**:578–9	NA	NA	Letter (related to included paper)
Bayraktar S, Bayraktar Z. Central corneal thickness and intraocular pressure relationship in eyes with and without previous LASIK: comparison of Goldmann applanation tonometer with pneumatonometer. *Eur J Ophthalmol* 2005;**15**:81–8	Pneumotonometer	234	No candidate tonometer
Abbasoglu OE, Bowman RW, Cavanagh HD, McCulley JP. Reliability of intraocular pressure measurements after myopic excimer photorefractive keratectomy. *Ophthalmology* 1998;**105**:2193–6	Ocular blood flow	21	No candidate tonometer
Dielemans I, Vingerling JR, Hofman A, Grobbee DE, de Jong PT. Reliability of intraocular pressure measurement with the Goldmann applanation tonometer in epidemiological studies. *Graefes Arch Clin Exp Ophthalmol* 1994;**232**:141–4	GAT	62	No candidate tonometer
Duch S, Serra A, Castanera J, Abos R, Quintana M. Tonometry after laser in situ keratomileusis treatment. *J Glaucoma* 2001;**10**:261–5	Ocular blood flow	60	No candidate tonometer
Gardiner SK, Demirel S. Assessment of patient opinions of different clinical tests used in the management of glaucoma. *Ophthalmology* 2008;**115**:2127–31	None	101	No candidate tonometer
Hallberg P, Eklund A, Bäcklund T, Lindén C. Clinical evaluation of applanation resonance tonometry: a comparison with Goldmann applanation tonometry. *J Glaucoma* 2007;**16**:88–93	Applanation resonance tonometry	78	No candidate tonometer
Nardi M, Bartolomei MP, Romani A, Barca L. Intraocular pressure changes in secondary positions of gaze in normal subjects and in restrictive ocular motility disorders. *Graefes Arch Clin Exp Ophthalmol* 1988;**226**:8–10	GAT	85	No candidate tonometer
Singh RP, Goldberg I, Graham SL, Sharma A, Mohsin M. Central corneal thickness, tonometry, and ocular dimensions in glaucoma and ocular hypertension. *J Glaucoma* 2001;**10**:206–10	Ocular blood flow	87	No candidate tonometer
Vaegan, RAM, Sanderson GF. Glaucoma affects steady state VEP contrast thresholds before psychophysics. *Optom Vis Sci* 2008;**85**:547–58	None	42	No candidate tonometer
Wang X, Shen J, McCulley JP, Bowman RW, Petroll WM, Cavanagh HD. Intraocular pressure measurement after hyperopic LASIK. *CLAO J* 2002;**28**:136–9	Ocular blood flow	15	No candidate tonometer
Zadok D, Tran DB, Twa M, Carpenter M, Schanzlin DJ. Pneumotonometry versus Goldmann tonometry after laser in situ keratomileusis for myopia. *J Cataract Refract Surg* 1999;**25**:1344–8	Pneumotonometer	17	No candidate tonometer
Schmidt KG, von Ruckmann A, Pillunat LE. Topical carbonic anhydrase inhibition increases ocular pulse amplitude in high tension primary open angle glaucoma. *Br J Ophthalmol* 1998;**82**:758–62	Ocular blood flow	33	No candidate tonometer

Study details	Technology	No. of participants	Reasons for exclusion
Georgopoulos GT, Diestelhorst M, Fisher R, Ruokonen P, Krieglstein GK. The short-term effect of latanoprost on intraocular pressure and pulsatile ocular blood flow. *Acta Ophthalmol Scand* 2002;**80**:54–8	Ocular blood flow	24	No candidate tonometer
Wolfs RC, Klaver CC, Vingerling JR, Grobbee DE, Hofman A, de Jong PT. Distribution of central corneal thickness and its association with intraocular pressure: the Rotterdam Study. *Am J Ophthalmol* 1997;**123**:767–72	None	395	No candidate tonometer
Shah S, Chatterjee A, Mathai M, Kelly SP, Kwartz J, Henson D, *et al*. Relationship between corneal thickness and measured intraocular pressure in a general ophthalmology clinic. *Ophthalmology* 1999;**106**:2154–60	None	868 (eyes)	No candidate tonometer
Troost R, Vogel A, Beck S, Schwenn O, Grus F, Pfeiffer N. Clinical comparison of two intraocular pressure measurement methods: SmartLens dynamic observing tonography versus Goldmann. *Graefes Arch Clin Exp Ophthalmol* 2001;**239**:889–92	SmartLens	80	No candidate tonometer
Munger R, Hodge WG, Mintsioulis G, Agapitos PJ, Jackson WB, Damji KF. Correction of intraocular pressure for changes in central corneal thickness following photorefractive keratectomy. *Can J Ophthalmol* 1998;**33**:159–65	None	318	No candidate tonometer
Hoffmann EM, Grus FH, Pfeiffer N. Intraocular pressure and ocular pulse amplitude using dynamic contour tonometry and contact lens tonometry. *BMC Ophthalmol* 2004;**4**:4	DCT, SmartLens	19	No candidate tonometer (prototype version of DCT)
Kontiola A. A new electromechanical method for measuring intraocular pressure. *Doca Ophthalmol* 1996;**93**:265–76	Electromechanical method	Pig eyes	Non-human population
Erdurmus M, Totan Y, Hepsen IF, Yagci R. Comparison of dynamic contour tonometry and noncontact tonometry in ocular hypertension and glaucoma. *Eye* 2009;**23**:663–8	DCT, NCT	104	Reference tonometer not used
Garcia-Resua C, Gonzalez-Meijome JM, Gilino J, Yebra-Pimentel E. Accuracy of the new ICare rebound tonometer vs. other portable tonometers in healthy eyes. *Optom Vis Sci* 2006;**83**:102–7	RT, TonoPen	65	Reference tonometer not used
Moreno-Montanes J, Garcia N, Fernandez-Hortelano A, Garcia-Layana A. Rebound tonometer compared with Goldmann tonometer in normal and pathologic corneas. *Cornea* 2007;**26**:427–30	RT	150	Reference tonometer not used
Moreno-Montanes J, Maldonado MJ, Garcia N, Mendiluce L, Garcia-Gomez PJ, Segui-Gomez M. Reproducibility and clinical relevance of the ocular response analyzer in nonoperated eyes: corneal biomechanical and tonometric implications. *Invest Ophthalmol Vis Sci* 2008;**49**:968–74	ORA	262	Reference tonometer not used
Kontiola A, Puska P. Measuring intraocular pressure with the Pulsair 3000 and Rebound tonometers in elderly patients without an anesthetic. *Graefes Arch Clin Exp Ophthalmol* 2004;**242**:3–7	RT	131	Reference tonometer not used
Herdener S, Pache M, Lautebach S, Funk J. Dynamic contour tonometry (DCT) versus Goldmann applanation tonometry (GAT) – a comparison of agreement and reproducibility. *Graefes Arch Clin Exp Ophthalmol* 2007;**245**:1027–30	DCT	23	Same study sample as in an included study

NA, not available.

Economic evaluations of ocular hypertension surveillance programmes (*Chapter 8*)

Reference	Reason for exclusion
Althin R, Grima DT, Dhawan R, Bernard LM. Considerations in developing model-based economic evaluations of glaucoma treatment. *J Glaucoma* 2006;**15**:541–7	Not surveillance strategy
Kymes SM, Kass MA, Anderson DR, Miller JP, Gordon MO, Ocular Hypertension Treatment Study Group (OHTS). Management of ocular hypertension: a cost-effectiveness approach from the Ocular Hypertension Treatment Study. *Am J Ophthalmol* 2006;**141**:997–1008	Not surveillance strategy
Peeters A, Schouten JS, Webers CA, Prins MH, Hendrikse F, Severens JL. Cost-effectiveness of early detection and treatment of ocular hypertension and primary open-angle glaucoma by the ophthalmologist. *Eye* 2008;**22**:354–62	Not surveillance strategy
Stewart WC, Stewart JA, Nasser QJ, Mychaskiw MA. Cost-effectiveness of treating ocular hypertension. *Ophthalmology* 2008;**115**:94–8	Not surveillance strategy
Tuck MW, Crick RP. The cost-effectiveness of various modes of screening for primary open angle glaucoma. *Ophthal Epidemiol* 1997;**4**:3–17	Not surveillance strategy
Coast J, Spencer IC, Smith L, Spry PG. Comparing costs of monitoring glaucoma patients: hospital ophthalmologists versus community optometrists. *J Health Serv Res Policy* 1997;**2**:19–25	Not OHT patients
Kobelt-Nguyen G, Gerdtham UG, Alm A. Costs of treating primary open-angle glaucoma and ocular hypertension: a retrospective, observational two-year chart review of newly diagnosed patients in Sweden and the United States. *J Glaucoma* 1998;**7**:95–104	Description of costs only
Spencer IC, Spry PG, Gray SF, Baker IA, Menage MJ, Easty DL, *et al.* The Bristol Shared Care Glaucoma Study: study design. *Ophthal Physiol Optics* 1995;**15**:391–4	Model design without results
van Gestel A, Severens JL, Webers CA, Beckers HJ, Jansonius NM, Schouten JS. Modeling complex treatment strategies: construction and validation of a discrete event simulation model for glaucoma. *Value Health* 2010;**13**:358–67	Model design without results
Gottlieb LK, Schwartz B, Pauker SG. Glaucoma screening. A cost-effectiveness analysis. *Surv Ophthalmol* 1983;**28**:206–26	Review article
Rosenbaum LJ. Management of patients with ocular hypertension: a cost-effectiveness approach from the Ocular Hypertension Treatment Study. *Am J Ophthalmol* 2007;**143**:191–2	Letter

Appendix 3

Characteristics of the included studies (*Chapter 5*)

Study details	No. participants/ eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Abraham 2008,[165] Australia	100/100	72 (NR); 55/45; NR; mixed population	15.4 (3.5)	RT (Icare)	RT: 2×4 (six readings); GAT: 2	RT: two ophthalmologists; GAT: two ophthalmologists
AlMubrad 2008,[193] Saudi Arabia	68/68	NR; 24/41; NR; not known to have OHT or glaucoma	14.8 (2.9)	NCT (CT80)	NCT: 4; GAT: 3	NCT: NR; GAT: NR
Alvarez 2004,[127] USA	137/137	NR; NR; NR; mixed population	15.7 (5.2)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 5; GAT: 2	Transpalpebral: patient; GAT: one trained clinician
Bafa 2001,[153] UK	99/99	68 (NR); 63/36; NR; not known to have OHT or glaucoma	NR	TonoPen (XL)	TonoPen: 4; GAT: NR	TonoPen: one examiner; GAT: one examiner
Bandyopadhyay 2002,[197] India	203/203	61 (9); 92/111; NR; not known to have OHT and glaucoma	12.9 (3.5)	TonoPen (XL)	TonoPen: NR; GAT: NR	TonoPen: NR; GAT: NR
Barleon 2006,[161] Germany	107/197	NR; NR; NR; mixed population	18.2 (7.9)	DCT (Pascal)	DCT: 3; GAT: 3	DCT: one ophthalmologist; GAT: NR
Baskaran 2006,[198] India	72/132	36 (15); 34/38; NR; not known to have OHT or glaucoma	14.8 (3.4)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 1; GAT: 3	Transpalpebral: patient; GAT: one ophthalmologist
Boles 1992,[147] Italy	45/90	NR; NR; NR; not known to have OHT or glaucoma	16.2 (3.4)	NCT (Keeler Pulsair)	NCT: 4; GAT: NR	NCT: patient; GAT: NR
Brigatti 2005,[137] USA	36/72	63 (16); NR; NR; mixed population	18.2 (NR)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: 2×3 readings	Transpalpebral: patient; GAT: one physician
Brusini 2006,[145] Italy	178/178	67 (13); NR; NR; glaucoma	19.4 (5.4)	RT (Icare)	RT: 1 (six readings); GAT: 3	RT: one ophthalmologist; GAT: NR
Carrim 2009,[157] UK	62/124	73 (12); 31/31; NR; OHT including glaucoma suspect and glaucoma	18.3 (4.8)	TonoPen (XL)	TonoPen: 4; GAT: NR	TonoPen: one of four nurses; GAT: one senior ophthalmologist
Ceruti 2009,[148] Italy	300/300	NR; 144/156; NR; mixed population	17.1 (5.7)	DCT (Pascal)	DCT: 3; GAT: NR	DCT: one ophthalmologist; GAT: NR
Chauhan 1988,[128] UK	102/102	44[a] (NR); 54/48; NR; mixed population	13.8 (3.5)	NCT (Mark II, Reichert)	NCT: 4; GAT: 4	NCT: NR; GAT: NR
Cheng 2005,[176] China	62/123	34 (6); NR; NR; not known to have OHT or glaucoma	15.0 (1.9)	NCT, PPT (Proview Eye Pressure Monitor)	NCT: 3; transpalpebral: 3; GAT: NR	NCT: NR; transpalpebral: one experienced optometrist; GAT: NR
Cho 1997,[173] China	22/44	22 (1); 7/15; NR; not known to have OHT or glaucoma	14.2 (2.7)	NCT (Nidek NT-2000 with AutoPuffControl)	NCT: 3; GAT: 3	NCT: one experienced optometrist; GAT: NR
Christoffersen 1993,[216] Norway	132/258	42 (NR); 100/32; NR; not known to have OHT or glaucoma	13.3 (NR)	TonoPen (XL)	TonoPen: 1; GAT: 2	TonoPen: two GPs; GAT: one ophthalmologist
Davies 2006,[154] UK	42/42	21 (3); 24/18; NR; not known to have OHT or glaucoma	16.2 (3.4)	RT (Icare)	RT: 1 (six readings); GAT: 2	RT: NR; GAT: one of two optometrists
Detorakis 2010,[203] Greece	55/55	NR; 15/40; white; not known to have OHT and POAG under treatment	NR	DCT (Pascal)	DCT: 3; GAT: NR	DCT: NR; GAT: NR

Study details	No. participants/eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Doyle 2005,[123] France	75/75	NR; NR; NR; mixed population	NR	DCT (Pascal)	DCT: 2; GAT: 2	DCT: two ophthalmologists; GAT: NR
Duba 2004,[186] Switzerland	50/100[b]	35 (NR); 12/8; NR; not known to have OHT or glaucoma	15.1 (2.2)	DCT (Pascal)	DCT: NR; GAT: NR	DCT: NR; GAT: NR
Ehongo 2009,[183] Belgium	23/23	40 (12); 12/11; NR; not known to have OHT or glaucoma	13.1 (2.5)	ORA	ORA: 2 (four readings); GAT: 2	ORA: NR; GAT: NR
Eichenbaum 2007,[132] USA	267/267	NR; NR; NR; mixed population	17.4 (3.1)	Perkins (hand-held applanation tonometer), DCT (Pascal)	Perkins: 1, DCT: 1; GAT: 1	DCT and Perkins: one examiner; GAT: NR
Fernandes 2005,[106] Portugal	46/46	23 (3); 32/14; NR; not known to have OHT or glaucoma	13.4 (2.3)	RT (Icare)	RT: 2 (six readings); GAT: 2	RT: one experienced optometrist; GAT: one experienced ophthalmologist
Fogagnolo 2010,[218] Italy/Spain	350/350	60 (13); 177/173; NR; mixed population	14.9 (3.6)	DCT (Pascal)	DCT: 2; GAT: 3	DCT: one experienced investigator per site; GAT: NR
Garcia Resua 2005,[191] UK	58/58	21 (3); 40/18; NR; not known to have OHT or glaucoma	10.8 (3.0)	Transpalpebral tonometer (TGDc-01 PRA), NCT (Reichert Xpert), TonoPen (XL), Perkins	Transpalpebral: 5; NCT: 3; TonoPen: 4, Perkins: 3; GAT: 3	NCT, transpalpebral, TonoPen and Perkins: one optometrist; GAT: NR
Garzozi 2001,[206] Israel	97/149	31 (8); 54/43; NR; not known to have OHT or glaucoma	13.4 (1.5)	TonoPen (XL), NCT	TonoPen: NR; NCT: NR; GAT: NR	NCT and TonoPen: NR; GAT: NR
Gunvant 2007,[133] USA	119/238	36 (NR); 52/67; NR; mixed population	14.5 (3.1)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: NR	Transpalpebral: patient; GAT: one trained examiner
Gupta 2006,[196] India	65/130	48 (10); NR; NR; glaucoma	NR	NCT (Nidek Nt-3000)	NCT: 3; GAT: 3	NCT: one experienced optometrist; GAT: one ophthalmologist
Halkiadakis 2009,[202] Athens	100/100	67 (10); 59/41; white (100); mixed population	17.9 (4.7)	DCT (Pascal)	DCT: 2; GAT: 2	DCT: two ophthalmologists; GAT: one ophthalmologist
Hansen 1995,[217] Denmark	113/113	64 (NR); 62/51; NR; not known to have OHT or glaucoma	16.5 (6.1)	NCT (Xpert)	NCT: 3; GAT: 1	NCT: one examiner; GAT: one examiner
Herdener 2008,[159] Germany	50/50	36 (17); 29/21; NR; not known to have OHT or glaucoma	14.7 (1.9)	DCT (Pascal)	DCT: 4; GAT: 4	DCT: one experienced ophthalmologist; GAT: NR
Herse 2005,[168] Australia	107/203	25 (NR); NR; NR; not known to have OHT or glaucoma	13.9 (2.2)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: 3	Transpalpebral: patient; GAT: one experienced clinician
Horowitz 2004,[169] Australia	138/276[b]	70 (NR); NR; NR; under treatment for glaucoma	16.1 (6.2)	TonoPen (XL)	TonoPen: 2; GAT: 2	TonoPen: two ophthalmologists; GAT: NR

Study details	No. participants/ eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Jackson 1995,[166] Australia	73/146	NR; 38/34; NR; not known to have OHT or glaucoma	NR	TonoPen, Perkins	TonoPen: NR; Perkins: NR; GAT: 3	TonoPen and Perkins: three general practitioners; GAT: one ophthalmologist
Johannesson 2008,[210] Sweden	76/150	58 (NR); 42/34; NR; mixed population	19.4 (6.7)	RT (Icare), DCT (Pascal)	RT: 1 (six readings); DCT: 6; GAT: 6	DCT and RT: one student; GAT: one clinician
Jorge 2002,[200] Portugal	35/35	25 (2); 22/13; NR; not known to have OHT or glaucoma	12.3 (2.2)	NCT (Reichert AT550), Perkins	NCT: 3, Perkins: 2; GAT: 2	NCT and Perkins: possibly a consultant ophthalmologist but paper is unclear; GAT: one trained practitioner
Jorge 2003,[199] Portugal	22/44	NR; 11/11; NR; glaucoma	19.0 (2.8)	NCT (Reichert AT550)	NCT: 3; GAT: 2	NCT: one examiner; GAT: one experienced examiner
Kaufmann 2003,[187] USA	31/62	39 (NR); 10/15; NR; not known to have OHT and POAG under treatment	NR	DCT (Pascal)	DCT: 3; GAT: 3	DCT: one examiner; GAT: one examiner
Kaufmann 2004,[104] USA	150/228	NR; NR; NR; not known to have OHT and POAG under treatment	NR	DCT (Pascal)	DCT: 3; GAT: 3	DCT: four examiners; GAT: four examiners
Kirwan 2008,[214] Ireland	67/125	NR; 45/22; NR; myopia	NR	ORA	ORA: 4, GAT: 4	ORA: NR; GAT: NR
Ko 2005,[204] Taiwan	170/170	67 (NR); 74/96; Asian; mixed population	15.6 (3.7)	NCT (Topcon CT60)	NCT: 3; GAT: 3	NCT: one ophthalmologist; GAT: one ophthalmologist
Kooner 1992,[140] USA	76/152	NR; NR; white (4), black (148); not known to have OHT or glaucoma	15.8 (3.8)	TonoPen (Oculab)	TonoPen: 2; GAT: 2	TonoPen: three ophthalmologists; GAT: three ophthalmologists
Kotecha 2005,[155] UK	130/130	61 (13); 52/78; NR; known OHT with and without topical ocular hypotensive therapy	19 (4.6)	DCT (Pascal)	DCT: 3; GAT 2	DCT: one of two technicians; GAT: one clinician
Kotecha 2006,[110] UK	144/144	58 (15); NR; NR; mixed population	17.3 (4.1)	ORA	ORA: 3; GAT: 2	ORA: one optometrist; GAT: NR
Kotecha 2010,[129] UK	100/100	52 (18); 38/62; NR; mixed population	15.9 (5.0)	DCT (Pascal), ORA	DCT: 3; ORA: 3; GAT: 2	DCT and ORA: one optometrist and two others (non-ophthalmologists); GAT: NR
Ku 2006,[215] New Zealand	116/116	64 (16); 65/41; NR; mixed population	15.9 (3.8)	DCT (Pascal)	DCT: 1; GAT: 1	DCT: NR; GAT: two experienced examiners
Kynigopoulos 2008,[185] Switzerland	50/50	36 (11); 37/12; Caucasian; not known to have OHT	15.8 (2.8)	ORA	ORA: 4; GAT: NR	ORA: one experienced technician; GAT: NR
Lafaut 2007,[122] Belgium	78/78	54 (NR); NR; NR; mixed population	17.1 (4.4)	NCT (Nidek NT-4000 with APC)	NCT: 3; GAT: 2	NCT: NR; GAT: NR
Lam 2004,[175] China	31/31	NR; 14/17; NR; not known to have OHT or glaucoma	16.0 (2.3)	NCT (Nidek NT-4000 PP)	NCT: 3; GAT: 3	NCT: possibly a consultant ophthalmologist but paper is unclear; GAT: NR
Lam 2004,[172] China	102/102	54 (13); NR; NR; mixed population	19.1 (4.7)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: 3	Transpalpebral: patient; GAT: two of the investigators
Lam 2007,[174] China	125/125	23 (3); 61/64; Chinese; not known to have OHT or glaucoma	5.9 (2.4)	ORA	ORA: 4, GAT: 3	ORA: possibly a consultant ophthalmologist but paper is unclear; GAT: one practitioner

Study details	No. participants/ eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Lanza 2008,[143] Italy	118/118	42 (14); NR; NR; not known to have OHT or glaucoma	15.0 (2.4)	DCT (Pascal)	DCT: NR; GAT: 2	DCT: NR; GAT: two examiners
Li 2004,[138] USA	91/91	NR; 48/38; NR; mixed population	17.2 (NR)	PPT (Proview Eye Pressure Monitor), TonoPen	Transpalpebral: 2; TonoPen: 2; GAT: 2	TonoPen: one consultant ophthalmologist; transpalpebral: patient; GAT: skilled clinicians
Liu 2006,[171] China	30/60[b]	27 (6); 20/10; NR; not known to have OHT or glaucoma	16.1 (1.3)	DCT (Pascal)	DCT: 3; GAT: 3	DCT: one examiner; GAT: one examiner
Mackie 1996,[156] UK	45/89	NR; 21/24; NR; POAG required treatment	18.0 (4.9)	NCT (Keeler Pulsair 2000)	NCT: 4; GAT: 2	NCT: one of two optometrists; GAT: one of three ophthalmologists
Marchini 2002,[146] Italy	20/20	61 (13); 10/10; NR; mixed population	18.1 (4.7)	Ocuton S	Ocuton S: 3; GAT: NR	Ocuton S: two operators; GAT: NR
Martinez-de-la-Casa 2005,[190] Spain	85/147	64 (16); NR; NR; mixed population	18.1 (5.4)	RT (Icare)	RT: 3; GAT: 3	RT: one experienced ophthalmologist; GAT: NR
Martinez-de-la-Casa 2006,[189] Spain	90/146	61 (14); NR; NR; mixed population	19.1 (5.1)	RT (Icare), DCT (Pascal)	RT: 2 (six readings), DCT: 2; GAT: 2	DCT and RT: one experienced ophthalmologist; GAT: NR
Medeiros 2006,[111] USA	78/153	54 (15); 43/35; white (31), African American (44), Asian (2), Hispanic (1); not known to have OHT or glaucoma	15.3 (3.3)	ORA	ORA: 2, GAT: 2	ORA: one trained technician; GAT: one of three investigators
Medeiros 2007,[134] USA	94[b]/176	58 (13); NR; African American; mixed population	16.7 (4.1)	DCT (Pascal)	DCT: 2; GAT: 2	DCT: one experienced ophthalmologist; GAT: NR
Meyer 2004,[164] Germany	88/176	49 (16); 55/33; NR; not known to have OHT or glaucoma	14.9 (3.2)	Transpalpebral tonometer (TGDc-01 PRA)	DCT: 1; GAT: 1	Transpalpebral: NR; GAT: NR
Milla 2009,[192] Spain	100/100	44 (11); NR; NR; not known to have OHT or glaucoma	14.8 (NR)	DCT (Pascal)	DCT: 3; GAT: 3	DCT: NR; GAT: NR
Morita 2010,[130] Japan	32/60	NR; 23/9; NR; mixed population	NR	DCT (Pascal), ORA	DCT: 3; ORA: 3; GAT: NR	DCT and ORA: one ophthalmologist; GAT: NR
Morledge-Hampton 2006,[136] USA	30/30	NR; 22/8; white (14), black (15), Hispanic (1); mixed population	15.8 (4.4)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 2; GAT: 3	Transpalpebral: one ophthalmologist; GAT: NR
Murase 2009,[180] Japan	69/69	59 (13); 35/34; NR; not known to have OHT or glaucoma	13.0 (2.5)	NCT (Reichert PT-100)	PNCT: 3; GAT: 3	NCT: three examiners; GAT: NR
Nakamura 2006,[107] Japan	45/45	52 (20); 30/15; NR; mixed population	14.1 (4.2)	RT (Icare), NCT (Canon TX-10), TonoPen (XL)	RT: 3; NCT: 3; TonoPen: 3; GAT: 3	NCT: orthoptist (number NR); RT: one ophthalmologist; GAT: one experienced observer
Naruse 2004,[178] Japan	22/36	37 (9); 15/7; NR; not known to have OHT or glaucoma	14.7 (2.1)	NCT (Nidek NT-3000), PPT (Proview Eye Pressure Monitor)	NCT: 3, transpalpebral: 5; GAT: 2	NCT: one examiner, transpalpebral: patient; GAT: one physician

Study details	No. participants/eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Naruse 2005,[177] Japan	101/101	42 (17); NR; NR; mixed population	14.7 (4.1)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 5; GAT: 3	Transpalpebral: patient; GAT: one ophthalmologist
Ogbuehi 2006,[195] KSA	60/60	23 (2); 21/39; NR; not known to have OHT or glaucoma	13.4 (2.3)	NCT (Topcon CT80)	NCT: 4; GAT: 3	NCT: one examiner; GAT: one ophthalmologist
Ogbuehi 2008,[194] KSA	72/72	23 (3); 26/46; NR; not known to have OHT or glaucoma	13.9 (2.2)	NCT (Pulsair Easy Eye)	NCT: 3; GAT: 3	NCT: one ophthalmologist; GAT: one of three clinicians
Oncel 2009,[212] Turkey	60/60	41 (16); 24/36; NR; not known to have OHT or glaucoma	14.8 (3.1)	ORA, DCT (Pascal), NCT (Topcon CT80)	ORA: 3; DCT: 3, NCT: 3; GAT: 2	ORA, NCT and DCT: NR; GAT: NR
Pache 2005,[163] Germany	100/100	24 (19); 54/46; NR; not known to have glaucoma or OHT	15 (3)	DCT (Pascal)	DCT: NR; GAT: NR	DCT: two ophthalmologists; GAT: NR
Pakrou 2008,[167] Australia	153/292	60 (21); 72/81; NR; mixed population	17.6 (8.7)	RT (Icare)	RT: 3 (six readings); GAT: 5	RT: one of two ophthalmology residents; GAT: one of two consultant ophthalmologists
Papastergiou 2008,[201] Greece	46/46	30 (6); 24/22; NR; not known to have OHT or glaucoma	15.9 (2.6)	DCT (Pascal), NCT (Niked NT-3000)	DCT: 2; NCT: 2; GAT: 2	NCT and DCT: NR; GAT: NR
Pelit 2009,[151] Italy	135/135	56 (9); 69/66; NR; not known to have OHT or glaucoma	15.1 (2.4)	DCT (Pascal), NCT (Canon TX-10)	DCT: 3, NCT: 3; GAT: 3	DCT and NCT: one ophthalmologist; GAT: NR
Pepose 2007,[135] USA	33/66	40 (11), 19/14; Caucasian, not known to have OHT or glaucoma	13.8 (3.3)	ORA, DCT (Pascal)	ORA: NR; DCT: NR; GAT: NR	ORA and DCT: NR; GAT: NR
Popovich 1997,[139] USA	220/421	65 (NR); 216/4; NR; mixed population	16.8 (6.0)	NCT (Reichert Xpert)	NCT: 3; GAT: 1	NCT: one of three trained technicians; GAT: one of five ophthalmologists
Pourjavan 2007,[181] Belgium	28/52	40 (NR); 15/13; NR; healthy volunteers	15.9 (2.8)	DCT (Pascal)	DCT: 3; GAT: 2	DCT: one ophthalmologist; GAT: one ophthalmologist
Punjabi 2006,[131] US	501/906	59 (9); 226/275; NR; mixed population	17.0 (NR)	DCT (Pascal)	DCT: 3; GAT: 2 or 3	DCT: two or more technicians; GAT: two technicians
Rai 2005,[124] USA	135/135	60 (NR); 74/61; white (108), black (19), Asian (3), other (5); various type of glaucoma	NR	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: NR	Transpalpebral: patient and technician; GAT: one examiner
Realini 2009,[141] USA	85/170	NR; 55/30; white (81), other (4); mixed population	NR	DCT (Pascal)	DCT: 2; GAT: 2	DCT: NR; GAT: NR
Regine 2006,[149] Italy	100/200	69 (10); 60/40; NR; mixed population	NR	NCT (Nidek NT-4000)	NCT: 3; GAT: 3	NCT: two ophthalmologists; GAT: NR
Rehman 2008,[209] Sweden	45/45	NR; 27/18; NR; treated glaucoma or OHT	17.2 (5.4)	RT (Icare)	RT: 1 (six readings); GAT: NR	RT: NR; GAT: one of three experienced ophthalmic nurses
Rietveld 2005,[208] Netherland	45/45	62 (NR); 21/24; NR; mixed population	19.9 (5.3)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: NR	Transpalpebral: one ophthalmic technician; GAT: one ophthalmologist

Study details	No. participants/ eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Roszkowska 2009,[150] Italy	35/70	36 (11); 20/15; NR; not known to have OHT or glaucoma with normal and PRK thinned corneas	14.6 (1.8)	DCT (Pascal)	DCT: 2; GAT: 2	DCT: one examiner; GAT: NR
Roukonen 2007,[160] Germany	223/245	68 (14); NR; NR; glaucoma	19.8 (0.5)	RT (Icare)	RT: 1 (six readings); GAT: 3	RT: one ophthalmologist; GAT: NR
Sacu 2004,[126] Austria	68/68	65 (13); 40/24; NR; glaucoma	15.3 (3.1)	Ocuton S	Ocuton S: 6; GAT: 3	Ocuton S: patient; GAT: NR
Sahin 2007,[211] Turkey	61/61	57 (21); 32/29; NR; glaucoma	18.3 (3.4)	RT (Icare)	RT: 3 (six readings); GAT: 3	RT: one ophthalmologist; GAT: one of three examiners
Salvetat 2007,[144] Italy	101/101	62 (10); NR; NR; POAG	20.8 (6.1)	TonoPen (XL), DCT (Pascal)	TonoPen: 3, DCT: 3; GAT: 3	DCT and TonoPen: one ophthalmologist; GAT: NR
Schipper 2000,[188] Switzerland	25/35	37 (10); NR; NR; myopia	13 (3.1)	TonoPen	TonoPen: 3; GAT: 3	TonoPen: NR; GAT: NR
Shemesh 2007,[205] Israel	43/43	31 (10); 25/18; NR; not known to have OHT or glaucoma	12.1 (1.6)	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: 3	Transpalpebral: two experienced examiners; GAT: NR
Sullivan-Mee 2009,[142] USA	63/126[b]	64 (10); 2/58; white (30), Hispanic (19), American Indian (4), black (7); mixed population	64.4 (9.6)	DCT (Pascal), ORA	DCT: 2, ORA: 4; GAT: 2	DCT and ORA: two optometrists; GAT: NR
Tai 2005,[125] Taiwan	70/140	57 (12); 30/40; NR; mixed population	NR	PPT (Proview Eye Pressure Monitor)	Transpalpebral: 3; GAT: 3	Transpalpebral: patient; GAT: one physician
Tonnu 2005,[75,152] UK	105/105	65 (NR); NR; NR; not known to have OHT or glaucoma	17.2 (4.3)	TonoPen (XL), NCT (Canon TX-10)	TonoPen: 3, NCT: 3; GAT: 3	TonoPen and NCT: one medical student; GAT: one medical student
Troost 2005,[162] Germany	68/68	63 (NR); 34/34; NR; glaucoma	NR	Transpalpebral tonometer (TGDc-01)	Transpalpebral: 3; GAT: 3	Transpalpebral: one examiner; GAT: one examiner
Troost 2005,[158] Germany	20/40	24 (NR); 12/8; NR; healthy volunteers	13 (NR)	Transpalpebral tonometer (TGDc-01)	Transpalpebral: 3; GAT: 3	Transpalpebral: one examiner; GAT: one examiner
van der Jagt 2005,[207] Netherland	103/103	58 (12); 61/42; NR; mixed population	16.5 (6.4)	TonoPen (XL), Transpalpebral tonometer (TGDc-01), RT (Icare)	TonoPen: 3, transpalpebral: 3, RT: 3 (six readings); GAT: 3	RT, transpalpebral and TonoPen: one radiology technician; GAT: NR
Van de Velde 1995,[182] Belgium	100/100	64 (NR); NR; NR; mixed population	14.9 (3.1)	NCT (Nidek NT-1000)	NCT: 2; GAT: 1	NCT: one technician; GAT: one experienced ophthalmologist
Vandewalle 2009,[184] Belgium	92/92	62 (15); 60/32; NR; mixed population	15.1 (4.8)	RT (Icare), DCT (Pascal), ORA	RT: 4 (six readings); DCT: 2, ORA: 2; GAT: 2	RT, ORA and DCT: one examiner; GAT: one experienced examiner

Study details	No. participants/ eyes	Patient demographics: age (years), mean (SD); gender, F/M; race; diagnosis	GAT IOP (mmHg), mean (SD)	Technology	Number of measurements (comparator and GAT)	Examiner number/type
Vasconcelos De Moraes 2009,[213] Brazil	20/40	54 (NR); 15/5; Caucasian (12), others (8); POAG	17.3 (4.4)	DCT (Pascal)	DCT: 2; GAT: 2	DCT: one of two ophthalmologists; GAT: one of two ophthalmologists
Wells 2003,[170] Australia	85/170[b]	58 (20); NR; NR; not known to have OHT or glaucoma	21.8 (5.8)	Ocuton S	Ocuton S: 3; GAT: NR	Ocuton S: patient; GAT: one researcher
Yaoeda 2005,[179] Japan	175/175	50.9 (13.5); NR; NR; mixed population	NR	NCT (Nidek NT-4000)	NCT: 3; GAT: 3	NCT: one examiner; GAT: one examiner

APC, automatic puff control; F, female; M, male; NR, not reported; PRK, photorefractive keratectomy; POAG, primary open-angle glaucoma.
a Median.
b Not specifically stated.

Appendix 4

Systematic review of the agreement and reliability of tonometers: further analysis (*Chapter 5*)

Full results of the sensitivity analyses and subgroup/factor analyses as detailed in *Chapter 5*, *Data analysis* are given below by tonometer.

Dynamic contour tonometry

The full results of all analyses (main and sensitivity) are reported in *Table 60*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates were marginally different for the mean difference (1.7 mmHg; 95% CI 1.6 mmHg to 1.7 mmHg) and random error (2.2 mmHg; 95% CI 2.1 mmHg to 2.2 mmHg). The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. Exclusion of studies classified as low quality similarly did not have a substantial impact on the estimates of either the mean difference or the random error and therefore the LoA. Using studies in which only one eye per participant was included led to a mean difference of 1.9 mmHg (95% CI 1.4 mmHg to 2.5 mmHg) and a random error of 2.5 mmHg (95% CI 2.2 mmHg to 2.7 mmHg). The sensitivity analyses, which sought to address the impact of differing numbers of repeated observations (SMD and adjustment for repeated measurements analyses), did not lead to substantially different estimates of the mean difference although, as expected, the latter (based on a single measurement of both GAT and DCT across studies) had a slightly larger random error estimate of 2.7 mmHg (95% CI 2.5 mmHg to 2.9 mmHg) versus 2.4 mmHg (95% CI 2.1 mmHg to 2.7 mmHg). None of the sensitivity analyses had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 60*.

Clinical factor analyses

A meta-analysis investigating the impact of manufacturer was not possible as there was a sole manufacturer. Analyses exploring the IOP, impact of refractive surgery and examiner type were carried out involving 32, 21 and 18 studies, respectively. It was not clear that refractive surgery influenced the mean difference: 2.2 mmHg (95% CI –1.2 mmHg to 5.5 mmHg) for operative participants versus 2.0 mmHg (95% CI 1.6 mmHg to 2.5 mmHg) for non-operative patients. There was no suggestion of an impact on the mean difference of examiner type: ophthalmologist 1.5 mmHg (95% CI 0.8 mmHg to 2.3 mmHg) versus non-ophthalmologist 1.7 mmHg (95% CI 1.1 mmHg to 2.3 mmHg).

Two studies[131,210] reported results for IOP subgroups. In the study by Johannesson and colleagues.[210] the mean IOP difference between DCT and GAT was 2.5 mmHg for patients with the lowest IOP (< 16 mmHg), 1.3 mmHg for mid-range IOP and −0.8 mmHg for those with the highest IOP (> 23 mmHg). Punjabi and colleagues,[131] however, reported very little difference between high- and low-IOP subgroups, with a mean IOP difference of 1.8 mmHg for patients in the primary open-angle glaucoma group (all of whom had IOP > 22 mmHg) and a mean IOP

difference of 1.7 mmHg in the normal-tension glaucoma group (IOP ≤ 22 mmHg). Four further studies[130,141,148,218] carried out similar subgroup analyses of mean IOP difference by categorising patients as either healthy or glaucomatous, but this is not directly equivalent to an IOP subgroup analysis.

A total of 26 out of the 32 DCT studies in the meta-analysis (81%) reported CCT results, of which 14 related the results to difference in IOP measurements between DCT and GAT. Eleven studies[110,134,141,143,144,161,163,202,203,210,215] reported varying levels of correlation (from no correlation to a correlation coefficient of 0.46 in one study[141]) and the other three studies[123,148,189] reported results directly relating to CCT subgroups. Martinez-de-la-Casa and colleagues[189] reported the mean IOP difference between DCT and GAT as 5.5 mmHg for patients in the lowest CCT tertile (≤ 530 μm), 4.2 mmHg for mid-range CCT and 3.5 mmHg for those in the highest CCT tertile (> 565 μm). In the Doyle and Lachkar study,[123] the mean difference was 2.6 mmHg for patients in the lowest CCT tertile (≤ 530 μm), 0.1 mmHg for mid-range CCT and 0.1 mmHg for those in the highest CCT tertile (> 580 μm). Ceruti and colleagues[148] reported the mean IOP difference in healthy subjects as 3.7 mmHg for patients with low CCT (≤ 530 μm), 2.9 mmHg for mid-range CCT and 2.3 mmHg for those in the highest CCT tertile (> 560 μm). Using the same CCT thresholds as in their analysis of healthy patients, the Ceruti and colleagues study[148] also reported mean differences in patients with open-angle glaucoma of 4.2, 2.3 and 0.7 mmHg and in patients with angle-closure glaucoma of 3.5, 2.0 and 1.3 mmHg for low, mid-range and high CCT, respectively.

Three DCT studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. Duba and Wirthlin[186] reported a preoperative mean difference of 1.9 mmHg and a 3-month post-laser-assisted in situ keratomileusis (LASIK) mean difference of 4.2 mmHg. Pepose and colleagues[135] reported a preoperative mean difference of 3.3 mmHg and a 1-week post-LASIK mean difference of 4.5 mmHg. Liu and colleagues[171] reported a mean IOP difference prior to LASIK surgery of 1.3 mmHg compared with 5 mmHg 1 week postoperatively and 5.45 mmHg after 4 weeks. A further study[150] found a mean difference of 2.4 mmHg in non-operated corneas and a mean difference of 1.5 mmHg in corneas thinned by photorefractive keratectomy, but these groups were independent samples and not necessarily comparable.

Non-contact tonometry

The full results of all analyses (main and sensitivity) are reported in *Table 61*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates were marginally different for the mean difference (0.1 mmHg; 95% CI 0.0 mmHg to 0.2 mmHg) and random error (1.8 mmHg; 95% CI 1.7 mmHg to 1.8 mmHg). The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. Exclusion of low-quality studies did not have a substantial impact on the estimates for either the mean difference or the random error. Only analysing studies in which one eye per participant was used led to a mean difference of 0.5 mmHg (95% CI 0.0 mmHg to 0.9 mmHg) and random error of 2.0 mmHg (95% CI 1.8 mmHg to 2.3 mmHg). The impact on the LoA is also shown in *Table 61*. The sensitivity analyses, which sought to address the impact of differing numbers of repeated observations (SMD and adjustment for repeated measurements analyses), did not lead to substantially different estimates of the mean difference although random error was substantially larger (2.6 mmHg vs 2.1 mmHg). None of the sensitivity analyses had a marked impact on the observed heterogeneity.

TABLE 60 Meta-analysis of agreement between DCT and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (DCT – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	32	3952	1.8	1.4 to 2.2	97%	2.4	2.1 to 2.6	95%	–2.9 to 6.5
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	32	3952	1.8	1.4 to 2.2	96%	2.5	2.3 to 2.7	95%	–3.1 to 6.7
Fixed-effects model with mean correlation	32	3952	1.7	1.6 to 1.7	97%	2.2	2.1 to 2.2	95%	–2.6 to 5.9
Fixed-effects model with minimum correlation	32	3952	1.7	1.6 to 1.8	96%	2.2	2.2 to 2.3	95%	–2.7 to 6.1
Additional exclusion criteria									
Main analysis excluding low-quality studies	23	2624	1.7	1.2 to 2.1	96%	2.2	2.0 to 2.5	95%	–2.7 to 6.1
Main analysis excluding multiple eye studies	18	2203	1.9	1.4 to 2.5	96%	2.5	2.2 to 2.7	91%	–2.9 to 6.8
SMD									
Main analysis with SMD	31	3853	0.5	0.4 to 0.6	96%				
Main analysis of mean difference excluding studies in which SMD was not calculable	31	3853	1.7	1.3 to 2.1	96%				
Adjustment for repeated measurements									
Main analysis with SE adjusted for repeated measurements	24	3215	1.8	1.3 to 2.2	96%	2.7	2.5 to 2.9	87%	–3.5 to 7.1
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	24	3215	1.8	1.3 to 2.2	97%	2.4	2.1 to 2.7	96%	–2.9 to 6.5

TABLE 60 Meta-analysis of agreement between DCT and GAT (mmHg unless otherwise stated) (*continued*)

Description of analysis	No. of studies	No. of eyes	Mean difference (DCT – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Clinical factors analyses									
IOP									
Studies with > 33% of patients with IOP > 21 mmHg	5	723	2.0	0.4 to 3.5	99%	2.6	2.5 to 2.8	29%	−3.1 to 7.1
Studies with < 33% of patients with IOP > 21 mmHg	27	3229	1.8	1.4 to 2.1	96%	2.3	2.1 to 2.6	95%	−2.8 to 6.3
CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 µm									
Refractive surgery									
Studies with operative patients	4	205	2.2	−1.2 to 5.5	100%	1.9	1.2 to 2.5	92%	−1.5 to 5.8
Studies with non-operative patients	17	2584	2.0	1.6 to 2.5	95%	2.5	2.3 to 2.6	82%	−2.8 to 6.8
Examiner type									
Studies with ophthalmologist examiners	13	1452	1.5	0.8 to 2.3	98%	2.2	2.0 to 2.5	92%	−2.9 to 5.9
Studies with non-ophthalmologist examiners	5	1009	1.7	1.1 to 2.3	92%	2.5	2.3 to 2.8	67%	−3.3 to 6.7
No studies with patient as examiner									

Manufacturer: not applicable as there was only one manufacturer – SMT Swiss Microtechnology AG

SE, standard error.

TABLE 61 Meta-analysis of agreement between NCT and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (NCT – GAT)	95% CI	I^2	Random error	95% CI	I^2	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	26	2477	0.2	–0.1 to 0.6	95%	2.1	1.8 to 2.3	95%	–3.8 to 4.3
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	26	2477	0.3	–0.1 to 0.6	94%	2.3	2.1 to 2.6	94%	–4.3 to 4.8
Fixed-effects model with mean correlation	26	2477	0.1	0.0 to 0.2	95%	1.8	1.7 to 1.8	95%	–3.4 to 3.6
Fixed-effects model with minimum correlation	26	2477	0.1	0.0 to 0.2	94%	2.1	2.0 to 2.1	94%	–3.9 to 4.2
Additional exclusion criteria									
Main analysis excluding low-quality studies	21	1704	0.3	–0.1 to 0.7	95%	1.9	1.7 to 2.2	95%	–3.5 to 4.1
Main analysis excluding multiple eye studies	15	1032	0.5	0.0 to 0.9	95%	2.0	1.8 to 2.3	91%	–3.5 to 4.4
SMD									
Main analysis with SMD	23	2145	0.1	0.0 to 0.2	93%				
Main analysis of mean difference excluding studies in which SMD was not calculable	23	2145	0.3	0.0 to 0.6	94%				
Adjustment for repeated measurements									
Main analysis with SE adjusted for repeated measurements	22	2080	0.3	–0.1 to 0.7	93%	2.6	2.4 to 2.8	87%	–4.8 to 5.4
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	22	2080	0.3	–0.1 to 0.7	95%	2.2	1.9 to 2.4	93%	–4.0 to 4.5

TABLE 61 Meta-analysis of agreement between NCT and GAT (mmHg unless otherwise stated) (*continued*)

Description of analysis	No. of studies	No. of eyes	Mean difference (NCT−GAT)	95% CI	I^2	Random error	95% CI	I^2	95% LoA
Subgroup analyses									
IOP: not applicable as all studies have <33% of patients with IOP >21 mmHg									
CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 μm									
Refractive surgery									
Studies with operative patients	3	305	−2.2	−5.0 to 0.7	100%	1.5	0.9 to 2.1	96%	−5.0 to 0.7
Studies with non-operative patients	7	631	0.5	−0.1 to 0.9	82%	2.0	1.5 to 2.4	94%	−3.4 to 4.4
Examiner type									
Studies with ophthalmologist examiners	4	477	1.0	0.3 to 1.6	89%	2.4	1.8 to 3.0	94%	−3.7 to 5.6
Studies with non-ophthalmologist examiners	6	819	0.4	−0.5 to 1.3	97%	2.4	2.0 to 2.8	92%	−4.3 to 5.1
Studies with patient as examiner	1	68	−0.5	−0.9 to −0.1	NA	1.7	1.4 to 1.9	NA	−3.7 to 2.7
Manufacturer									
Canon	2	240	0.3	−1.7 to 2.4	97%	2.6	2.4 to 2.9	0%	−4.8 to 5.5
Keeler	3	229	0.1	−0.6 to 0.9	86%	2.2	1.3 to 3.2	96%	−4.3 to 4.6
Nidek	8	542	0.2	−0.6 to 1.1	95%	2.1	1.7 to 2.5	91%	−3.9 to 4.4
Reichert	7	842	0.2	−0.7 to 1.1	97%	2.1	1.7 to 2.5	93%	−3.9 to 4.3
Topcon	4	355	0.4	−0.1 to 0.8	74%	2.0	1.3 to 2.7	96%	−3.5 to 4.2

NA, not applicable; SE, standard error.

Clinical factor analyses

It was not possible to carry out a subgroup meta-analysis investigating the impact of IOP and CCT because of the way that the data were reported in individual studies. Analyses exploring the impact of refractive surgery, examiner type and manufacturer were carried out and included 10, 11 and 24 studies, respectively. It was possible, although not clear, that receiving refractive surgery impacted on the mean difference; however, this was based on a small number of studies with operative patients. There was no suggestion of an impact on the mean difference according to examiner type [ophthalmologist 1.0 mmHg (95% CI 0.3 mmHg to 1.6 mmHg) and non-ophthalmologist 0.4 mmHg (95% CI –0.5 mmHg to 1.3 mmHg)] or self-tonometry –0.5 mmHg (–95% CI –0.9 mmHg to –0.1 mmHg). Manufacturer-based estimates of mean difference were broadly similar.

Two NCT studies[182,196] reported results for IOP subgroups. In Van de Velde and Zeyen,[182] the mean IOP difference between NCT and GAT was 0.8 mmHg for patients with the lowest IOP (< 12 mmHg), 0.4 mmHg for mid-range IOP and 0.6 mmHg for those with the highest IOP (> 18 mmHg). Conversely, Gupta and colleagues[196] reported a mean IOP difference of 0.4 mmHg for patients with IOP of ≤ 18 mmHg and –1.7 mmHg for those with the highest IOP (> 18 mmHg). One study[152] carried out a regression of the mean difference on the underlying IOP (average of NCT and GAT values), which suggested that values of –1 to 2 mmHg were plausible.

A total of 6 out of the 26 NCT studies reported CCT results, three of which reported correlation between CCT levels and the IOP difference between GAT and NCT. The correlation coefficient in the Papastergiou and colleagues study[201] was 0.34, Murase and colleagues[180] reported a correlation coefficient of 0.53 and Tonnu and colleagues[152] reported a value of 0.25. However, none of the six studies reported results that directly related to CCT subgroups.

Three studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. Garzozi and colleagues[206] reported mean differences of 0.1 mmHg preoperatively with 0.2 mmHg 1 year after photorefractive keratectomy, Naruse and colleagues[178] reported a preoperative mean difference of –1.6 mmHg with a 1-week post-LASIK mean difference of –2.2 mmHg and Cheng and colleagues[176] reported a mean difference of 0.5 mmHg preoperatively with a 3-month post-LASIK difference of –4.5 mmHg.

Ocuton S

The full results of all analyses (main and sensitivity) are reported in *Table 62*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates (mean correlation) were marginally different for the mean difference (2.4 mmHg; 95% CI 1.9 mmHg to 3.0 mmHg) and random error (3.5 mmHg; 95% CI 3.1 mmHg to 3.9 mmHg). It was not necessary to input any correlation values. None of the studies was classified as low quality. Using the studies in which only one eye per participant was included led to a mean difference of 4.4 mmHg (95% CI 0.5 mmHg to 8.2 mmHg) and random error of 3.2 mmHg (95% CI 1.4 mmHg to 5.0 mmHg). The sensitivity analyses to address the impact of differing numbers of repeated observations (SMD only) led to a SMD of 0.6 mmHg (95% CI –0.2 mmHg to 1.5 mmHg) although the I^2 was still very large (97%). The adjustment of the repeated measurements led to a marginally smaller mean difference and as anticipated a slightly larger random error (5.2 mmHg; 95% CI 4.3 mmHg to 6.2 mmHg). None of the sensitivity analyses had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 62*.

Clinical factor analyses

A subgroup meta-analysis investigating the impact of CCT, refractive surgery, examiner type and manufacturer was not performed either because of the way that the data were reported in individual studies or because there were no studies representing a particular clinical factor.

Three studies were included in IOP analysis with the suggestion of a difference in mean difference depending on IOP level (–0.5 mmHg vs 4.4 mmHg). No Ocuton S studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. One Ocuton S study reported CCT results[146] but, although the CCT results were reported in relation to IOP measurements, it did not report results for any CCT subgroups. It was not possible to conduct an examiner analysis as none of the studies had an ophthalmologist examiner. All studies used a tonometer by the same manufacturer.

Ocular response analyser

The full results of all analyses (main and sensitivity) are reported in *Table 63*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates (mean correlation) were marginally different for the mean difference (1.2 mmHg; 95% CI 1.1 mmHg to 1.4 mmHg) and random error (2.7 mmHg; 95% CI 2.6 mmHg to 2.9 mmHg). The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. Exclusion of the sole low-quality study[184] did not have a substantial impact on the estimates of either the mean difference or the random error. Using studies in which only one eye per participant was included led to a mean difference of 1.8 mmHg (95% CI 1.0 mmHg to 2.7 mmHg) and random error of 2.8 mmHg (95% CI 2.4 mmHg to 3.2 mmHg). The sensitivity analyses addressing the impact of differing numbers of repeated observations (SMD) led to a SMD of 0.4 mmHg (95% CI 0.3 mmHg to 0.6 mmHg) although the I^2 was still very large (88%). The adjustment of the repeated measurements did not alter the mean difference estimate although, as anticipated, there was a slightly larger random effect estimate. None of the sensitivity analyses had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 63*.

Clinical factor analyses

A subgroup meta-analysis investigating the impact of IOP, CCT and manufacturer was not performed either because of the way that the data were reported in individual studies or because studies represented only one subgroup. Analyses exploring the impact of refractive surgery and examiner type were carried out involving 11 and 6 studies, respectively. Neither refractive surgery nor examiner type appeared to have any impact although only one study included ophthalmologist examiners. None of the studies used patient examiners.

No ORA studies reported results for IOP subgroups. However, Morita and colleagues[130] carried out a subgroup analysis of mean IOP difference by categorising patients as either healthy or glaucomatous, but this is not equivalent to an IOP subgroup analysis. One study[110] regressed the difference between ORA and GAT values to the underlying (average) IOP value. The estimated correlation was 0.41 with a slope of –0.3 (mmHg).

A total of 5 out of the 12 studies reported results for CCT with two reporting low correlation between CCT levels and the IOP difference between ORA and GAT. These studies were those by Lam and colleagues,[174] with a correlation coefficient of 0.17, and Kotecha and colleagues,[110] with a value of 0.1.

TABLE 62 Meta-analysis of agreement between Ocuton S and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (Ocuton S – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	3	164	2.7	–1.2 to 6.6	98%	3.5	2.4 to 4.6	88%	–4.1 to 9.6
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	Not applicable – no imputation required								
Fixed-effects model with mean correlation	3	164	2.4	1.9 to 3.0	98%	3.5	3.1 to 3.9	88%	–4.5 to 9.3
Fixed-effects model with minimum correlation	Not applicable – no imputation required								
Additional exclusion criteria									
Main analysis excluding low-quality studies	Not applicable – no low-quality studies								
Main analysis excluding multiple eye studies	2	84	4.4	0.5 to 8.2	97%	3.2	1.4 to 5.0	92%	–2.0 to 10.7
SMD									
Main analysis with SMD	3	164	0.6	–0.2 to 1.5	97%				
Main analysis of mean difference excluding studies in which SMD was not calculable	Not applicable – SMD calculable in all studies								
Adjustment for repeated measurements									
Main analysis with SE adjusted for repeated measurements	1	64	2.4	1.1 to 3.7	NA	5.2	4.3 to 6.2	NA	–7.9 to 12.7
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	1	64	2.4	1.4 to 3.4	NA	4.1	3.4 to 4.8	NA	–5.7 to 10.5
Clinical factors analyses									
IOP									
Studies with >33% of patients with IOP >21 mmHg	1	80	–0.5	–1.3 to 0.4	NA	4.0	3.4 to 4.6	NA	–8.3 to 7.3
Studies with <33% of patients with IOP >21 mmHg	2	84	4.4	0.5 to 8.2	97%	3.2	1.4 to 5.0	92%	–2.0 to 10.7

Refractive surgery: not applicable as no studies have operative patients

Examiner type: not applicable as no studies have ophthalmologist examiners

Manufacturer: not applicable as all Ocuton S tonometers in the review were manufactured by EPSa Elektronik & Präzisions-Bau Saalfeld GmbH

NA, not applicable; SE, standard error.

Two studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. Pepose and colleagues[135] reported a preoperative mean difference of 1.6 mmHg and a 1-week post-LASIK mean difference of 1.1 mmHg. Kirwan and O'Keefe[214] reported a pre-LASIK mean difference of 0.7 mmHg with a 3-month postoperative mean difference of 3.5 mmHg and, similarly, a pre-laser-assisted subepithelial keratectomy mean difference of 0.2 mmHg with a 3-month postoperative mean difference of 2.8 mmHg.

Perkins

The full results of all analyses (main and sensitivity) are reported in *Table 64*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates (mean correlation) were quite different for the mean difference (–0.4 mmHg; 95% CI –0.5 mmHg to –0.2 mmHg) because of the impact of one study.[166] This was not the case for the random error (2.1 mmHg; 95% CI 1.3 mmHg to 2.8 mmHg), which was similar. The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. None of the studies was classified as low quality. Excluding the one study which used data from both eyes of the same individual from the analysis did affect the results. However, this was driven by the unusual results of this study.[166] The sensitivity analyses to address the impact of differing numbers of repeated observations (SMD) led to a SMD of 0.0 mmHg (95% CI –0.1 to 0.1 mmHg) although the I^2 was still very large (70%). The adjustment for repeated measurements was not conducted as none of the studies reported the appropriate data. None of the sensitivity analyses had an impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 64*.

Clinical factor analyses

No subgroup meta-analyses were performed either because of the way that the data were reported in individual studies or because of a lack of studies representing different groups.

None of the Perkins tonometer studies reported results for IOP subgroups, nor did any study report CCT results. There were also no studies that reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. An analysis of examiner or manufacturer was not carried out. General practitioners used the Perkins tonometer in the Jackson and colleagues study,[166] which had a high mean difference and random error relative to the other studies.

Rebound tonometry

The full results of all analyses (main and sensitivity) are reported in *Table 65*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates (mean correlation) were marginally different for the mean difference (1.4 mmHg; 95% CI 1.3 mmHg to 1.5 mmHg) and random error (1.8 mmHg; 95% CI 1.8 mmHg to 1.9 mmHg). No imputation of the correlation value was needed. Exclusion of low-quality studies led to a slightly lower mean difference whereas the random error was the same. Using studies in which only one eye per participant was included led to a mean difference of 0.5 mmHg (95% CI 0.0 mmHg to 1.0 mmHg) and random error of 2.7 mmHg (95% CI 2.4 mmHg to 3.1 mmHg). The sensitivity analyses to address the impact of differing numbers of repeated

TABLE 63 Meta-analysis of agreement between ORA and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (ORA – GAT)	95% CI	I^2	Random error	95% CI	I^2	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	12	1052	1.5	0.9 to 2.2	93%	2.8	2.5 to 3.1	85%	−3.9 to 7.0
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	12	1052	1.5	0.9 to 2.2	92%	3.2	2.8 to 3.6	87%	−4.8 to 7.8
Fixed-effects model with mean correlation	12	1052	1.2	1.1 to 1.4	93%	2.7	2.6 to 2.9	85%	−4.1 to 6.6
Fixed-effects model with minimum correlation	12	1052	1.2	1.0 to 1.4	92%	3.0	2.9 to 3.1	87%	−4.7 to 7.0
Additional exclusion criteria									
Main analysis excluding low-quality studies	11	960	1.4	0.7 to 2.0	92%	2.7	2.4 to 3.0	81%	−3.9 to 6.6
Main analysis excluding multiple eye studies	7	593	1.8	1.0 to 2.7	92%	2.8	2.4 to 3.2	81%	−3.7 to 7.4
SMD									
Main analysis with SMD	12	1052	0.4	0.3 to 0.6	88%				
Main analysis of mean difference excluding studies in which SMD was not calculable	12	1052	1.5	0.9 to 2.2	93%				
Adjustment for repeated measurements									
Main analysis with standard error (SE) adjusted for repeated measurements	8	847	1.7	0.8 to 2.7	94%	3.3	3.0 to 3.7	71%	−4.8 to 8.3
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	8	847	1.7	0.8 to 2.7	95%	3.0	2.6 to 3.3	81%	−4.1 to 7.5

continued

TABLE 63 Meta-analysis of agreement between ORA and GAT (mmHg unless otherwise stated) *(continued)*

Description of analysis	No. of studies	No. of eyes	Mean difference (ORA – GAT)	95% CI	I^2	Random error	95% CI	I^2	95% LoA
Clinical factors analyses									
IOP: not applicable as all studies have <33% of patients with IOP >21 mmHg									
CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 µm									
Refractive surgery									
Studies with operative patients	2	156	2.3	–0.0 to 4.7	98%	1.9	1.3 to 2.5	85%	–1.4 to 6.1
Studies with no operative patients	9	866	1.8	0.9 to 2.6	95%	2.9	2.6 to 3.2	79%	–3.9 to 7.4
Examiner type									
Studies with ophthalmologist examiners	1	30	0.4	–0.2 to 1.0	NA	1.6	1.2 to 2.0	NA	–2.7 to 3.5
Studies with non-ophthalmologist examiners	5	566	1.7	0.6 to 2.9	95%	2.9	2.7 to 3.1	45%	–4.0 to 7.4
No studies with patient as examiner									
Manufacturer: not applicable as all ORA tonometers in the review were manufactured by Reichert									

NA, not applicable; SE, standard error.

TABLE 64 Meta-analysis of agreement between Perkins and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (Perkins − GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	4	504	−1.2	−2.8 to 0.4	99%	2.1	1.3 to 2.8	97%	−5.2 to 2.8
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	Not applicable – no imputation required								
Fixed-effects model with mean correlation	4	504	−0.4	−0.5 to −0.2	99%	1.8	1.7 to 1.9	97%	−3.9 to 3.2
Fixed-effects model with minimum correlation	Not applicable – no imputation required								
Additional exclusion criteria									
Main analysis excluding low-quality studies	Not applicable – no low-quality studies								
Main analysis excluding multiple eye studies	3	360	0.0	−0.4 to 0.3	71%	1.5	1.1 to 2.0	89%	−3.0 to 3.0
SMD									
Main analysis with SMD	3	360	0.0	−0.1 to 0.1	70%				
Main analysis of mean difference excluding studies in which SMD was not calculable	3	360	0.0	−0.4 to 0.3	71%				

Adjustment for repeated measurements: not applicable as no studies reported relevant results for repeatability

Clinical factors analyses

IOP: not applicable as all studies have <33% of patients with IOP >21 mmHg

CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 μm

Refractive surgery: not applicable as no studies have operative patients

Examiner type: not applicable as no studies have ophthalmologist examiners

Manufacturer: not applicable as all Perkins tonometers in the review were manufactured by Kowa

observations (SMD) led to a SMD of 0.2 mmHg (95% CI 0.1 to 0. 3 mmHg). None of the sensitivity analysis had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 65*.

Clinical factor analyses

A subgroup meta-analysis investigating the impact of CCT, refractive surgery and manufacturer was not performed because of the way that the data were reported in individual studies for CCT and refractive surgery, and because there was only a sole manufacturer. Analyses exploring the IOP and examiner type were carried out involving 13 and 11 studies, respectively. The underlying IOP level did not appear to have an impact on mean difference or random error although the random error for the higher IOP group was larger. There was no clear evidence of an impact on the mean difference of examiner: ophthalmologist 0.8 mmHg (95% CI 0.0 mmHg to 1.5 mmHg) and non-ophthalmologist 1.3 mmHg (95% CI 0.5 mmHg to 2.1 mmHg). None of the studies used patient examiners.

Three RT studies reported results for IOP subgroups. In Sahin and colleagues,[211] the mean IOP difference between RT and GAT was 0 mmHg for patients with IOP within the normal range (< 20 mmHg) and 1.9 mmHg for those with high IOP (≥ 20 mmHg). Johannesson and colleagues[210] reported a mean difference of 2.4 mmHg for patients with the lowest IOP (< 16 mmHg), 2.4 mmHg for mid-range IOP and 1.3 mmHg for those with the highest IOP (> 23 mmHg). Pakrou and colleagues[167] was the only study in the whole review to use the 21-mmHg threshold of OHT to define subgroups. In this study, patients with IOP < 21 mmHg had a mean difference of 0.7 mmHg compared with –0.7 mmHg for those with high IOP.

A total of 8 of the 14 RT studies reported CCT results, five of which related CCT levels to the IOP difference between GAT and RT. One of these studies[189] presented results relating directly to CCT subgroups with a mean IOP difference between RT and GAT of 0.4 mmHg for patients in the lowest CCT tertile (≤ 530 μm), 2.3 mmHg for mid-range CCT and 1.5 mmHg for those in the highest CCT tertile (> 565 μm). The other studies reported correlation between IOP difference and CCT, with correlation coefficients of 0.42 in Johannesson and colleagues,[210] 0.3 in Sahin and colleagues,[211] 0.63 in Brusini and colleagues[145] and 0.47 in Nakamura and colleagues.[107]

No RT studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery.

TonoPen

The full results of all analyses (main and sensitivity) are reported in *Table 66*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed effect estimates (mean correlation) were marginally different for the mean difference (–0.1 mmHg; 95% CI –0.2 mmHg to 0.0 mmHg) and random error (2.1 mmHg; 95% CI 2.0 mmHg to 2.2 mmHg). The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. Exclusion of low-quality studies did not have a substantial impact on the estimates for either the mean difference or the random error. Using studies in which only one eye per participant was included led to a mean difference of 0.3 mmHg (95% CI –0.7 mmHg to 1.2 mmHg) and random error of 3.0 mmHg (95% CI 2.5 mmHg to 3.5 mmHg). The sensitivity analyses to address the impact of differing numbers of repeated observations (SMD) led to a SMD of 0.1 mmHg (95% CI 0.0 mmHg to 0.3 mmHg) although the I^2 was still very large (94%). The adjustment for repeated measurements led to a marginally smaller

TABLE 65 Meta-analysis of agreement between RT and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (RT – GAT)	95% CI	I^2	Random error	95% CI	I^2	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	14	1543	0.9	0.4 to 1.4	94%	2.6	2.1 to 3.2	98%	−4.3 to 6.1
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	Not applicable – no imputation required								
Fixed-effects model with mean correlation	14	1543	1.4	1.3 to 1.5	94%	1.8	1.8 to 1.9	98%	−2.3 to 5.0
Fixed-effects model with minimum correlation	Not applicable – no imputation required								
Additional exclusion criteria									
Main analysis excluding low-quality studies	8	660	0.5	−0.1 to 1.1	88%	2.6	2.2 to 3.0	86%	−4.6 to 5.6
Main analysis excluding multiple eye studies	10	855	0.5	0.0 to 1.0	85%	2.7	2.4 to 3.1	87%	−4.8 to 5.9
SMD									
Main analysis with SMD	11	1052	0.2	0.1 to 0.3	91%				
Main analysis of mean difference excluding studies in which SMD was not calculable	11	1052	0.9	0.3 to 1.5	92%				
Adjustment for repeated measurements: not applicable as no studies reported relevant results for repeatability									
Clinical factors analyses									
IOP									
Studies with >33% of patients with IOP >21 mmHg	3	474	0.8	−0.9 to 2.5	97%	3.1	2.7 to 3.5	81%	−5.3 to 6.9
Studies with <33% of patients with IOP >21 mmHg	10	926	1.0	0.5 to 1.5	92%	2.5	1.8 to 3.1	97%	−3.9 to 5.8
CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 μm									
Refractive surgery: not applicable as no studies have operative patients									
Examiner type									
Studies with ophthalmologist examiners	8	1065	0.8	0.0 to 1.5	96%	2.5	1.7 to 3.2	98%	−4.1 to 5.7
Studies with non-ophthalmologist examiners	3	299	1.3	0.5 to 2.1	83%	2.8	2.1 to 3.6	91%	−4.2 to 6.9
No studies with patient as examiner									
Manufacturer: not applicable as all RTs in the review were manufactured by the same company									

mean difference and slightly larger random effect as anticipated. None of the sensitivity analysis had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 66*.

Clinical factor analyses

A meta-analysis investigating the impact of CCT was not performed because of the way that the data were reported in individual studies. Analyses exploring the IOP, refractive surgery, examiner type and manufacturer were carried out involving 12, 3, 10 and 12 studies, respectively. There was no clear evidence that underlying IOP level had an impact on either the mean difference or random error although there was only a single study in the higher IOP group. Refractive surgery did not appear to have any impact although there was only one study with no operative participants available. There was a suggestion of an impact on the mean difference according to examiner: ophthalmologist –0.8 mmHg (95% CI –1.5 mmHg to –0.2 mmHg) and non-ophthalmologist –0.3 mmHg (95% CI –2.5 mmHg to 1.9 mmHg). GPs used the Perkins tonometer in the Jackson and colleagues study,[166] which had a high mean difference relative to the other studies. None of the studies had patients as examiners. There was no clear evidence of a difference in either the mean difference or the random error between manufacturers.

Results for IOP subgroups were reported in two studies. In the Kooner and colleagues study,[140] the mean IOP difference between TonoPen and GAT was 0.5 mmHg for patients with the lowest IOP (< 16 mmHg), –1.8 mmHg for mid-range IOP (16–30 mmHg) and –7.1 mmHg for the single patient with IOP > 30 mmHg. A similar pattern was observed by Horowitz and colleagues,[169] who reported a mean difference of 0.5 mmHg for patients with the lowest IOP (< 16 mmHg), 0.1 mmHg for mid-range IOP and –2.2 mmHg for those with the highest IOP (between 20 and 30 mmHg).

Three TonoPen studies reported CCT results, two of which reported the correlation between CCT levels and the difference in IOP between TonoPen and GAT. Both studies had an identical low correlation of 0.2.[144,152] No TonoPen study reported results directly in relation to CCT subgroups.

Two TonoPen studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery, both of which measured IOP in the central part of the cornea preoperatively and 1 month following photorefractive keratectomy. Schipper and colleagues[188] reported a preoperative mean difference of 0.4 mmHg with a postoperative mean difference of 1.2 mmHg and Garzozi and colleagues[206] reported a preoperative mean difference of 0.1 mmHg and a postoperative mean difference of 0.3 mmHg. Schipper and colleagues also reported a mean difference of 0.1 mmHg after 3 months, but only followed up 27 patients from the original cohort of 35.

Transpalpebral

The full results of all analyses (main and sensitivity) are reported in *Table 67*. The impact of the sensitivity analyses and subgroup analyses and a summary of the factor analyses are given below.

Sensitivity analyses

The fixed-effect estimates (mean correlation) were marginally different for the mean difference (–0.1 mmHg; 95% CI –0.3 mmHg to 0.0 mmHg) and random error (2.5 mmHg; 95% CI 2.4 mmHg to 2.5 mmHg). The imputed correlation value had little impact on the estimates for either random- or fixed-effect(s) meta-analyses. Exclusion of low-quality studies did not have a substantial impact on the estimates of either the mean difference or the random error

TABLE 66 Meta-analysis of agreement between TonoPen and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (TonoPen – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	14	1619	–0.2	–1.0 to 0.5	97%	3.1	2.5 to 3.7	98%	–6.2 to 5.8
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	14	1619	–0.2	–1.0 to 0.6	97%	3.1	2.6 to 3.6	97%	–6.3 to 5.8
Fixed-effects model with mean correlation	14	1619	–0.1	–0.2 to 0.0	97%	2.1	2.0 to 2.2	98%	–4.2 to 4.1
Fixed-effects model with minimum correlation	14	1619	–0.1	–0.2 to 0.0	97%	2.4	2.3 to 2.5	97%	–4.8 to 4.6
Additional exclusion criteria									
Main analysis excluding low-quality studies	10	1085	–0.5	–1.5 to 0.5	98%	3.0	2.3 to 3.8	98%	–6.5 to 5.5
Main analysis excluding multiple eye studies	7	810	0.3	–0.7 to 1.2	97%	3.0	2.5 to 3.5	95%	–5.6 to 6.1
SMD									
Main analysis with SMD	9	1056	0.1	0.0 to 0.3	94%				
Main analysis of mean difference excluding studies in which SMD was not calculable	9	1056	0.4	–0.3 to 1.1	95%				
Adjustment for repeated measurements									
Main analysis with SE adjusted for repeated measurements	10	998	0.1	–0.7 to 0.9	94%	3.4	2.9 to 3.8	90%	–6.5 to 6.7
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	10	998	0.1	–0.8 to 0.9	96%	3.0	2.5 to 3.5	94%	–5.7 to 5.9

TABLE 66 Meta-analysis of agreement between TonoPen and GAT (mmHg unless otherwise stated) *(continued)*

Description of analysis	No. of studies	No. of eyes	Mean difference (TonoPen – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Clinical factors analyses									
IOP									
Studies with >33% of patients with IOP >21 mmHg	1	101	−0.5	−1.4 to 0.4	NA	4.5	3.9 to 5.1	NA	−9.3 to 8.3
Studies with <33% of patients with IOP >21 mmHg	11	1274	0.2	−0.4 to 0.9	96%	2.7	2.1 to 3.3	98%	−5.1 to 5.5
CCT: not applicable as all studies reporting CCT have >33% of patients with CCT <555 µm									
Refractive surgery									
Studies with operative patients	2	253	0.8	−0.1 to 1.6	96%	1.1	0.7 to 1.4	91%	−1.3 to 2.8
Studies with non-operative patients	1	103	0.8	0.1 to1.5	NA	3.6	3.1 to 4.1	NA	−6.2 to 7.8
Examiner type									
Studies with ophthalmologist examiners	4	468	−0.8	−1.5 to −0.2	85%	2.9	2.0 to 3.8	97%	−6.5 to 4.8
Studies with non-ophthalmologist examiners	6	665	−0.3	−2.5 to 1.9	99%	3.6	2.8 to 4.4	95%	−7.4 to 6.7
No studies with patient as examiner									
Manufacturer									
Intermedics Intraocular	1	152	−0.7	−1.2 to −0.2	NA	3.0	2.6 to 3.3	NA	−6.5 to 5.1
Medtronic	3	217	1.5	−0.7 to 3.7	96%	3.3	1.6 to 4.9	97%	−4.9 to 7.9
Mentor	8	976	−0.1	−0.7 to 0.6	95%	2.9	2.1 to 3.6	98%	−5.7 to 5.6

NA, not applicable; SE, standard error.

(a slight increase in the former and a reduction in the latter). Using studies in which only one eye per participant was included led to a mean difference of –0.9 mmHg (95% CI –2.0 mmHg to 0.2 mmHg) and random error of 3.7 mmHg (95% CI 2.9 mmHg to 4.6 mmHg. The sensitivity analyses to address the impact of differing numbers of repeated observations (SMD) led to a SMD of 0.1 mmHg (95% CI –0.1 mmHg to 0.3 mmHg) although the I^2 was still very large (97%). The adjustment of the repeated measurements led to a smaller mean difference and as anticipated a slightly larger random effect. None of the sensitivity analysis had a marked impact on the observed heterogeneity. The impact on the LoA of these analyses is also shown in *Table 67*.

Clinical factor analyses

A subgroup meta-analysis investigating the impact of CCT was not performed because of the way that the data were reported in individual studies. Analyses exploring the IOP, refractive surgery, examiner type and manufacturer were carried out involving 18, 12, 15 and 20 studies, respectively. It was not clear that underlying IOP level had an impact on mean difference. Refractive surgery did appear to have an impact although there were only three studies with operative participants. There was a suggestion of an impact on the mean difference according to examiner [ophthalmologist 3.6 mmHg (95% CI 1.3 mmHg to 5.9 mmHg), non-ophthalmologist –0.8 mmHg (95% CI –3.4 mmHg to 1.8 mmHg) and patient examiners –0.3 mmHg (95% CI –1.3 mmHg to 0.7 mmHg)] and also possibly on the random error. There was no evidence of a difference in either mean difference or random error between the two represented manufacturers.

Results for IOP subgroups were reported in two studies. Alvarez and colleagues[127] reported a mean difference of –0.4 mmHg in patients with normal IOP and 6.2 mmHg for patients with IOP of ≥ 22 mmHg. In the Tai and colleagues study[125] the mean IOP difference between the phosphene tonometer and GAT was 4.3 mmHg for patients with the lowest IOP (< 10 mmHg), 0.4 mmHg for patients with IOP within the normal range (10–20 mmHg), –2.7 mmHg for patients with higher IOP (20–30 mmHg) and –6.1 mmHg for patients with severe OHT (IOP > 30 mmHg). Gunvant and colleagues[133] carried out a similar subgroup analysis of mean IOP difference by categorising patients as either healthy or glaucomatous, but this is not directly equivalent to an IOP subgroup analysis. Only two[133,138] studies reported CCT results, both of which presented results in relation to the IOP difference between GAT and transpalpebral tonometry. Gunvant and colleagues[133] reported a low correlation coefficient of 0.10. Neither study reported results directly relating to CCT subgroups.

Three studies reported preoperative and postoperative results for a cohort of patients undergoing refractive surgery. Naruse and colleagues[178] reported a preoperative mean difference of 1.6 mmHg and a 1-week post-LASIK mean difference of 5 mmHg. A further two studies followed-up LASIK patients 3 months after surgery. Cheng and colleagues[176] reported a preoperative mean difference of 1.3 mmHg with a postoperative mean difference of 5.5 mmHg and Shemesh and colleagues[205] reported a preoperative mean difference of 0.2 mmHg with a postoperative mean difference of 1.8 mmHg.

TABLE 67 Meta-analysis of agreement between transpalpebral tonometry and GAT (mmHg unless otherwise stated)

Description of analysis	No. of studies	No. of eyes	Mean difference (PPT − GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Main analysis of mean difference and random error									
Random-effects model with mean correlation	20	1979	−0.5	−1.3 to 0.3	98%	3.3	2.8 to 3.7	97%	−6.9 to 5.9
Sensitivity analyses									
Model type and imputation method									
Random-effects model with minimum correlation	20	1979	−0.5	−1.3 to 0.3	97%	3.3	2.9 to 3.8	96%	−7.1 to 6.1
Fixed-effects model with mean correlation	20	1979	−0.1	−0.3 to 0.0	98%	2.5	2.4 to 2.5	97%	−4.9 to 4.7
Fixed-effects model with minimum correlation	20	1979	−0.3	−0.4 to −0.1	97%	2.6	2.5 to 2.7	96%	−5.4 to 4.8
Additional exclusion criteria									
Main analysis excluding low-quality studies	13	1232	−0.7	−1.6 to 0.3	98%	2.8	2.3 to 3.3	97%	−6.2 to 4.8
Main analysis excluding multiple eye studies	11	872	−0.9	−2.0 to 0.2	97%	3.7	2.9 to 4.6	97%	−8.2 to 6.4
SMD									
Main analysis with SMD	14	1380	0.1	−0.1 to 0.3	97%				
Main analysis of mean difference excluding studies in which SMD was not calculable	14	1380	0.3	−0.4 to 1.1	97%				
Adjustment for repeated measurements									
Main analysis with SE adjusted for repeated measurements	15	1386	−0.2	−1.2 to 0.7	96%	3.9	3.4 to 4.4	94%	−7.8 to 7.4
Main analysis with unadjusted SEs excluding studies in which adjusted SEs were not calculable	15	1386	−0.2	−1.1 to 0.6	97%	3.5	2.8 to 4.1	97%	−7.0 to 6.6

Description of analysis	No. of studies	No. of eyes	Mean difference (PPT – GAT)	95% CI	I²	Random error	95% CI	I²	95% LoA
Subgroup analyses									
IOP									
Studies with >33% of patients with IOP >21 mmHg	3	210	−1.5	−3.0 to 0.1	95%	2.7	1.8 to 3.5	91%	−6.7 to 3.7
Studies with <33% of patients with IOP >21 mmHg	15	1521	−0.1	−0.9 to 0.7	97%	3.3	2.7 to 3.9	97%	−6.5 to 6.3
CCT (not applicable as all studies reporting CCT have >33% of patients with CCT <555 μm)									
Refractive surgery									
Studies with operative patients	3	199	4.1	1.2 to 6.9	99%	1.8	−0.1 to 3.6	99%	0.6 to 7.6
Studies with non-operative patients	9	1034	−1.6	−3.0 to −0.1	98%	3.7	3.0 to 4.4	96%	−8.8 to 5.7
Examiner type									
Studies with ophthalmologist examiners	1	30	3.6	1.3 to 5.9	NA	6.5	4.9 to 8.1	NA	−9.1 to 16.3
Studies with non-ophthalmologist examiners	4	324	−0.8	−3.4 to 1.8	97%	4.0	1.7 to 6.4	98%	−8.7 to 7.1
Studies with patient as examiner	10	1174	−0.3	−1.3 to 0.7	97%	3.2	2.7 to 3.8	95%	−6.6 to 6.1
Manufacturer									
Bausch & Lomb	15	1534	−0.4	−1.3 to 0.5	98%	3.1	2.6 to 3.7	97%	−6.5 to 5.8
Ryazan State Instrument Making Enterprise	5	445	−0.9	−2.2 to 0.5	95%	3.6	2.5 to 4.8	96%	−8.0 to 6.2

NA, not applicable; SE, standard error.

Appendix 5

Recordability data by tonometer (*Chapter 5*)

Study	% recordable IOP (*n/N*)	Comments
DCT		
Barleon 2006[161]	93 (197/211)	14 eyes (3/69 healthy and 11/142 glaucoma) had difficulties with DCT
Johannesson 2008[210]	99 (149/150)	Four eyes of four people (19 measurements) for which not all six measurements were achieved. Mean quality value of measurement was also reported, which 'implied' mean of 0.84 extra measurements needed. Only one eye for which a mean IOP was not reported
Martinez-de-la-Casa 2006[189]	100 (146/146)	No eyes were excluded because of poor or unreliable measurements
Medeiros 2007[134]	100 (176/176)	Two patients needed a third DCT measurement because of excessive blinking
Milla 2009[192]	99 (99/100)	
Sullivan-Mee 2009[142]	98 (62/63)	Three subjects were excluded due to: reliable DCT readings could not be obtained for the second measurement set in one participant
NCT		
Boles 1992[147]	76 (68/90)	
Mackie 1996[156]	96 (43/45)	Any notably apprehensive patients who made the examination difficult by narrowing their palpebral aperture during measurements were also excluded
Ogbuehi 2008[194]	100 (72/72)	Some patients needed to have their IOP measurement repeated to get valid readings
Van de Velde 1995[182]	99 (99/100)	One eye with corneal oedema and unreliable NCT readings was excluded
Ocuton S		
Sacu 2004[126]	94 (64/68)	Excluded those with corneal disease, astigmatism > 1 D, contact lens, inflammation, recent surgery (< 3 months)
Wells 2003[170]	82 (70/85)	70 patients obtained one measurement with the Ocuton S; 61 patients obtained three measurements; 15 were unable to obtain any measurements
ORA		
Kynigopoulos 2008[185]	98 (49/50)	199/200 (99.5%) for the proportion of recordable eyes
Sullivan-Mee 2009[142]	98 (62/63)	Three subjects were excluded because reliable DCT readings could not be obtained for the second measurement set; the second set of ORA readings was lost as a result of computer malfunction; and one subject was post strabismus surgery (contraindication for inclusion in the study)
RT		
Johannesson 2008[210]	100 (150/150)	One eye for which not all six measurements were achieved. Mean quality value of measurement was also reported, which 'implied' a mean of 1.64 extra measurements needed. No eyes for which a mean IOP was not reported
Martinez-de-la-Casa 2005[190]	50 (18/36)	Estimate would seem to be the average across three examiners taking a measurement on 12 eyes. For valid measurement after two attempts it was 92% (33/36) and after three attempts it was 100%. A mean of 1.58 ± 0.67 attempts were needed to obtain a valid measurement
Martinez-de-la-Casa 2006[189]	100 (146/146)	No eyes were excluded because of poor or unreliable measurements
Pakrou 2008[167]	100 (143/143)	286/292 (97.9%) eyes had no error bar with RT (i.e. optimal SD) after a maximum of three attempts

Study	% recordable IOP (*n/N*)	Comments
TonoPen		
Bandyopadhyay 2002[197]	100 (208/208)	
Christoffersen 1993[216]	90 (132/146)	Reason for exclusion related to measurement for five patients (10 eyes) who were unable to cooperate with tonometry
Van der Jagt 2005[207]	100 (103/103)	
Transpalpebral		
Alvarez 2004[127]	91 (124/137)	13 subjects could not see the phosphene
Baskaran 2006[198]	95 (72/76)	Four patients were excluded because they could not see the phosphene ring during self-tonometry
Brigatti 2005[137]	76 (55/72)	Pressure phosphene could not be seen
Cheng 2005[176]	97 (60/62)	Two patients could not appreciate the phosphene phenomenon
Herse 2005[168]	95 (203/213)	Pressure phosphene could not be seen in 10/213 normal eyes
Lam 2004[172]	97 (99/102)	Three patients unable to perceive phosphenes
Li 2004[138]	95 (86/91)	
Naruse 2005[177]	93 (94/101)	'7 glaucoma patients were unable to detect the pressure phosphene and could not self-measure their IOP with the PPT'
Rai 2005[124]	95 (135/142)	

Appendix 6

Acceptability/practicality data by tonometer (*Chapter 5*)

Study	Reported (Y/N)	Acceptability Description	Reported (Y/N)	Practicality Description
DCT				
Doyle 2005[123]	N		Y	'We did not find the tonometer (DCT) easy to use'
Eichenbaum 2007[132]	Y	No difficulties with the introduction of DCT. Ready compliance when repeat measurement was required	N	
Johannesson 2008[210]	N		Y	0.84 extra measurements per eye were required for DCT
Ku 2006[215]	Y	Preference of tonometer: 36 (34%) prefer DCT, 15 (14%) GAT and 55 (52%) no preference. List of reasons given	N	
Lanza 2008[143]	Y	'Patient comfort was ensured' and 'pressure on the globe' was avoided	N	
Pache 2005[163]	Y	'GAT and DCT measurements were comparably well tolerated by the study participants'	N	
Salvetat 2007[144]	Y		Y	Measurements 'were not easy to perform and entailed a learning curve'
Vandewalle 2009[184]	Y	No patient reported more than moderate discomfort with any of the procedures	N	
NCT				
Cho 1997[173]	Y	The subjects were asked to comment on each technique. 'About 50% of subjects preferred NCT-2000 over the other tonometers' as air puff is 'weaker' and is non-contact tonometer. The remaining subjects had 'no particular comments on the test used'	Y	The operator found that NCT-2000 was the 'easiest and fastest' to operate
Mackie 1996[156]	N		Y	2, 3 and 3 minutes mean time for Pulsair 200, GAT and AO MkII, respectively
Ogbuehi 2008[194]	Y	Eight patients (11%) (Pulsair EasyEye) had anxiety about the procedure and needed to calm down	Y	IOP screening is quick with NCT
Ocuton S				
Marchini 2002[146]	Y	Safety was assessed clinically, including complaints reported by patients. Of the 80 subjects, 62 (78%) did not complain, 14 (17.5%) complained of foreign body sensation, 4 (5%) complained of burning. All lasted < 20 minutes and were resolved without consequence	Y	Participants were taught how to perform self-tonometry by practising at least three times

Study	Reported (Y/N)	Acceptability Description	Reported (Y/N)	Practicality Description
ORA				
Vandewalle 2009[184]	Y	No patient reported more than moderate discomfort with any of the procedures. No difference in patient preference was reported	N	
Perkins				
Eichenbaum 2007[132]	Y	'No difficulty with acceptance' although it is noted that 'the patients were not new to the practice'	N	
Jackson 1995[166]	Y	'All patients interviewed found tonometry acceptable'	Y	Time (mean) to measure IOP: Perkins: 96 seconds; GAT: 120 seconds.
RT				
Johannesson 2008[210]	N		Y	1.64 extra measurements per eye were required, increasing to 2.25 for thick CCT
Pakrou 2008[167]	Y	RT well tolerated by patients: 2 (5%) preferred GAT and 28 (74%) preferred RT and 8 (21%) had no preference. Mean response score for patient comfort was 2.03 in favour of RT ($p < 0.001$)	Y	1.36 extra measurements per eye for RT. Overall ease of use was reported, although it was noted that extra attention was required to position Icare and the operator correctly and to take the measurements as quickly as possible. However, no apparent learning effect when comparing first third of patients to final third. Also, no anaesthetic required. However, RT cannot be used for supine patients
Rehnman 2008[209]	Y	Stated that it was 'minimally invasive for the patient'	Y	'Very easy to use'
Roukonen 2007[160]	N		Y	In general, easy to use and no anaesthetic required. However, difficulty was reported with the required positioning of the tonometer
Sahin 2007[211]	N		Y	Noted that there is no requirement for anaesthetic and has disposable tips
van der Jagt 2005[207]	Y	10-point comfort scale used: mean (SD) 8.2 (1.0)	Y	'Easy to use and measurements could be performed in all subjects'
Vandewalle 2009[184]	Y	No patient reported more than moderate discomfort with any of the procedures	N	

Study	Reported (Y/N)	Acceptability Description	Reported (Y/N)	Practicality Description
TonoPen				
Bandyopadhyay 2002[197]	Y	Patient preference: 'which of these two procedures do you find more acceptable?': 30 (14.8%) preferred GAT, 64 (31.5%) preferred TonoPen, 100 (49.3%) found them equally acceptable, 9 (4.4%) had no preference	Y	Time (mean) to measure IOP: TonoPen, 15 seconds (SD 4); GAT, 54 seconds (SD 18)
Jackson 1995[166]	Y	'All patients interviewed found tonometry acceptable'	Y	Time (mean) to measure IOP: TonoPen, 50 seconds; GAT,120 seconds
Salvetat 2007[144]	Y	'TonoPen offered greater patient comfort' than Pascal	Y	TonoPen is 'easy to use, portable'
van der Jagt 2005[207]	Y	10-point comfort scale used: mean (SD) 7.4 (1.3)	Y	'Easy to use and measurements could be performed in all subjects'
Transpalpebral				
Brigatti 2005[137]	Y	No adverse events. Data not reported although 0–100 scale used	N	
Gunvant 2007[133]	Y	Analysis of covariance used to assess different levels of instruction/training/ experience resulting in no significant learning curve ($p = 0.14$)	Y	Clinicians can benefit from obtaining IOP data from between visits if home tonometry is used
Lam 2004[172]	Y	Rating scale (0–100), with the following definitions: 0, very great difficulty; 25, great difficulty; 50, some difficulty; 75, easy; 100, very easy to use at home. Results: easy to use PPT for home self-tonometry: 87.4 ± 16.3. Ease of perceiving phosphene: 78.0 ± 20.4	N	Initial training lasted 1 hour. The patient was then asked to practice for another hour
Li 2004[138]	Y	Grading: 1, very easy; 2, easy; 3, moderate; 4, difficult; 5, very difficult. Results: 70 (81%) described Proview as 'easy' or 'very easy', 6 (7%) described it as 'difficult'	N	
Morledge-Hampton 2006[136]	N		Y	Subjectivity of what a patient thinks a phosphene is makes it impractical for clinicians
Shemesh 2007[205]	Y	'With appropriate instruction, the device is easy to use and causes no discomfort'	N	
Tai 2005[125]	Y	92% rate as easy to use. 0–100 scale with mean of 79 for willingness to use at home	N	
van der Jagt 2005[207]	Y	10-point comfort scale used: mean (SD) 8.0 (0.9)	Y	'Easy to use and measurements could be performed in all subjects'

N, no; Y, yes.

Appendix 7

Reliability data by tonometer (*Chapter 5*)

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
DCT					
Doyle 2005[123]	DCT: two ophthalmologists; GAT: NR	DCT: two measurements; if reading ≥ 2 mmHg a third measurement was taken GAT: 2; if ≥ 2 mmHg of difference a third measurement was taken		GAT: 0.52 (SD) mmHg; DCT: 0.62 (SD) mmHg	
Fogagnolo 2010[218]	DCT: one experienced investigator per site; GAT: NR	DCT: measurements were repeated if Q > 3; this occurred in 8% of the readings and twice for 2% of the time GAT: 2; if ≥ 2 mmHg of difference a third measurement was taken	DCT: ICC 0.96 ($n = 350$ eyes) and test–retest variability was significant ($p = 0.01$, $F = 2.86 > 2.12$, 6 df)	DCT: CoV 5.0%; RC 3.24 mmHg	
Herdener 2008[159]	DCT: one experienced ophthalmologist; GAT: NR	DCT: NR; GAT: NR		DCT: short-term (same day) and long-term (2–8 days) reproducibility was 1.2 and 1.5 mmHg, respectively GAT: short-term (same day) and long-term (2–8 days) reproducibility was 1.1 and 1.2 mmHg, respectively	Sensitivity analyses excluding lower-quality DCT readings showed greater intraobserver reliability
Johannesson 2008[210]	DCT: one student; GAT: one clinician	DCT: 6; GAT: 6		95% CI (may be LoA) for RC: DCT ± 1.7 mmHg ($n = 149$); GAT ± 1.9 mmHg ($n = 150$)	IOP appeared to decrease (aside from tonometer) as multiple measurements were taken
Kaufmann 2004[104]	DCT: four examiners; GAT: four examiners	DCT: 3; GAT: 3	DCT: 0.65 mmHg residual variance estimate GAT: 1.10 mmHg residual variance estimate	DCT: 0.44 mmHg investigation plus interaction variance estimate GAT: 1.28 mmHg investigator plus interaction variance estimate	ANOVA model with examiner, subject and examiner–subject interaction as factors
Kotecha 2005[155]	DCT: one of two technicians; GAT: one optometrist	DCT: 3; GAT: 2	DCT: mean difference of 0.2 (95% LoA −4.9 to 5.3) mmHg. Based only on measurements 2 and 3 GAT: average readings between clinician and technician was mean difference of 0.4 (95% LoA −3.5 to 4.2) mmHg	Technician: DCT: RC 4.2 mmHg and 3.2 mmHg, respectively, for measurements 1 and 2, and 2 and 3; GAT: corresponding RC 1.6 mmHg Clinician: DCT: RC 3.3 mmHg and 2.6 mmHg, respectively; GAT: corresponding RC 1.7 mmHg	

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
Kotecha, 2010[129]	DCT: one optometrist and two others (non-ophthalmologists); GAT: NR	DCT: 3; GAT: 2	DCT: −0.20 (95% LoA ± 2.8) mmHg (between observer 1 and observers 2 and 3) GAT: mean difference of −0.80 (95% LoA ± 3.9) mmHg	Observer 1: DCT: 1.8 mmHg; GAT: 2.2 mmHg Observers 2 and 3: GAT: 2.3 mmHg; DCT: 2.0 mmHg ($n = 100$, all RC)	Only good-quality measurements accepted (DCT) – value 1 or 2
Pourjavan 2007[181]	DCT: one ophthalmologist; GAT: one ophthalmologist	DCT: 4; GAT: 2		DCT: ICC 0.78 (one measurement) and 0.91 (mean of three measurements)	
Roszkowska 2009[150]	DCT: one observer; GAT: NR	DCT: 2, GAT: 2		GAT: 0.8 mmHg, DCT: 1.4 mmHg (coefficient uncertain, probably RC)	
Sullivan-Mee 2009[142]	DCT: two optometrists; GAT: two optometrists	DCT: 2, GAT: 2	DCT: mean difference (SD) of 0.34 (1.16) (95% LoA −1.9 to 2.6) mmHg; CoV 6.6%; CCC 0.92 (95% CI 0.88 to 0.96) ($n = 60$ eyes) GAT: mean difference (SD) of 0.83 (1.14) (95% LoA −1.4 to 3.1) mmHg; CoV 7.8%; CCC 0.93 (95% CI 0.90 to 0.96)	Examiner 1 ($n = 30$ eyes): DCT: mean difference (SD) 1.18 (1.33) (95% LoA −1.4 to 3.8) mmHg, CoV 8.0%, CCC 0.92 (95% CI 0.86 to 0.97); GAT: mean difference (SD) 0.67 (1.54) (95% LoA −2.4 to 3.7) mmHg, CoV 11%, CCC 0.95 (95% CI 0.91 to 0.98) Examiner 2 ($n = 30$ eyes): DCT: mean difference (SD) 0.06 (1.62) (95% LoA −3.1 to 3.2) mmHg, CoV 9.3%, CCC 0.84 (95% CI 0.75 to 0.94); GAT: mean difference (SD) 0.64 (1.45) (95% LoA −2.2 to 3.5) mmHg, CoV 10%, CCC 0.85 (95% CI 0.75 to 0.95) Both examiners: DCT: RC of 2.0; GAT: RC of 2.5	RC calculated using measurements on both eyes
NCT					
AlMubrad 2008[193]	NCT: NR; GAT: NR	NCT: 4 (mean of last three used); GAT: 3		NCT: within-session mean difference (SD) 0.1 (1.2) and 0.2 (1.3) mmHg for the first two sessions, respectively (95% LoA −2.3 to 2.5 and −2.4 to 2.8 mmHg, respectively). Between-session 95% LoA −2.6 to 3.0 mmHg GAT: within-session mean difference (SD) 0.1 (1.1) and 0.0 (1.0) mmHg, respectively (95% LoA −2.2 to 2.3 and −2.0 to 2.0 mmHg, respectively). Between-session 95% LoA −2.2 to 2.8 mmHg	Intraobserver repeatability within session taken from second and third measurements (for both GAT and NCT) and between sessions 1 week apart. Unclear if a single observer or not

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
Hansen 1995[217]	NCT: one examiner; GAT: one examiner	NCT: 3; GAT: 3		NCT: 2.90 (Varmid), mean difference 0.10 (95% LoA −4.88 to 5.08) mmHg GAT: 0.98 (Varmid), mean difference 0.36 (95% LoA −2.76 to 3.48) mmHg	
Lafaut 2007[122]	NCT: one examiner; GAT: one examiner	NCT: 3; GAT: 3		NCT: within-session mean difference (SD) 1.2 (0.7) mmHg GAT: within-session mean difference (SD) 0.1 (0.2) mmHg	Pertains to subgroup only. Similar results for the other participants were also reported
Lam 2004[175]	NCT: possibly a consultant ophthalmologist but paper is unclear, GAT: NR	NCT: 3; GAT: 3		NCT: mean CoV (SD) 4.5% (3.4%) GAT: mean CoV (SD) 3.7% (1.8%)	
Mackie 1996[156]	NCT: one optometrist; GAT: one of three ophthalmologists	NCT: 4; GAT: 2		NCT: SDs of measurements from individual patients range from 2 to 6.7 mmHg	
Ogbuehi 2006[195]	NCT: one examiner; GAT: one ophthalmologist	NCT: 4 (only last three readings were averaged); GAT: 3. Readings taken in four sessions, two each for GAT and NCT		NCT: mean difference (SD) between both sessions of 0.1 (1.3) (95% LoA −2.5 to 2.7) mmHg GAT: mean difference (SD) between both sessions of 0.2 (1.4) (95% LoA −2.54 to 2.94) mmHg	
Ogbuehi 2008[194]	NCT: one ophthalmologist; GAT: one of three clinicians	NCT: 3; GAT: 3		NCT: within-session RC (first session) 1.8 mmHg and (second session) 1.7 mmHg; test–retest reproducibility 3.1 mmHg GAT: 1.7 (first session) and 1.9 (second session); test–retest reproducibility 2.5 mmHg	
Regine 2006[149]	NCT: two ophthalmologists; GAT: NR	NCT: 3; GAT: 3		NCT: RC 3.59 mmHg (right eye) GAT: RC 3.98 mmHg (right eye) Mean difference of 1.94 mmHg	Only 10 subjects
Tonnu 2005[152]	NCT: one examiner; GAT: one examiner	NCT: 3, GAT: 3		NCT: RC 3.2 mmHg GAT: RC 2.2 mmHg	

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
Ocuton S					
Marchini 2002[146]	Ocuton S: two operators and patient for self-tonometry; GAT: NR	Ocuton S: 3 measurements; if there was a difference ≥ 5 mmHg between one and the other two, a fourth measurement was taken GAT: NR	Ocuton S: first measurement between two observers 0.61 (95% CI 0.30 to 0.93) mmHg; second measurement between two observers 0.41 (95% CI 0.02 to 0.80) mmHg (both κ)	First observer: Ocuton S: 0.66 (95% CI 0.31 to 1.00) mmHg; second observer: Ocuton S: 0.42 (95% CI 0.06 to 0.78) mmHg (both κ) Self-tonometry: Ocuton S: mean difference (SD) 0.6 (2.1) (95% CI LoA −3.6 to 4.8) mmHg Each measurement was the mean of three consecutive readings	
Wells 2003[170]	Ocuton S: patient; GAT: 1 researcher	Ocuton S: 3, GAT: NR		ORA: RC 9.17 mmHg	
ORA					
Ehongo 2009[183]	ORA: NR; GAT: NR	ORA: 2 (8 readings); GAT: 2		ORA: before and after anaesthetised right eye were 1.44 and 1.11 (RE) mmHg	
Kotecha 2006[110]	ORA: one optometrist; GAT: NR	ORA: 3; GAT: 2		Appear to be intra (not explicitly stated). ORA: CoV 8.9% ($n = 144$)	
Kotecha 2010[129]	ORA: one optometrist and two others (non-ophthalmologists); GAT: NR	ORA: 3; GAT: 2	ORA: mean difference of 0.30 (95% LoA −4.2 to 3.6) mmHg (between observer 1 and observers 2 and 3)	ORA: observer 1: RC 4.4 mmHg; observers 2 and 3: RC 4.3 mmHg ($n = 100$ eyes)	Only good-quality measurements accepted (ORA)
Kynigopoulos 2008[185]	ORA: one experienced technician; GAT: NR	ORA: 4; GAT: NR		ORA: RC 2.22 mmHg; CoV 6.5%; ICC 0.89 (lower 95% CI 0.82) ($n = 49$)	
Sullivan-Mee 2009[142]	ORA: two optometrists; GAT: NR	ORA: 2 (8 readings); GAT: 2	ORA: mean difference (SD) of −0.04 (1.73) (95% LoA −3.4 to 3.4) mmHg; CoV 9.6%; CCC 0.89 (95% CI 0.84 to 0.94) ($n = 60$ eyes)	Examiner 1 ($n = 30$ eyes): ORA: mean difference (SD) 0.48 (1.66) (95% LoA −2.8 to 3.7) mmHg; CoV 9.9%; CCC 0.92 (95% CI 0.86 to 0.98) Examiner 2 ($n = 30$ eyes): ORA: mean difference (SD) 0.10 (1.75) (95% LoA −3.3 to 3.5) mmHg; CoV 10.1%; CCC 0.91 (95% CI 0.71 to 0.96) Both examiners: ORA: RC 3.9; GAT: RC 2.0	RC calculated using measurements on both eyes
RT					
Abraham 2008[165]	RT: two ophthalmologists; GAT: two ophthalmologists	RT: 6 (highest and lowest discarded); GAT: 2		RT: RC 2.38 mmHg	Unclear if RC based on data for one or both examiners
Davies 2006[154]	RT: NR; GAT: one of two optometrists	RT: 2; GAT: 2		RT: mean difference 0.46 (SD 2.61) (95% LoA ± 5.11 mmHg)	

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
Johannesson 2008[210]	RT: one student; GAT: one student	RT: 6; GAT: 6		RT: ± 2.0 mmHg (*n* = 150) 95% CI (may be LoA) for repeatability GAT: ± 1.9 mmHg (*n* = 150) 95% CI (may be LoA) for repeatability	IOP appeared to decrease (aside from tonometer) as multiple measurements were taken
Martinez-de-la-Casa 2005[190]	RT: one experienced ophthalmologist; GAT: NR	RT: 3; GAT: 3	RT: CC 0.82 (range 0.62–0.94) (*n* = 12 eyes); CoV 8.9%	RT: first examiner: CC 0.82 (range 0.62–0.94); second examiner: CC 0.73 (range 0.46–0.90); third examiner: CC 0.87 (range 0.72–0.96) 'Intra-subject variation coefficient': 8.9%	Only 12 observations
TonoPen					
Bafa 2001[153]	TonoPen: one examiner; GAT: one examiner	TonoPen: 4; GAT: 1		TonoPen: CoV 5% to < 10% (*n* = 2); CoV < 5% (*n* = 97)	
Horowitz 2004[169]	TonoPen: two ophthalmologists; GAT: NR	TonoPen: 2; GAT: 2		TonoPen: mean difference (SD) 0.74 (1.50) (95% LoA −2.26 to 3.74) mmHg; ICC 0.97 GAT: mean difference (SD) 0.13 (−1.75) (95% LoA −3.34 to 3.63) mmHg; ICC 0.95	
Tonnu 2005[152]	TonoPen: one examiner; GAT: one examiner	TonoPen: 3; GAT: 3		TonoPen: 4.3 (RC) mmHg GAT: 2.2 (RC) mmHg	Only measurement < 5% accepted for TonoPen
Transpalpebral					
Alvarez 2004[127]	Transpalpebral: patient; GAT: one trained clinician	Transpalpebral: each subject tried five different Proview devices, five measurements with each device (25 total). Up to 10 attempts per device were permitted to achieve five successful measurements GAT: 2 or 3 if > 2 mmHg difference for first 2		Transpalpebral: within-subject and within-device variance was 3.4 mmHg (620 means considered = means of each of five devices for each of 124 subjects) (18.2% of variation). RC of 5.11 mmHg	Five devices were all Proview and varied in a very minor way. Appeared to be from ANOVA analysis as variance is partitioned
Brigatti 2005[137]	Transpalpebral: patient; GAT: one physician	Transpalpebral: 3; GAT: 2 sets of 3 readings		Transpalpebral: mean difference (SD) 1.76 (1.76) mmHg; 0.71 (CC) for first and third readings GAT: mean difference (SD) 1.73 (1.42) mmHg; 0.94 (CC)	
Herse 2005[168]	Transpalpebral: patient; GAT: one experienced clinician	Transpalpebral: 3; GAT: 3		Transpalpebral: 4.21 (RC) mmHg; mean difference (SD) 0.1 (2.1) mmHg GAT: 0.96 (RC) mmHg; mean difference (SD) 0.1 (0.6) mmHg	

Study	Examiner number/type	Number of measurements	Interobserver	Intraobserver	Comments
Lam 2004[172]	Transpalpebral: patient; GAT: two of the investigators	Transpalpebral: 3; GAT: 3. Two different visits 1 week apart. Patients asked to practice at home		Transpalpebral: CoV 7.3% ($n=194$) GAT: CoV 4.4% (unclear if intraobserver)	Median of the three readings was used for comparison among tonometers. Random order. Masked investigators
Morledge-Hampton 2006[136]	Transpalpebral: one ophthalmologist; GAT: NR	Transpalpebral: 2; GAT: 3. First and second measurements used in main analysis		Transpalpebral: 0.82 (CC) for first and second readings GAT: 0.97 (CC) for first and second readings May not be intraobserver as number of observers not reported	The third GAT reading was compared with the first two, which showed evidence of a tonometry effect (i.e. lowering of IOP)
Naruse 2005[177]	Transpalpebral: patient; GAT: one ophthalmologist	Transpalpebral: 3 (sets of five consecutive readings), GAT: 3		Transpalpebral: mean difference (SD) 0.2 (0.54) (95% LoA −0.83 to 1.27) mmHg ($n=26$); RC 1.07 mmHg	Second and third measurements compared
Rai 2005[124]	Transpalpebral: patient and technician; GAT: one examiner	Transpalpebral: 3; GAT: NR		Transpalpebral: 0.83 (CC) patient GAT: 0.78 (CC) examiners	
Troost 2005[158]	Transpalpebral: two examiners; GAT: one examiner	Transpalpebral: 3,; GAT: 3	Transpalpebral: mean difference of −1 (95% LoA −8 to 6) mmHg		

ANOVA, analysis of variance; CC, correlation coefficient; df, degrees of freedom; kappa, kappa statistic; NR, not reported; Q, quality score; Varmid, variance of difference between the middle reading and the average of first and last readings.

Appendix 8

Using statistical models to develop monitoring criteria (*Chapter 6*)

This appendix describes possible models of IOP or visual field or RNFL parameter changes over time, taking into account both measurement error and long-term true changes.

The apparent change in measurements over time can be thought of as being made up of three components:

1. a short-term variation: a combination of analytic variability and day-to-day biological fluctuation around a stable average
2. the average true long-term change of the whole group and
3. a variation in long-term true change between individuals (as would be seen with the theoretical average of a large number of measurements per individual).

The models that will be used in the project are extensions of the simple model of measurement error that assumes no trend. This simple model is described below using IOP measurement. The same model can be applied to visual field or RNFL measures.

Measurement error with no trend (simple model)

Let T_i be the true long-term IOP for individual i but let measurements of true IOP be imperfect. The measurement of IOP IOP_{it} for patient i at time t will have a measurement error, ω_{it}, so that for times t_1 and t_2:

$$IOP_{i1} = T_i + \omega_{i1}, \ IOP_{i2} = T_i + \omega_{i2} \qquad \text{[Equation 17]}$$

or more generally:

$$IOP_{it} = T_i + \omega_{it} \qquad \text{[Equation 18]}$$

Assume that the true value T_i is normally distributed, $N(\mu_T, \sigma_T^2)$, and that the error ω_{it} is also normally distributed: $N(\mu, \sigma_\omega^2)$. Also assume that T_i and ω_{it} are independent and the error terms ω_{it} for $t = 1, \dots n_i$ are independent from each other.

In this model σ_T^2 is the *between-subject* variability and σ_ω^2 is the *within-subject* variability.

Differences in IOP between time points can be used to estimate the within-person variability. For example, using baseline and subsequent times we can compute σ_ω^2 as:

$$\text{Var}(\Delta IOPi) = \text{Var}(IOP_{it}) + \text{Var}(IOPi_0) = 2\sigma_\omega^2 \qquad \text{[Equation 19]}$$

Therefore, we can estimate σ_ω^2 (SD_ω^2) as the square of the SD of any difference divided by 2.

Extension 1: linear model with an increasing variance

As the true IOP may change with time, the simple model needs to be extended. One approach is to assume that each individual has a linear increase over time but that that increase varies between individuals, that is, each person i has a rate of increase b_i with time and this follows a normal distribution $N(\beta, {}_1{}^2)$. In this case the model becomes:

$$IOP_{it} = T_i + \beta_{it} + \omega_{it} \qquad \text{[Equation 20]}$$

with T_i again being the true IOP measurement at baseline. Note that two special cases of this model are (1) $N(0,\sigma_1{}^2)$, with no average increase of the population but allowing individual changes, and (2) $N(\beta,0)$, that is, everyone has the same change over time.

There are now three terms and three components for the variance: the population baseline variation (between subject), the individual variation in trend and the measurement error (within-subject variability), that is:

$$var(IOP_{it}) = \sigma T^2 + \sigma_1{}^2 t^2 + \sigma_\omega{}^2 \qquad \text{[Equation 21]}$$

with the formula for the variance of the change from baseline equal to:

$$var(\Delta IOP_{it}) = \sigma_1{}^2 t^2 + 2\sigma_\omega{}^2 \qquad \text{[Equation 22]}$$

Extension 2: non-linear increase in variance (direct method)

Extension 1 assumes that the IOP and its SD will increase approximately linearly with time. This is likely to be true only for short ranges of time. An alternative is to use a separate parameter for each time point, that is, for each time point there is an average increase, and a distribution to that increase, so we have b_{it} for each time t and these follow a normal distribution $N(\beta_t, \sigma_1{}^2)$. The model now becomes:

$$IOP_{it} = T_i + \beta_{it} + \omega_{it} \qquad \text{[Equation 23]}$$

There are three terms and three components for the variance: the population baseline variation (between subject), the individual variation in trend and the measurement error (within-subject variability), that is:

$$var(IOP_{it}) = \sigma_T{}^2 + \sigma_{1t}{}^2 + \sigma_\omega{}^2 \qquad \text{[Equation 24]}$$

With this model the formula for the variance of the change from baseline is derived as:

$$var(\Delta IOP_{it}) = \sigma_{1t}{}^2 + 2\sigma_\omega{}^2 \qquad \text{[Equation 25]}$$

Hence, we can estimate the variance in long-term true change by subtracting twice the within-person variability, that is:

$$\sigma_{1t}{}^2 = var(\Delta IOP_{it}) - 2\sigma_\omega{}^2 \qquad \text{[Equation 26]}$$

where $\sigma_\omega{}^2$ is estimated from short-term variability studies.

Appendix 9

Optimal frequency of monitoring intraocular pressure and visual fields: further analysis (*Chapter 6*)

Parameter estimates obtained using the direct method

The direct method of estimation has a total of two parameters per time point, which is less useful for prediction as well as for the inclusion of its results into economic modelling. In this particular example, it has the disadvantage of estimating negative variances (meaningless) because of the way that the variance at different time points is calculated (noise subtracted from observed variance). For these reasons we focused on the three models obtained from the linear random effects model.
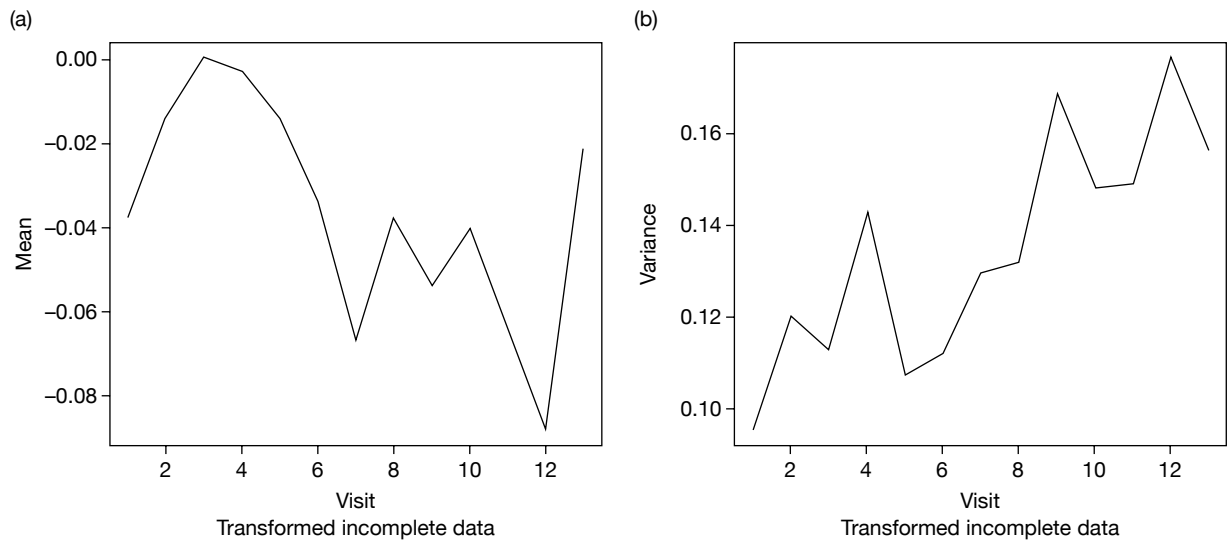
FIGURE 67 Mean and variance of transformed IOP change score from baseline using the non-imputed data. (a) Mean of IOP change from baseline; (b) variance of IOP change from baseline.
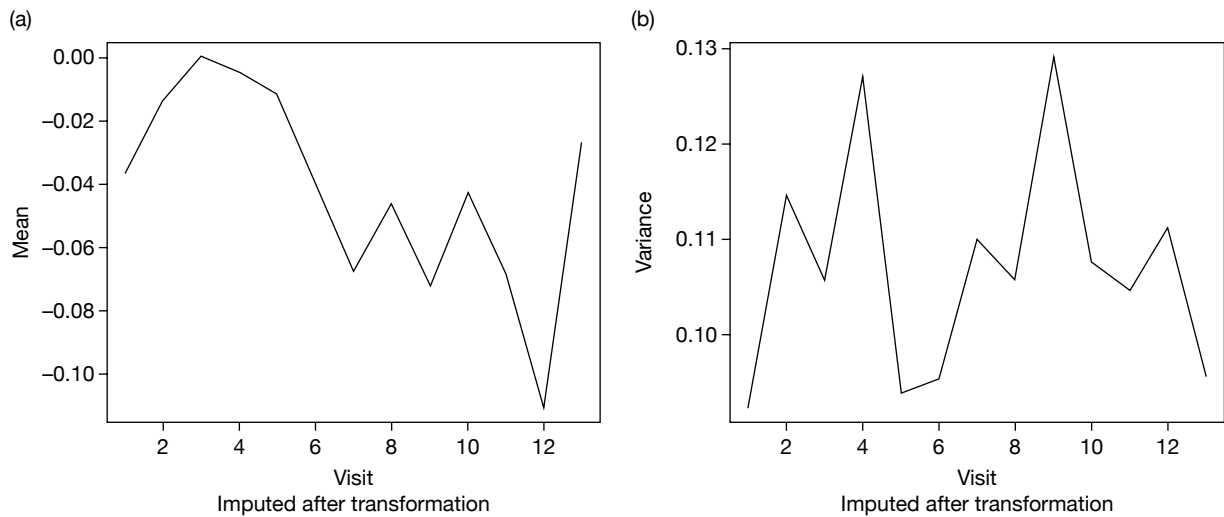
FIGURE 68 Mean and variance of transformed IOP change score from baseline using the (incomplete) median imputation data. (a) Mean of IOP change from baseline; (b) variance of IOP change from baseline.
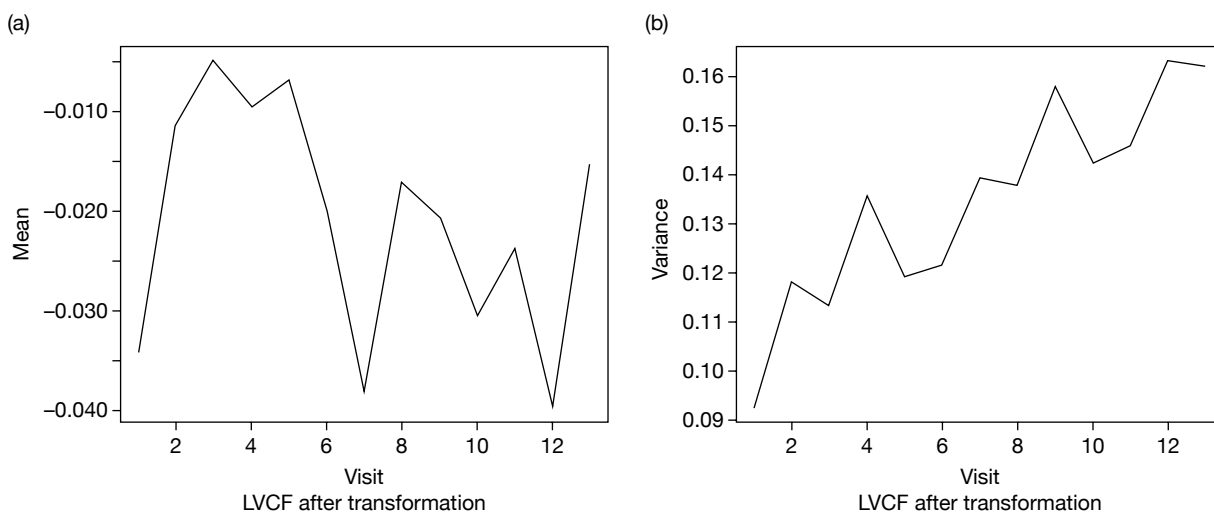
FIGURE 69 Mean and variance of transformed IOP change score from baseline using the LVCF imputed data. (a) Mean of IOP change from baseline; (b) variance of IOP change from baseline.
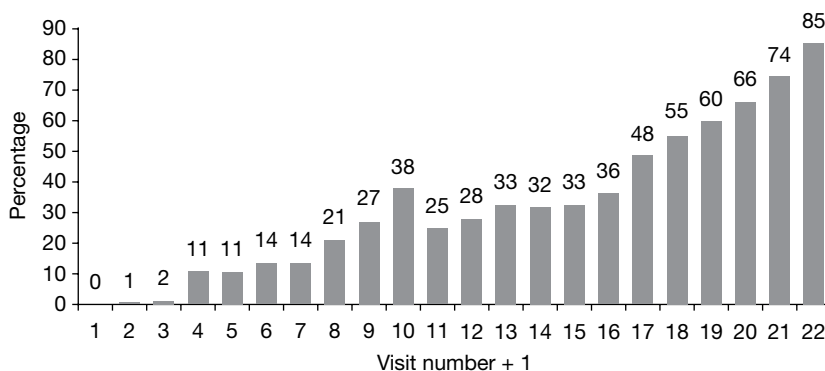
FIGURE 70 Missing data distribution for the placebo group, Rotterdam data set.

TABLE 68 Direct method outputs for transformed IOP

(a) Model 1: using LVCF to impute missing data

	Visit number, t^a												
Output	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Mean change	−0.034	−0.011	−0.005	−0.010	−0.007	−0.020	−0.038	−0.017	−0.021	−0.031	−0.024	−0.040	−0.015
Variance of change	−0.008	0.018	0.013	0.035	0.019	0.021	0.039	0.037	0.057	0.042	0.045	0.063	0.061

a Visit frequency every 4 months.
Baseline mean = 4.97.
Baseline variance = 0.0556.
Noise = 0.0503.

(b) Model 2: using the incomplete median imputed data

	Visit number, t^a												
Output	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Mean change	−0.037	−0.013	0.000	−0.005	−0.012	−0.040	−0.068	−0.046	−0.072	−0.043	−0.069	−0.111	−0.027
Variance of change	−0.014	0.008	−0.001	0.021	−0.013	−0.011	0.004	−0.001	0.023	0.001	−0.002	0.005	−0.011

a Visit frequency every 4 months.
Baseline mean = 4.97.
Baseline variance = 0.0527.
Noise = 0.0532.

(c) Model 3: using incomplete (non-imputed) data

	Visit number, t^a												
Output	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Mean change	−0.037	−0.014	0.001	−0.003	−0.014	−0.034	−0.067	−0.038	−0.054	−0.040	−0.063	−0.088	−0.021
Variance of change	−0.003	0.022	0.015	0.044	0.009	0.014	0.032	0.034	0.071	0.050	0.051	0.079	0.058

a Visit frequency every 4 months.
Baseline mean = 4.97.
Baseline variance = 0.0568.
Noise = 0.049.

TABLE 69 Empirical distribution from the established betas (β)

Slope value (β)	Probability, i.e. $P(X \leq x)$ [using normal $F(x)$]	Empirical probability, i.e. $\hat{F}(x) = \%$ betas $\leq x$
−0.00075	0.0000	0
−0.00070	0.0001	0.0065
−0.00065	0.0004	0.0065
−0.00060	0.0009	0.0131
−0.00055	0.0021	0.0196
−0.00050	0.0047	0.0261
−0.00045	0.0098	0.0392
−0.00040	0.0192	0.0392
−0.00035	0.0353	0.0392
−0.00030	0.0613	0.0588
−0.00025	0.1001	0.0915
−0.00020	0.1545	0.1307
−0.00015	0.2255	0.1895
−0.00010	0.3120	0.3203
−0.00005	0.4104	0.3987
0.00000	0.5147	0.5294
0.00005	0.6181	0.6993
0.00010	0.7136	0.7516
0.00015	0.7960	0.8039
0.00020	0.8624	0.8431
0.00025	0.9122	0.8824
0.00030	0.9472	0.9150
0.00035	0.9701	0.9477
0.00040	0.9840	0.9542
0.00045	0.9920	0.9673
0.00050	0.9962	0.9804
0.00055	0.9983	0.9869
0.00060	0.9993	0.9935
0.00065	0.9997	1

Appendix 10

Systematic review of economic evaluations of surveillance strategies for ocular hypertension: methods and results (*Chapter 8*)

Aims

The aim was to systematically review the available evidence on the cost-effectiveness of strategies for monitoring individuals with OHT who are at risk of developing glaucoma and to inform the methodology of the proposed economic evaluation. Studies that considered the UK NHS perspective were of particular interest.

Methods

Search strategy for identification of published reports

Extensive sensitive electronic searches were conducted to identify reports of published and ongoing studies of economic evaluations of SOH. In addition, reference lists of retrieved papers were scrutinised to identify additional potentially relevant studies.

The databases searched were MEDLINE (1950 to October Week 4 2010), EMBASE (1980 to Week 41 2010), MEDLINE In-Process & Other Non-Indexed Citations (25 October 2010), Health Management Information Consortium (1979 to October 2010), the NHS Economic Evaluations Database (NHS EED) (October 2010) and the HTA database for relevant evidence syntheses (October 2010).

Full details of the search strategies used are reproduced in *Appendix 1*.

Selection process

Inclusion criteria

■ Interventions: surveillance strategies of patients with OHT.
■ Study population: adults.
■ Study type: full economic evaluations – studies that compare the costs and outcomes of two or more strategies for monitoring individuals with OHT. An eligible strategy can include a situation in which no active or organised monitoring service is available.
■ Outcomes: studies had to include either costs (regardless of the way that these were estimated) or effects (regardless of the way that these were estimated).

Screening

Two reviewers (KB, AA-B) independently scanned titles and abstracts for inclusion of potentially relevant studies. Full-text copies of all potentially relevant reports were retrieved. The full-text reports were assessed against the inclusion criteria by two reviewers independently (AT, AA-B). Any disagreements were resolved by consensus or arbitration by a third party.

Results

In total, 161 papers were identified from the initial search. Of these, 11 were selected as potentially relevant and full-text copies were retrieved. On screening the full text of these retrieved studies; however, no studies were identified that fulfilled the inclusion criteria for the systematic review. A list of excluded full-text studies can be found in *Appendix 2*. A flow diagram of the selection process is shown in *Figure 71*.
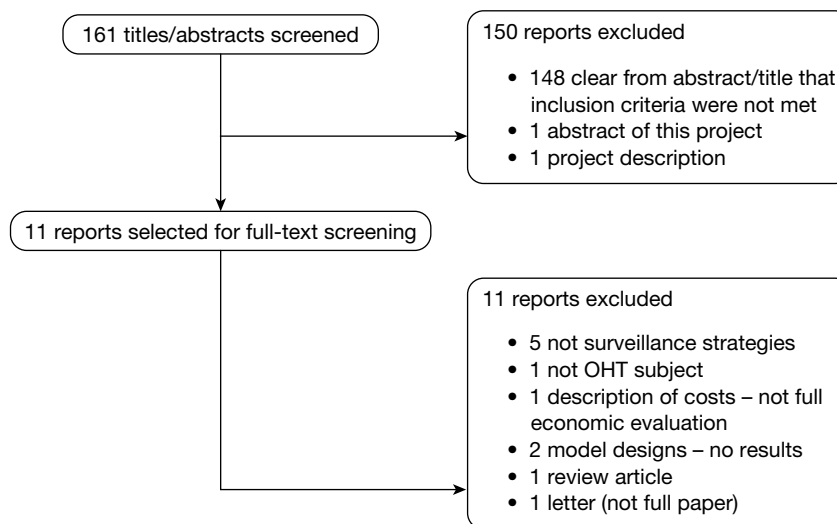
FIGURE 71 Flow diagram of the selection process.

Appendix 11

Project protocol

1 Title: Optimal Surveillance Regimes for Individuals with Ocular Hypertension: Modelling and Economic Evaluation

Protocol version 3, 19 July 2010

2 BACKGROUND

2.1 Burden of Glaucoma

Glaucoma is a chronic progressive optic neuropathy leading to impaired vision and sometimes blindness if untreated. The underlying pathophysiology is an accelerated loss of retinal ganglion cells, compared with the normal ageing process, leading to a characteristic pattern of retinal nerve fibre layer atrophy and optic nerve head changes which ultimately result in visual loss. Open angle glaucoma (OAG) is the most common form of glaucoma, affecting about 2% of the population aged over 40 in the UK. In the UK glaucoma is second to age related macular degeneration as a main cause of blindness.[1–3]

OAG is diagnosed primarily by glaucomatous optic neuropathy (characteristic changes of the optic disc) and a compatible visual field defect, in the presence of an open, normal appearing, anterior chamber angle. Early diagnosis is difficult as there is an overlap between normal and glaucomatous change. Progressive changes determine whether OAG is truly present. OAG is treated by interventions to lower the IOP. These include a variety of topical ocular hypotensive medications (eye drops), or laser trabeculoplasty or glaucoma drainage surgery or a combination of these.

2.1.1 Ocular Hypertension as a risk factor for OAG

Risk factors for developing OAG include raised IOP, greater age, non-white ethnicity (those of black ethnicity being a higher risk group), having a first-degree relative with OAG, myopia and diabetes.[4] IOP is the most important risk factor and is the only one which is treatable. Ocular hypertension, generally defined as an IOP of ≥ 21 mmHg (2 standard deviations above the mean) used to be considered as a part of the definition of OAG, but population studies have consistently found that many people with OAG have an IOP below this level.[5–9] The risk of developing OAG, and for worsening of existing disease, increases with increasing IOP,[10–12] and those presenting with advanced glaucoma at diagnosis have higher IOPs.[13] The population prevalence estimates for ocular hypertension range from 4.5% to 9.4% for those aged over 40[14] with prevalence increasing with age.[14,15]

The effectiveness of treatment in ocular hypertension has been evaluated in two large trials of medical treatment (Ocular Hypertension Treatment Study [OHTS][16] and the European Glaucoma Prevention Study [EGPS]).[17] OHTS, an unmasked randomised controlled trial, involved 1637 participants, 818 participants treated with any commercially available ocular hypotensive drugs (eye drops) to achieve 20% reduction from baseline IOP and an untreated control group. OHTS demonstrated that medical treatment to lower IOP reduced the incidence of the development of a glaucomatous visual field defect or glaucomatous optic neuropathy from 9.4% to 4.5% over a

five year period in people with ocular hypertension (defined in OHTS as IOP ≥ 24 mmHg). EGPS, a multi-centre randomised double masked controlled trial randomised 1081 participants aged ≥ 30 years to a topical ocular hypotensive eye drop (dorzolamide) or placebo eye drop. EGPS failed to detect a statistically significant difference between treatment and placebo in reducing the incidence of OAG. In this study, treatment reduced the IOP by 15–22% throughout the 5 years of the trial, and interestingly an IOP lowering effect of between 9 and 19% reduction in IOP was observed in the placebo group.[17]

Combining these data and data from eight smaller trials a Cochrane systematic review and meta-analysis by Vass and colleagues evaluating different topical medications against placebo or untreated controls showed about a 40% reduction in the incidence of glaucomatous visual field defects with treatment at five years (odds ratio 0.62 (95% confidence interval 0.5 to 0.8) for people with ocular hypertension.[18]

In the UK there is great variability in clinical practice regarding standards of care (follow-up and treatment) of ocular hypertension. Across Europe the annual direct costs associated with managing ocular hypertension and glaucoma are estimated, in 2003, as €455 per person for ocular hypertension increasing to €595 for early glaucoma rising to €969 for end stage disease.[19] These estimates are however very imprecise as they are based on retrospectively collected patient data from only 194 patient records collected across four countries, including three centres in the UK. Costs increased with increasing disease stage with the main cost driver of the total direct healthcare cost of glaucoma care being the cost of glaucoma medications. In the USA a cost–utility analysis, taking a societal perspective, using the OHTS data reported that treatment for individuals with an IOP ≥ 24 mmHg and a ≥ 2% annual risk of developing glaucoma treatment would be cost-effective if society were willing to pay $50,000 for an additional quality adjusted life year (QALY).[20]

While it is important to treat early disease to reduce the costs and disability associated with OAG, only a minority of those with ocular hypertension develop OAG and of these not all are at risk of visual impairment in their lifetime. The challenge is therefore to identify the groups most at risk of developing significant OAG, and to determine effective and efficient monitoring criteria to identify those individuals who are most likely to benefit from treatment.

2.1.2 Risk prediction for those with ocular hypertension

The OHTS study group identified baseline factors that predicted the development of OAG based on the entire OHTS sample, both treated and untreated individuals. These were older age, larger vertical or horizontal cup to disc ratios, higher IOP, greater pattern standard deviation of visual field loss and thinner corneas (central corneal thickness).[21] Medeiros and colleagues[22] validated the OHTS-derived predictive models in an independent observational cohort of 126 untreated subjects with ocular hypertension (Diagnostic Innovations in Glaucoma Study).[23]

A predictive model[24] has recently been published building upon that of Medeiros and colleagues.[22] This most recent prediction tool was developed and validated using data from the two largest treatment trials for ocular hypertension published to date (OHTS and EGPS). Briefly, the tool was developed from the observation group of the OHTS and then tested on the placebo group of EGPS for discrimination ability, and for systematic overestimation/ underestimation of predicted risk. A pooled prediction tool was then developed based on the observation group of the two studies, this allows the estimation of the five-year risk of an ocular hypertensive individual developing OAG. The risk tool incorporates the following criteria – age, IOP, central corneal thickness, vertical cup to disc ratio (a measure of glaucomatous optic

neuropathy) and pattern standard deviation (a measure of visual field loss). Based on this model the group have made freely available a quantitative risk calculator which can be accessed at http://ohts.wustl.edu/risk.

Validation of models by their creators typically produce overestimates of their performance.[25] We therefore propose to systematically identify and review existing risk prediction tools, and validate the performance of the most promising models (including that of OHTS/EGPT collaboration) in independent datasets. Validation of several models in the same data will facilitate comparison of predictive ability and transportability.

2.1.3 Current detection and monitoring strategies for ocular hypertension in the UK

In the UK, the majority of people are identified as ocular hypertension during a 'sight' test usually to obtain glasses at a community based optometrist. People aged 60 years and older, and people with a self reported family history of glaucoma 40 years and older, qualify for a free 'sight test' under the General Ophthalmic Services in England and in Scotland, since April 2007, all are eligible for a free test. The constituents of this 'sight' test are historic involving an examination of the fundus and the optometrist may elect to measure the IOP and test for a visual field defect. In the UK there is considerable debate about the role and optimal organisation of a monitoring service for those at risk of glaucoma.

In England a recent report on commissioning community eye care services has identified that Primary Care Trusts should commission high quality cost-effective services to offer patients sustainable choices close to home.[26] Guidelines for the management of ocular hypertension and OAG by the National Institute for Health and Clinical Excellence (NICE) have been recently published. Further research was recommended to evaluate the clinical effectiveness and cost effectiveness of using different monitoring intervals to detect disease progression.[28]

The issue of monitoring people with ocular hypertension is not to measure changes in the level of IOP in response to treatment but initially: (a) to establish that ocular hypertension is truly present and not a measurement error; then (b) to decide whether monitoring is necessary. One would wish to monitor those individuals at risk of developing glaucoma (with a measurement of IOP, visual field loss and optic nerve assessment) and to decide not to monitor those with negligible risk; (c) to establish whether prophylactic treatment is advisable; and (d) to detect those individuals progressing to glaucoma in a timely manner.

There is currently insufficient evidence to guide clinicians with respect to the optimal inter-test spacing and frequency of testing necessary to identify clinically significant changes in IOP or the optimal frequency and spacing of testing required to detect clinically significant changes in the visual field or optic nerve over a relevant time period, which could be up to five years. There is uncertainty whether delaying treatment for those with ocular hypertension and in some cases missing the conversion to glaucoma is deleterious to patient reported outcome. It is widely accepted that monitoring should include a measure of visual function, commonly standard automated perimetry, and a measure of structural changes in the optic nerve and retinal nerve fibre layer (RNFL) to estimate RNFL loss that is characteristic of glaucoma. This may be by clinical examination, photography or automated estimates of RNFL loss. Commercial instruments such as the Heidelberg Retina Tomograph (HRT),[28] the GDx, and a modified version GDX VCC[29] are available providing quantitative structural information of the RNFL. These instruments are used in some centres for glaucoma detection and to monitor progression.

3   PLANNED INVESTIGATION

3.1   Aim

To validate a predictive model to estimate the risk of developing open angle glaucoma (OAG) for individuals with ocular hypertension (raised intraocular pressures [IOP]), and to develop models to determine optimal thresholds for surveillance, and frequency for monitoring according to IOP, patient and clinical characteristics, taking into account NHS and societal costs, the health and wider benefits, as well as public preferences for process of care, health, and other outcomes.

The objectives are to:

1. Identify and validate the most relevant prediction tool(s) for the risk of developing OAG in individuals with ocular hypertension;
2. Determine thresholds for initiating surveillance;
3. Determine the optimal monitoring criteria (tests, frequency of monitoring, and determinants of progression to OAG);
4. Develop feasible, preferable and reliable testing and monitoring regimes for people with ocular hypertension in a UK context;
5. Model the impact of different plausible surveillance regimes for monitoring ocular hypertension on costs, diagnostic yield, health outcomes, as well as public preferences for process of care, health, and other outcomes and relative efficiency;
6. Make recommendations for surveillance regimes for ocular hypertension; and
7. Identify future research needs.

The research involves three linked work packages:

(A) Risk prediction: (Objective 1 and 2) Led by Deeks, Birmingham;
(B) Monitoring criteria: (Objective 3) Led by Perera (Oxford), collaborating with Crabb/Garway Heath (City/MEH University);
(C) Modelling evaluation of alternative surveillance regimes (Objectives 4–7): Led by Burr, Aberdeen.

3.2   Research Methods

This research aims to determine effective and efficient monitoring criteria to identify those individuals who will most benefit from treatment as well as to identify those who may not need treatment or follow-up.

The choice of monitoring tests and criteria should consider four factors:[30]

1. Clinical validity: The monitoring strategy (including test types and frequency of testing) should be either a measure of the clinically relevant outcome or a good predictor of the clinically relevant outcome;
2. Responsiveness: The test should change promptly in response to changes in therapy;
3. Large signal to noise ratio: The test should differentiate clinically important changes over time from background measurement variability (short-term biological fluctuations and technical measurement error);
4. Practicality: Ideally the test should be non-invasive, cheap, and simple to do. Ideally test results should be immediately available and suitable for patient self-monitoring.

Several approaches will be used across three interlinked work packages to achieve our objectives: expert and consumer opinions; population preferences; systematic reviews; individual patient data analysis (and possible meta-analysis); and decision analytic modelling.

Information from the *systematic reviews* will inform the clinical testing strategy in terms of validity and practicality, specifically looking at the reliability of a range of IOP measurement devices (tonometers). Goldmann applanation tonometry, a contact tonometer, is currently the most widely used tonometer by ophthalmologists, but if surveillance is to be considered in primary care, tonometers which do not touch the cornea may be more applicable. In addition a very thick or thin cornea can lead to measurement error in tonometry, including Goldmann applanation tonometry. New tonometers are available which account for the biomechanical properties and thickness of the cornea. In addition, non-invasive self measurement devices are in development and may be highly appropriate and relevant as monitoring devices. The systematic review (section **3.2.4 component ii**) is thus an important component to determine reliable and acceptable tonometers for a primary care setting which might be more acceptable to service users than hospital based monitoring.

It is important to note, however, that the individual patient data to be used in our analysis (section **3.2.2**) includes IOP measured by Goldmann applanation tonometry in all datasets and therefore the proposed analyses of risk prediction and optimal monitoring frequency (sections **3.2.3**, and **3.2.4**) will be informative about the comparative reliability of alternative tonometers.

The proposed *individual patient data analysis* (section **3.2.4**) will inform on the biological variability in IOP, and the responsiveness and the signal–noise ratio of the tests of visual function and optic nerve damage. This information will then be used in an economic modelling valuation, together with estimates of costs and benefits derived from the reviews and the preference surveys, to determine effective and efficient monitoring regimes for individuals with ocular hypertension.

The proposed *population preference study* (section **3.3.5 component i**), using discrete choice experiment methodology, will inform us about preferences for alternative monitoring regimes, including preferences for where monitoring might best be placed (hospital eye service or community optometry, or alternative community based setting) and on the likely uptake of alternative ways of organising a service.

3.2.1 Systematic review strategy

Extensive electronic searches will be conducted to identify reports of published, unpublished and ongoing studies and will include abstracts from conference proceedings and other grey literature sources. The search strategies will be designed to be highly sensitive, including appropriate subject heading and text word terms that reflect both the clinical content and type of study required for each component of the research project. This will involve identification of:

- studies that report on risk prediction tools (Work Package A);
- studies that report on the reliability of the selected IOP measurement techniques (Work Package B; Component ii);
- economic evaluations of surveillance programmes for ocular hypertension (Work Package C; Component ii)
- focussed searches to identify parameter values for the economic model (Work Package C; Component ii)

There will be no language restriction but searches will be restricted to 1987 onwards for the IOP measurement techniques (to reflect technological advances). The following databases will be

searched for all components: Medline (1966 onwards), Medline Extra, Embase (1980 onwards), Biosis (1985 onwards), Science Citation Index (1981 onwards), Conference Papers Index (2005 onwards), ISI Proceedings (2005 onwards) and ZETOC Conference Abstracts (2005 onwards). The Health Management Information Consortium Database (1979 onwards) and NHS Economic Evaluations Database (NEED) will be searched for economic literature. The Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effectiveness (DARE), and HTA database will be searched for relevant systematic reviews and HTA reports and the National Research Register, UKCRN Portfolio Database, Current Controlled Trials and Clinical Trials for ongoing studies. In addition, an internet search using Copernic Agent will be undertaken and will include relevant professional organisations, manufacturers and conference proceedings of the Association for Research in Vision and Ophthalmology (ARVO) (2005–7) and the American Academy of Ophthalmology (AAO) (2005–7). Reference lists of all included studies will also be scanned for additional reports. The reviews will be undertaken in accordance with the Centre for Reviews and Dissemination[31] and Cochrane Collaboration guidelines.[32]

3.2.2 Individual Patient Datasets (IPD)

Access to the following four datasets has been negotiated for this study. These include two randomised controlled trials (Moorfields Eye Hospital, London and Rotterdam Eye Hospital, the Netherlands) and two observational datasets – one hospital based (Dunfermline, Scotland) and one based on a community optometry led monitoring scheme (Nottingham, UK). The holders are applicants and collaborators on this project. A summary of the available data in these datasets is provided in *Table 1*. Specific details of each dataset are provided in the subsequent text.

Data from randomised controlled trials

Moorfields Eye Hospital (MEH) dataset The MEH dataset comprises data from a randomised controlled trial (RCT) of medical treatment with betaxolol eye drops versus placebo in ocular hypertension.[33] 356 ocular hypertensives were randomised during the period 1992–1996, only eye per patient was included. Eligible participants included those aged > 35 years with an IOP, by Goldmann Applanation Tonometry, between 22 mmHg and 35 mmHg and no significant ocular or neurological pathology that would affect visual field performance; visual acuity of 6/12 or better, and normal visual fields on Humphrey standard automated perimetry (SAP) (mean AGIS[34] score two baseline fields of 0). Participants were followed up four monthly, with Humphrey SAP, IOP and optic disc imaging, in the study eye for a minimum of 2 years, see *Table 1*. Treatment status was noted at each visit. Conversion to glaucoma was based on predefined criteria, if a patient developed a defect the visual fields test was repeated within one month, and repeated 3–4 months later. Conversion was confirmed if the defect was present on three consecutive reliable field tests. Any participants reaching the conversion end point had

TABLE 1 IPD for analysis (work package A and B)

Source	Patient group	Study design	Measures collected	Length of follow-up
Moorfields Eye Hospital, London	356 Ocular Hypertensives	RCT of treatment with Betaxolol 0.5% eye drops and placebo	IOP, Visual fields, Optic Nerve imaging [2 measures], Central corneal thickness	5 years [10 years on 200]
Rotterdam	409 Ocular Hypertensives	RCT of treatment with Betaxolol 0.25% eye drops, Timolol 0.5% and placebo	IOP, Visual fields [2 measures], Optic Nerve imaging [3 measures], Central cornea thickness	6–10 years
Dunfermline	350 Ocular Hypertensives	Registry with hospital based monitoring	IOP, Visual fields, Optic Nerve imaging	1–7 years
Nottingham	128 Ocular Hypertensives	Registry with community based monitoring	IOP, Visual fields Optic nerve imaging [3 measures], Central corneal thickness	1–5 years

IOP lowered by a topical ocular hypotensive, other than betaxolol. The treatment code remained unbroken until the end of the study in 1998 with 255 participants completing the study. The MEH data provides valuable RCT data, additionally, post trial surveillance with repeated measures of SAP and imaging with HRT every six months providing 10 year follow up data in 205 individuals.

Rotterdam The Rotterdam IPD comprises data from a placebo controlled, double-masked, prospective, 3 arms randomised controlled trial (placebo, Timolol 0.5% bid, betaxolol 0.25% bid) which was completed in August 2008. This ongoing study randomised 409 participants, including both eyes, with raised IOP, ≥ 22 mmHg and ≤ 32 mmHg by Goldmann Applanation Tonometry, and normal visual fields on Humphrey automated perimetry between November 1997 and March 2001. Participants were deemed eligible if of white ethnic origin, and had a best corrected Snellen visual acuity of at least 20/40. Participants with any significant coexisting ocular or systemic disease or any use of ocular hypotensives in the preceding three months were excluded.[35] The follow-up entails 6-monthly visual field analysis and imaging of the optic disc using scanning laser polarimetry (GDx at baseline, but follow up using a later model GDx VCC) in all participants, and HRT images in about 100 participants and stereo-optic disc photographs in 80 participants. To date, 360 participants remain in the study, 24 have converted to OAG.

Data from observational cohorts

Nottingham The Nottingham data set involves a cohort of 128 ocular hypertensives followed in a community scheme run by locally trained optometrists. Some are on treatment but most are not. The scheme started in 2003. Patients are reviewed once a year with visual fields, Goldman applanation tonometry and clinical disc examination. Treatment state is recorded at each visit. In addition all patients had stereo-photography taken at baseline for reference and a HRT performed at baseline and 5 years. Measurements on both eyes are available.

Dunfermline The Dunfermline data set collated on an electronic patient record of all referrals with confirmed ocular hypertension from 2000, with data on both eyes of an estimated 350 people with ocular hypertension accrued. Low risk cases are discharged after two visits (at one year), those who develop glaucoma remain under review, and those deemed as high risk of developing glaucoma remain under review. Humphrey SAP data are available on both eyes of all patients (except those unable to complete the test).

3.2.3 Work package A: Risk prediction (objectives 1 and 2) – Lead – Deeks, Birmingham

This will include several components (i) systematic identification and critical appraisal of risk prediction tools for the progression of ocular hypertension to OAG (ii) validate models identified in the systematic review that predict progression to glaucoma using individual patient data (IPD) from identified cohorts (RCTs and observational) of people with raised IOP, assessing comparative predictive ability and transportability; (iii) Compare agreement with the published predictive models:

Component (i) Systematic Review to identify the risk prediction tools (Led by Burr, Aberdeen):

- Two reviewers will view the titles and abstracts of all the reports, identified by the electronic searches, and screen independently. A full copy of all possible or definitely relevant studies will be obtained and inspected by two reviewers, independently, to assess whether they meet the inclusion criteria. Where a difference in opinion exists, the two reviewers will consult a third party. Initial screening of the titles and abstracts will be undertaken in Aberdeen and in Birmingham independently to identify articles for retrieval.

■ *Critical appraisal (Led by Deeks, Birmingham):* The identified prediction tools will be assessed according to the data and methods used in their development, and the quality of validation studies. Key components which will be assessed are the completeness of recording baseline characteristics and outcomes, the use of prospective (rather than retrospective) methodology, and avoidance of the treatment paradox. In development studies the modelling used to create a predictive score will be evaluated for statistical soundness and validity, particularly for the determination of cut-points. In validation studies we will consider whether the score has been applied exactly as it was derived, and review the measures used to report predictive ability.

Component (ii) Validation in the two observational cohorts (Nottingham and Dunfermline) and the trial datasets (Rotterdam & Moorfields)

We will independently validate the identified model(s) in several datasets and explore if additional risk factors can be identified. Predictive equations will be applied to the available data for each individual in each dataset. Three challenges which will need to be addressed in this process will be (a) the non-availability of key predictive variables in some of the datasets and (b) in subsets of individuals within a dataset, and (c) the use of treatment to affect prognosis. For (b) we will consider the role of multiple imputation techniques where data are missing, whereas for (a) variables will be omitted from the predictive scores, for (c) we will adjust for the expected effect of interventions estimated from systematic reviews.

Component (iii) Comparison of predictive ability

Models will be compared both in terms of their calibration (using calibration plots) and discriminating ability (using ROC analysis). The results of the validation of the identified predictive models will be considered across the datasets obtained to judge transportability of models across settings.

3.2.4 Work package B: Monitoring criteria (objective 3) Lead – Perera (Oxford), in collaboration with Crabb/Garway Heath (City/MEH University)

Component (i) Exploratory analyses of the two randomised controlled trials

Data from the randomised controlled trials at Moorfields and Rotterdam will provide information to determine the optimal monitoring criteria to both ascertain ocular hypertension and for identifying clinically important rates of change. The basis of these rely on estimating the within-person variability, between-person variability, and the natural trend of IOP as well as variability of relevant clinical measures of glaucoma (visual field, optic disc and RNFL) within specified periods.

Exploratory analyses of the two trials will be carried out to determine the feasibility of combining their data. In case combining is not appropriate (e.g. significant differences in average long term effect, or period variation), the analysis will be limited to the dataset providing the most information. We aim to carry out two independent analyses, one focusing in individuals under no ocular hypotensive drugs (e.g. placebo) and a second on individuals under stable medication.

As the data arise from placebo controlled RCTs, the apparent change in IOP and visual field and RNFL measurements over time (after the initial response to therapy) can be thought of as made up of 3 components:

1. The *average true long-term change* in IOP, visual field and RNFL measures of the whole group
2. A *short-term variability*: that is a combination of analytic variability and a week-to-week biological fluctuation around a stable average
3. The *long-term variability*: A *variation in long-term true change* between individuals (as would be seen with the theoretical average of a large number of measurements per individual).

These elements will be estimated using two different methods (a) a direct method and (b) a linear mixed-effects model. For a discussion of the methods used we focus only on IOP; however the same methods will be employed to estimate the different components of variability for visual field and RNFL parameters.

Direct Method of Estimation

The average change of the group can be estimated from the group average at each time point. Therefore, a direct method for estimating long term variability uses the variance of the differences between the baseline value and each subsequent time point, taking IOP as an example, this can be calculated as:

$$(\text{IOP at time } i - \text{IOP at baseline})^2$$

The times i are equivalent to the different follow up times where data were recorded – 4 months for the Moorfields RCT and 6 months for the Rotterdam RCT – with a maximum of 10 years follow-up for either trial. These will be calculated only after the stable baseline which we have taken as the first measurement in each trial (4 or 6 months) after treatment for the medicated and placebo group to reduce the noise created by the initial response to therapy. The long term variability can be estimated by subtracting the short-term variability from the variability of the change.

Two methods will be used to estimate short-term variation in IOP. First, the run-in period IOP used to determine eligibility to the trials will provide an estimate of short-term variability. Second, we will use linear extrapolation backward from the longer-term measures, to estimate the apparent variance at time 0. This second method will be the only one used to estimate short term variability for visual field and RNFL measures.

Modeled Method of Estimation

The above components will be estimated using a mixed longitudinal model. The main assumption of this model is that the IOP for each individual has a linear increase over time but that the rate of increase varies between individual (See Appendix for details).

Bland–Altman plots will be used to test if between- and within-person variability levels are dependent on different starting IOP levels within the 22 mmHg and 35 mmHg range.

Modelling of true and false positive rates

Estimates obtained using the above methods will allow modelling of the number whose 'true' IOP level exceed pre-defined threshold(s). The average change in IOP per year and the estimated within-person variability at different times (e.g. 1, 2, and 5 years) will allow us to estimate the proportion whose true value would, and would not, have drifted beyond a specific level.

In the economic model (work package C component ii) these data will be used to inform the estimates of progression of IOP to visual field loss over time and also provide information on our

ability to correctly identify that such changes have actually occurred. Within the economic model the implication of progression of IOP to visual field loss and consequent effects on costs and quality of life will be modeled.

Component (ii) Systematic review to determine the reliability of alternative IOP measurement devices (tonometers)

Objective: To compare the reliability between alternative tonometers.

Methods:

■ Health technologies to be assessed: Contact, non-contact tonometry and self-measurement tonometers, including Goldmann Applanation Tonometry, Tonopen, Ocular Response Analyser, Dynamic Contour Tonometry, Rebound Tonometry, Transpalpebral Tonometry, Perkins Tonometry and Pressure Phosphene Tonometry.
■ Target Population: Adults aged 18 years.
■ Inclusion criteria: We will include studies which assess the reliability of for two or more of the alternative tonometers compared to the Goldmann applanation tonometer. Directly comparative studies reported since 1987 of people suspected of having raised IOP will be eligible for inclusion.

Two reviewers will view the titles and abstracts of all the reports, identified by the electronic searches, and screen independently. A full copy of all possible or definitely relevant studies will be obtained and inspected by two reviewers, independently, to assess whether they meet the inclusion criteria. Where a difference in opinion exists, the two reviewers will consult a third party. We will undertake dual data extraction onto data extraction forms developed for the reviews. Discrepancies will be resolved by discussion or by a third party.

We will quality assess included studies using a modified checklist based upon the one produced by Craig *et al.* and Whiting *et al.*[36,37] We will extract and summarise any reliability measures reported (including but not limited to Kappa, intracorrelation coefficients and Bland–Altman limits of agreement approach). Where appropriate, estimates from comparable studies will be formally combined.

This review will inform the economic modelling with regards to the identification of ocular hypertension and tonometry devices for surveillance pathways.

3.2.5 Work package C: Economic modelling evaluation of alternative surveillance regimes (objectives 4–7)

Component (i) Determining the care pathways for the economic model

We plan to elicit the preference of the public using a discrete choice experiment (DCE), an approach we have considerable experience with and have successfully used to elicit preferences of patients with glaucoma for the outcomes of treatment.[38] The DCE will assess the individuals' preferences for a monitoring service for the detection of glaucoma within individuals with OHT. In particular we will assess:

■ willingness to pay for specific attributes of monitoring for detecting early, treatable glaucoma in those individuals at risk due to raised IOP and other characteristics (e.g. age, gender and general health)
■ willingness to pay for different configurations of monitoring services
■ probability of take-up of alternative monitoring regimes

■　Feed the estimates from above into an economic evaluation model that forms part of the wider project

DCEs are an attribute measure of benefit. The technique has the basic premise that any goods or service can be described by its characteristics (attributes) and second that the extent to which an individual values a good or service and therefore takes it up depends upon the levels of these attributes.[39,40] The technique involves presenting choices to individuals that vary with respect to the level of the attribute. In this application the DCE will be applied to assess probability of take-up and the value of IOP surveillance regimes and their components identified in **work packages A, B** and **component ii** of **work package B**.

Defining attributes and levels

To define attributes and levels for the DCE an Advisory Panel, a Focus Group and Questionnaire Piloting will be used.

Advisory Panel

An Advisory Panel will be convened for the main project to develop the care pathways for the economic model with participants not directly involved in the project. This will include two potential users of monitoring strategies, one optometrist, one non-clinical health service manager, one specialist nurse, one ophthalmologist, and one community optometrist. The Advisory Panel will develop the potential 'attributes' (factors influencing the take-up of alternative monitoring regimes) and plausible 'levels' for the DCE based on information available generated during the development of the application for funding for this work, the applicants and collaborators meeting and their own expertise.

Patient Focus Group

A Patient Focus Group will also be convened to further develop the attributes and levels. Potential participants will be identified through attendance to the Grampian Glaucoma Referral and Monitoring Scheme Service led by Mr Azuara Blanco, Consultant Ophthalmologist, Aberdeen Royal Infirmary which includes patients with ocular hypertension. This service started in June 2004 and has been developed to improve the diagnostic accuracy of glaucoma to reduce unnecessary referrals to the hospital glaucoma clinic and to develop a community-monitoring scheme for people at risk of developing glaucoma but who do not require treatment.

Initially, around 15 people will be approached; we aim to have a group of 6 to 8 people. Individuals will receive an invitation letter by post inviting them to take part in the focus group, together with an information leaflet describing the study and what is involved. If after reading the information sheet, individuals wish to participate, they will be asked to return the form attached to the invitation letter provided in a reply paid envelop or telephone the study co-ordinator (contact details will be provided in the invitation letter). After this, participants will be contacted by the interviewer to arrange the focus group meeting (date and time) and a confirmation letter will be sent. Participants will be encouraged to make contact with the research team for clarification of queries or further information. Participants will express their views on how a monitoring regime could be organised, including the setting where it might take place, testing frequency, monetary value and time spent attending for monitoring. We will use the framework approach to analyse the data obtained.

Determining choice sets

Following defining of the attributes and levels, experimental design techniques will be used to reduce the number of possible choice sets to a manageable size whilst still being able to estimate utility scores for any possible monitoring scenario. Experimental design techniques maximise some statistical measure of efficiency, usually D-efficiency. This ensures precision around the

estimated parameters.[40] The choice sets will ask individuals to choose between alternative profiles that vary with respect to the levels of the given attribute. These choices will reflect the criteria of minimal overlap, level balance and orthogonality. D-efficiency scores will be used to ensure an optimal set of choices are presented.[41] *Table 2* gives a hypothetical example of a DCE question.

Questionnaire Pre-piloting

Pre-piloting of the questionnaire will be conducted among members of the Health Services Research Unit. The purpose of this pre-piloting will be to test its rationality and validity.

A price proxy will be included so that willingness to pay, a monetary measure of benefit, can be indirectly estimated.[42] It is believed that a reasonable number of attributes to be included within a DCE is around 6 as more than this might lead to an unmanageable number of options by the respondent.[43] The more attributes the less likely the individual to adopt compensatory decision making, which is crucial when estimating marginal rates of substitution. Examples of potential attributes for the present experiment are: the setting where the monitoring might take place (e.g. hospital, community); testing frequency (e.g. 6 months, 12 months, 24 months); monetary value (e.g. £20, £40); time spent each year for monitoring sessions (e.g. 2 hours, 5 hours). Other attribute/s might arise from test characteristics (e.g. discomfort, anxiety).

Questionnaire piloting

Piloting of the questionnaire will be conducted in a sub-sample of the target populations. Pilot sample size will be big enough to conduct preliminary regression analysis (e.g. around 30 responses per subgroup) for the general population survey. This will be provided by the private company (see below, section on surveys) as part of the commercial agreement. Piloting will allow the detection of potential inconsistencies (e.g. positive relations where negative ones are anticipated). Furthermore, we will pilot the patient survey in 5 to 10 participants from two observational cohorts of patients in Nottingham and Dunfermline (see details of dataset below) to check participants responses (e.g. how participants responded to rationality tests).

The Surveys

We will conduct a survey among members of the general population – to elicit society's preferences.

The questionnaires will include a set of DCE choices (around 16) with alternative monitoring schedules that vary with respect to the levels of the given attribute, questions on patients' basic demographics and general health (EQ-5D).

Survey of the general population – a population preferences study

Adults over 18 years old living in the UK will be sampled from the general population. The survey will be conducted through an online research company: Research Now. This company has been conducting online research since 2001 and they provide a full service of field work management, data verification, quality control and conduct of the survey. The company was

TABLE 2 Example of a DCE question: which monitoring service would you choose?

	Monitoring service A	Monitoring service B	No monitoring C
Frequency of testing	Yearly	Every three months	–
Time costs per year	2 hours	5 hours	–
Money costs per year	£20	£40	–
Place of testing	Hospital eye clinic	Local optician	–
Which option would you choose?	Service A ❑	Service B ❑	No monitoring C ❑

chosen on the basis of a tendering process and is a company members of the research team have used previously in a National Preventative Research Initiative funded project that used a Discrete Choice Experiment to investigate lifestyle interventions to prevent obesity.

Sample and setting

Sampling will be done from Research Now's actively-managed online access panel. These panellists are recruited by email and online marketing from over 300 diverse online affiliate partners and targeted website advertising. Panel members actively joined the Valued Opinions panels by completing a registration survey (www.valuedopinions.co.uk). They are judged to have 'opted in' by accepting the terms and conditions and by taking part in surveys with the option to unsubscribe at any time. Panel registrations are automatically checked at the time of registration for duplicate e-mail addresses with further checks (e.g. at the time of incentive fulfilment, address files are checked for duplicate addresses) for fraudulent panellists. Basic demographic information (e.g. age, gender, region, household demographics) is collected when signing up to the panel and updated at the end of every survey. An active panel member is defined as a member that has participated in at least one survey, or updated his/her profile data or registered to join the panel within the last 12 months. At April 2009, panel membership across Valued Opinions family country panels around the world was 2,560,262 (these are individual members and not household). Panel management is compliant with all relevant market research standards, local data protection and privacy laws.

For this survey, sample selection will be based on the sample size needs (see section below on sample size). Once the sample has been selected, an e-mail invitation will be automatically generated to potential respondents. The invitation will contain key information for panellists to understand the commitment for the survey e.g. survey topic, survey length, incentive, and the length of time the survey will be open for. Panellists will access the survey through a unique link stated in the invitation e-mail. The content of the survey will be similar to the survey of patient based preferences study concerned with lifestyle interventions to prevent obesity, previously approved by the North of Scotland Research Ethics Committee 2 and will be fully piloted prior to field work. The invitation will also contain a link to the company's privacy policy, an opportunity to unsubscribe from the panel and a link to a member of Research Now's staff for any queries.

Panellists are rewarded for the time they take to complete the survey through a structured incentive scheme. They receive a cash reward for participating in individual surveys – the amount is clearly stated in the invitation e-mail and related to the survey length, interest and complexity (range between 50p–£5). Once a panellist reward balance reaches £10, they can redeem a voucher which is valid at national retailers.

Each panellist will be assigned an individual ID, allowing the company to monitor panellist activity and distinguish between contact rate (e.g. those who were initially contacted and did/did not complete the survey) and completion rate (e.g. those who completed the survey and did not drop out). The company follows UK data protection procedures. They have secure servers for collecting survey data. Sampling is carried out regularly with highly encrypted links to the database servers. All survey data will be anonymous and linked to the panel database by a unique ID so the panellists' identity is always protected.

Sample sizes

The following considerations govern the necessary sample size:

1. Regression analysis needs a sample size sufficiently larger than the number of independent variables (e.g. DCE number of attributes and levels);[9]
2. For each predetermined subgroup of the main sample, a sample size of 30–100 is sufficient;[10]

A number of factors have been identified which may influence preferences for monitoring for those with ocular hypertension. These are socio-economic status (high, medium and low); age (18–40, 41–60, >60); and general health (3 categories according to EQ-5D scores collected simultaneously). The private company would sample individuals until the targeted sample size is obtained. A purposive quota sampling approach will be employed to identify sufficient participants in each of the defined subgroups. We believe that a sample size of 800 will be enough to conduct the analysis for the general population survey.

Analysis

Appropriate logistic regression techniques will be used to analyse the response data.[40] The decision on which model to use is an empirical one and will depend, to certain extend, on the data. Following Amaya-Amaya and colleagues[43] we will start with a standard multinomial logit model, then test that restricted assumptions of this model are met, to finally decide on the proper model according to this. For instance, if unobserved characteristics are present and the data likely to be clustered then a nested logit specification might be more appropriate. Moreover, if for particular attributes there is suspicion that tastes might vary considerably across members of the population, then a random parameters model might be adequate. We anticipate that a few model options will be tested and a final model chosen according to data characteristics. The theoretical validity (e.g. the extent to which the coefficients move in the direction one would expect) will be evaluated from the regression results.

The analysis described above will allow consideration of the relative importance of the individual characteristics of the service; as well as estimation of the probability of take-up of alternative monitoring regimes. The inclusion of a price proxy will allow estimation of the monetary value of any given service. This willingness to pay will be used within the economic evaluation modelling (**section 3.2.5, component ii**). In effect, monetary values will be attached to attributes and levels of surveillance processes and individual health outcomes. Furthermore, the model outputs will be expressed in terms of these processes and health outcomes and therefore net benefits (in monetary values) will be estimated for each strategy considered. The analysis will build upon the work conducted for a previous Health Technology Assessment by our group[46] but we believe this approach has not been applied to the economic evaluation of surveillance regimes in ophthalmology. The DCE will also be developed iteratively with the economic evaluation model and we believe that this also has not been done before. More explicitly, the monetary values estimated for alternative surveillance methods from the DCE will be incorporated into an economic model, with the imprecision of estimates used explored using probabilistic sensitivity analysis.

Component (ii) Economic modelling using the results of the systematic reviews and primary datasets to determine the effectiveness, cost-effectiveness, cost–utility and cost–benefit of differing surveillance regimes for people with ocular hypertension

We will use an individual sampling model (e.g. a discrete event simulation) approach as the basis of the economic evaluation of the alternative surveillance regimes considered. We believe this will give a more refined approach than that of a Markov model when addressing questions such as the optimal length of time between surveillance contacts as the length of time between surveillance contacts becomes a model parameter rather than stipulated by the model structure. Moreover, the capacity of health care services to supply the some or all of the proposed strategies may be constrained and this approach gives high flexibility to model these issues. The model structure will be based on the findings of the other parts of this project including a review of previous economic evaluations of surveillance regimes. For example from **component (ii)** of **work**

package B we will obtain information on the type of test(s) used for surveillance and their signal to noise ratio). Similarly, from the risk prediction algorithm (**work package A**) we will obtain the risk of developing glaucoma; from the DCE we will obtain estimates of the public preferences for different surveillance strategies which vary strategy according to their characteristics defined in terms of the attributes and levels of the DCE. The results of the DCE along with discussion with the advisory panel (experts and consumer representatives) will be used to identify potentially feasibility surveillance strategies to be compared within the model.

The model will sample adults who are each defined in terms of certain characteristics (based upon the factors considered by the predictive models and the sub-groups considered by the DCE). According to these characteristics, each individual will have a risk of developing glaucoma (and also to subsequently develop visual impairment). These risks will be informed by the analyses conducted in the other elements of the research and additional focused searches to identify other relevant parameter values to populate other aspects of the model. The model will itself will describe the pathways of care that individuals may follow from following initial identification as 'at risk' to alternative testing strategies, compared with current practice alternatives. The model will be populated with cost data derived from consultation with NHS providers and the searches (**section 3.2.1**). Other sources of cost data would be the NHS Reference costs for health care interventions or unit costs for health care service staff (e.g. PSSRU). We will use a NHS and societal perspective for the economic analysis. In addition to NHS costs we will attempt to estimate the travel costs and costs due to absenteeism caused by the interventions and/or disease as proxy for indirect costs. Such costs will be presented separately from NHS costs. Benefits will be measured in three ways (i) a measure of clinical effectiveness (e.g. number of glaucoma cases avoided) (ii) QALYs derived from utility values derived from the EQ-5D responses included as part of the DCE questionnaire, a focused search of the literature and our existing quality of life research briefly described in **Section 3.2.5(i); (ii)** willingness to pay derived from the incorporation of the DCE results into the economic model. The economic model will thus be presented as (i) a cost-effectiveness analysis (CEA) (ii) a cost–utility analysis (CUA) and (iii) a cost–benefits analysis (CBA). Different limitations apply to each of these proposed economic evaluation approaches; for example, the CEA may be intuitively easier for a clinical audience but the narrow measure of benefit used may not capture preferences for health and process outcomes. In the CUA, QALYs will be used which will capture preferences for changes in health related outcomes and in the CBA willingness to pay estimates derived from the DCE will be used to capture individuals preferences for both process and health outcomes. The results from these three analyses will be compared to see whether consistent conclusions would be drawn. If conclusions are not consistent then the reasons for this inconsistency will be sought as this is potentially important information to present to decision makers (for example the CUA and CBA may imply different conclusions because of the preferences individual place on different way the surveillance service might organised which are captured in the willingness to pay values but not in the QALYs).

Costs and benefits will be discounted at 3.5%.[47] The analysis will also include probabilistic and deterministic sensitivity analyses will be conducted to explore the importance of key areas of uncertainty. These might include, for instance, the use of alternative utility values and/or unit costs. Structural uncertainty will also be explored if required.[48]

Component (iii) Identification of future research needs

A value of information analysis[49] will be conducted to identify the expected value of perfect information over the expected lifetime of the surveillance strategies considered and the value of further research to identify more precise and reliable estimates of parameters of the economic model.

4   ETHICAL ARRANGEMENTS

Part of our proposed study involves analysing individual patient data from datasets collected for other purposes as described in **section 3.2.2**. This data will be anonymised and we will follow the guidance laid out in the MRC Ethics Series Personal Information in Medical Research and Section 60 of the Health and Social Care Act regarding use, storage and investigator responsibilities. We are also proposing to survey a sample of the UK population using a research marketing company (**section 3.2.5i**). The responses to this survey will be anonymous i.e. we will not be able to link responses to names and addresses. In addition, we will ask for the informed consent from the advisory panel members (**section 3.2.5i**) to confidentially use their panel discussions to inform the other study components and the results of the study. We will also seek NHS Research Ethics Committee advice and approval to conduct the patient survey and we will abide by the 1998 Data Protection Act.

5   EXPERTISE AND ADVISORY PANEL

The project will be led from the Health Services Research Unit (HSRU), the Health Economics Research Unit (HERU), University of Aberdeen and the Eye department NHS Grampian. The units have extensive experience in systematic reviews, statistical and economic modelling, discrete choice experiments and a track-record of research in the field of glaucoma (MR, LV, JC, RH, JB, and AAB). Additionally, HSRU has previous experience of conducting IPD meta-analyses including co-ordinating a large international collaboration involving 28 datasets (which included extensive economic modelling).[50]

Burr as Principal Investigator will be project manager ensuring that the three work packages are complementary and working to complementary time scales. This will involve weekly project management meetings in Aberdeen (Burr, Vale, Cook, Hernandez, Azuara-Blanco and the project research fellow) and communication with each of the work package leads by email and telephone.

The clinical research unit at MEH and the Vision Science Unit of City University (TGH, RW, DC, and AK) provide clinical and research expertise in epidemiology, Cochrane reviews, and on the optimal frequency and spacing of testing to detect change in people with ocular hypertension and glaucoma. Expertise in the statistical methods for monitoring will be provided by the Monitoring in Chronic Diseases group in the Department of Primary Health Care in the University of Oxford (PG, RP). JD is Professor of Health Statistics at the University of Birmingham and provides expertise in systematic reviews and predictive modelling. Other collaborators have agreed to provide data for the individual patient data analysis; Roshini Sanders, a Consultant Ophthalmologist leading the glaucoma service at, Fife Acute Hospitals Trust, Scotland; Hans G Lemij, MD, PhD, a glaucoma specialist at the Rotterdam Eye Hospital in Rotterdam, The Netherlands, with a primary interest in research and education in glaucoma imaging, structure–function relations in glaucoma and in glaucoma shared care. The co-applicants from Nottingham have particular expertise in developing co-management care pathways for glaucoma monitoring involving optometrists both in a primary care, and hospital setting.

The applicants and collaborators will meet at the start of the project and again in month 12, when the **work packages A** and **B** are complete to assist in the interpretation of the evidence from the systematic reviews, DCE and IPD analysis. The group will meet for a third time when the economic modelling is completed.

An advisory panel will be convened at the start of the project. Participants will be invited from local optometry practices, and the eye department at Aberdeen Royal Infirmary aiming for a panel of seven comprising: two potential users of a surveillance strategy, one optometrist, one health service manager, a specialist nurse, an ophthalmologist and a community optometrist. The panel will meet in month 2 to develop the pathways for the modelling, and determine the ranges of levels to be included in the DCE. The advisory panel will meet again in month 12 and will comment on and agree the proposed care pathways based on the outputs of the three work packages. The refined pathways will determine the structure of the final economic model.

6 SERVICE USERS

The novel part of this application is the discrete choice experiment to determine preferences for the process of any potential surveillance service for those with ocular hypertension. In addition we will identify two potential users of a service from patients attending a monitoring scheme in Grampian for those with ocular hypertension.

7 REFERENCES

1. Bamashmus MA *et al.* Causes of blindness and visual impairment in the West of Scotland. *Eye* 2004;**18**:257–61.

2. Bunce C *et al.* Leading causes of certification for blindness and partial sight in England & Wales. *BMC Public Health* 2006;**6**:8 Mar 2006.

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

4. Burr JM *et al.* The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technol Assess* 2007;**11**:1–190.

5. Bonomi L *et al.* Prevalence of glaucoma and intraocular pressure distribution in a defined population: The Egna-Neumarkt study. *Ophthalmology* 1998;**105**:209–15.

6. Hollows FC *et al.* Intra-ocular pressure glaucoma and glaucoma suspects in a defined population. *Br J Ophthalmol* 1966;**50**:570–86.

7. Leibowitz HM *et al.* The Framingham Eye Study monograph. *Surv Ophthalmol* 1980;**24**:335–610.

8. Sommer A *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.

9. Wensor MD *et al.* The prevalence of glaucoma in the Melbourne Visual Impairment Project. *Ophthalmology* 1998;**105**:733–9.

10. Heijl A *et al.* Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol* 2002;**120**:1268–79.

11. Kass MA *et al.* Intraocular pressure and visual field progression in open-angle glaucoma. *Am J Ophthalmol* 2000;**130**:490–1.

12. Leske MC *et al.* Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999;**106**:2144–53.

13. Grodum K *et al.* A comparison of glaucoma patients identified through mass screening and in routine clinical practice. *Acta Ophthalmol Scand* 2002;**80**:627–31.

14. Leske MC. The epidemiology of open-angle glaucoma: a review. *Am J Epidemiol* 1983;**118**:166–91.

15. Varma R *et al.* Prevalence of open-angle glaucoma and ocular hypertension in Latinos: The Los Angeles Latino Eye Study. *Ophthalmology* 2004;**111**:1439–48.

16. Kass MA *et al.* The Ocular Hypertension Treatment Study: A randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:701–13.

17. Miglior S *et al.* Results of the European Glaucoma Prevention Study. *Ophthalmology* 2005;**112**:366–75 .

18. Vass C *et al.* Medical interventions for primary open angle glaucoma and ocular hypertension. *Cochrane Database Syst Rev* 2007, Issue 4. Art. No.: CD003167. DOI: 10.1002/14651858.CD003167.pub3.

19. Traverso CE *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.

20. Kymes SM *et al.* Management of Ocular Hypertension: A Cost-effectiveness Approach From the Ocular Hypertension Treatment Study. *Am J Ophthalmol* 2006;**141**:997–1008.

21. Gordon MO *et al.* The Ocular Hypertension Treatment Study: Baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;**120**:714–20.

22. Medeiros FA *et al.* Validation of a predictive model to estimate the risk of conversion from ocular hypertension to glaucoma. *Arch Ophthalmol* 2005;**123**:1351–60.

23. Medeiros FA *et al.* Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma. *Am J Ophthalmol* 2005;**139**:1010–8.

24. Validated Prediction Model for the Development of Primary Open-Angle Glaucoma in Individuals with Ocular Hypertension. *Ophthalmology* 2007;**114**:10–9.

25. Altman DG *et al.* What do we mean by validating a prognostic model? *Stat Med* 2000;**19**:453–73.

26. *Commissioning toolkit for community based eye care services* [report on the Internet]. UK Department of Health; 2007 [accessed May 2009]. Available from: URL:http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_063978.

27. *NICE Guideline Group Remit: Glaucoma* [document on the Internet]. National Institute for Health and Clinical Excellence; 2009 [accessed Jan 2008] Available from : URL: http://www.nice.org.uk/nicemedia/pdf/CG85NICEGuideline.

28. *Heideleberg Retina Tomograph (HRT)* [webpage on the Internet]. Heidelberg Engineering [accessed January 2008]. Available from: URL:http://www.heidelbergengineering.com/.

29. *Clinical solutions: GDx* [webpage on the Internet]. Carl Zeiss Meditec [accessed January 2008] Available from: URL: http://www.meditec.zeiss.com/gdxvcc.

30. Glasziou PA *et al.* Evidence-based medical monitoring: from principles to practice. Oxford: Blackwells; 2008.

31. *Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews*. CRD Report No 4. York: NHS Centre for Reviews and Dissemination; 2001.

32. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6* [document on the Internet]. The Cochrane Library; 2006 [accessed Jan 2008]. Available from: URL: http://www3. interscience.wiley.com/homepages/106568753/handbook.pdf.

33. Kamal D *et al.* Results of the betaxolol versus placebo treatment trial in ocular hypertension. *Graefe's Arch Clin Exp Ophthalmol* 2003;**241**:196–203.

34. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology* 1994;**101**:1445–55.

35. Reus NJ *et al.* The prevalence of glaucomatous defects with short-wavelength automated perimetry in patients with elevated intraocular pressures. *J Glaucoma* 2005;**14**:26–9.

36. Craig JV, Lancaster GA, Williamson PR, Smyth RL. Temperature measured at the axilla compared with rectum in children and young people: systematic review. *BMJ* 2000 April 29;320(7243):1174–1178.

37. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003 Nov 10;**3**:25.

38. Burr JM *et al.* Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. *Optomet Vis Sci* 2007;**84**:797–808.

39. Ryan M *et al.* Methodological issues in the monetary valuation of benefits in healthcare. *Expert Rev Pharmacoecon Outcomes Res* 2003;**3**:717–27.

40. Louviere J *et al. Stated choice methods: analysis and application*. Cambridge: Cambridge University Press; 2000.

41. Burgess L *et al.* Optimal designs for 2k choice experiments. *Comm Stat Theory Method* 2003;**32**:2185–206.

42. Ryan M *et al.* Using discrete choice experiments in health economics. In: Jones A, editor. *The Elgar companion to health economics*. Bodmin: MPG Books; 2006.

43. Amaya-Amaya M, Gerard K and Ryan M. Discrete Choice Experiments in a nutshell. In *Using Discrete Choice Experiments to Value Health and Health Care*. M. Ryan, K. Gerard and M Amaya-Amaya (eds.), Springer, 2008. Page 32.

44. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal*. June, 2008 Website: www.nice.org.uk/media/B52/A7/ TAMethodsGuideUpdatedJune2008.pdf Accessed: May 2009.

45. Porteous T *et al.* Preferences for self-care or professional advice for minor illness: a discrete choice experiment. *Br J Gen Pract* 2006;**56**:911–7.

46. Ryan M *et al.* Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy* 2003;**2**:55–64.

47. McCormack K *et al.* Laparoscopic surgery for inguinal hernia repair: Systematic review of effectiveness and economic evaluation. *Health Technol Assess* 2005;**9**:iii–80.

48. *Guide to the methods of technology appraisal* [document on the Internet]. London: National Institute for Clinical Excellence; 2004. [accessed Jan 2008] Available from: URL: http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf

49. Bojke L *et al.* Defining and characterising structural uncertainty in decision analytic models [document on the Internet]. Centre for Health Economics, University of York; 2006 [accessed Jan 2008]. Available from: URL: http://www.york.ac.uk/inst/che/pdf/rp9.pdf.

50. Briggs A, S *et al. Decision modelling for health economic evaluation.* Oxford: Oxford University Press; 2006.

51. McCormack K *et al.* Value of updating a systematic review in surgery using individual patient data. *Br J Surg* 2004;**91**:495–9.

Health Technology Assessment programme

Director,
Professor Tom Walley, CBE,
Director, NIHR HTA programme,
Professor of Clinical Pharmacology,
Department of Pharmacology and Therapeutics,
University of Liverpool

Deputy Director,
Professor Hywel Williams,
Professor of Dermato-Epidemiology,
Centre of Evidence-Based Dermatology,
University of Nottingham

Prioritisation Group

Members

Chair,
Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, Department of
Pharmacology and Therapeutics,
University of Liverpool

Professor Imti Choonara,
Professor in Child Health,
Academic Division of Child
Health, University of Nottingham
Chair – Pharmaceuticals Panel

Dr Bob Coates,
Consultant Advisor – Disease
Prevention Panel

Dr Andrew Cook,
Consultant Advisor – Intervention
Procedures Panel

Dr Peter Davidson,
Director of NETSCC, Health
Technology Assessment

Dr Nick Hicks,
Consultant Adviser – Diagnostic
Technologies and Screening Panel,
Consultant Advisor–Psychological
and Community Therapies Panel

Ms Susan Hird,
Consultant Advisor, External
Devices and Physical Therapies
Panel

Professor Sallie Lamb,
Director, Warwick Clinical Trials
Unit, Warwick Medical School,
University of Warwick
Chair – HTA Clinical Evaluation
and Trials Board

Professor Jonathan Michaels,
Professor of Vascular Surgery,
Sheffield Vascular Institute,
University of Sheffield
Chair – Interventional Procedures
Panel

Professor Ruairidh Milne,
Director – External Relations

Dr John Pounsford,
Consultant Physician, Directorate
of Medical Services, North Bristol
NHS Trust
Chair – External Devices and
Physical Therapies Panel

Dr Vaughan Thomas,
Consultant Advisor –
Pharmaceuticals Panel, Clinical
Lead – Clinical Evaluation Trials
Prioritisation Group

Professor Margaret Thorogood,
Professor of Epidemiology, Health
Sciences Research Institute,
University of Warwick
Chair – Disease Prevention Panel

Professor Lindsay Turnbull,
Professor of Radiology, Centre for
the MR Investigations, University
of Hull
Chair – Diagnostic Technologies
and Screening Panel

Professor Scott Weich,
Professor of Psychiatry, Health
Sciences Research Institute,
University of Warwick
Chair – Psychological and
Community Therapies Panel

Professor Hywel Williams,
Director of Nottingham Clinical
Trials Unit, Centre of Evidence-
Based Dermatology, University of
Nottingham
Chair – HTA Commissioning
Board
Deputy HTA Programme Director

HTA Commissioning Board

Chair,
Professor Hywel Williams,
Professor of Dermato-Epidemiology,
Centre of Evidence-Based Dermatology,
University of Nottingham

Deputy Chair,
Professor Jon Deeks,
Department of Public Health and
Epidemiology,
University of Birmingham

Programme Director,
Professor Tom Walley, CBE,
Professor of Clinical Pharmacology,
Department of Pharmacology and Therapeutics,
University of Liverpool

Members

Professor Judith Bliss,
Director of ICR-Clinical Trials
and Statistics Unit, The Institute of
Cancer Research

Professor David Fitzmaurice,
Professor of Primary Care
Research, Department of Primary
Care Clinical Sciences, University
of Birmingham

Professor John W Gregory,
Professor in Paediatric
Endocrinology, Department of
Child Health, Wales School of
Medicine, Cardiff University

Professor Steve Halligan,
Professor of Gastrointestinal
Radiology, Department of
Specialist Radiology, University
College Hospital, London

Professor Angela Harden,
Professor of Community and
Family Health, Institute for
Health and Human Development,
University of East London

Dr Martin J Landray,
Reader in Epidemiology, Honorary
Consultant Physician, Clinical
Trial Service Unit, University of
Oxford

Dr Joanne Lord,
Reader, Health Economics
Research Group, Brunel University

Professor Stephen Morris,
Professor of Health Economics,
University College London,
Research Department of
Epidemiology and Public Health,
University College London

Professor Dion Morton,
Professor of Surgery, Academic
Department of Surgery, University
of Birmingham

Professor Gail Mountain,
Professor of Health Services
Research, Rehabilitation and
Assistive Technologies Group,
University of Sheffield

Professor Irwin Nazareth,
Professor of Primary Care and
Head of Department, Department
of Primary Care and Population
Sciences, University College
London

Professor E Andrea Nelson,
Professor of Wound Healing and
Director of Research, School of
Healthcare, University of Leeds

Professor John David Norrie,
Director, Centre for Healthcare
Randomised Trials, Health
Services Research Unit, University
of Aberdeen

Dr Rafael Perera,
Lecturer in Medical Statisitics,
Department of Primary Health
Care, University of Oxford

Professor Barney Reeves,
Professorial Research Fellow
in Health Services Research,
Department of Clinical Science,
University of Bristol

Professor Peter Tyrer,
Professor of Community
Psychiatry, Centre for Mental
Health, Imperial College London

Diagnostic Technologies and Screening Panel

Members

Chair,
Professor Lindsay Wilson Turnbull,
Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary

Professor Judith E Adams,
Consultant Radiologist, Manchester Royal Infirmary, Central Manchester & Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester

Mr Angus S Arunkalaivanan,
Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham

Dr Diana Baralle,
Consultant and Senior Lecturer in Clinical Genetics, University of Southampton

Dr Stephanie Dancer,
Consultant Microbiologist, Hairmyres Hospital, East Kilbride

Dr Diane Eccles,
Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital

Dr Trevor Friedman,
Consultant Liason Psychiatrist, Brandon Unit, Leicester General Hospital

Dr Ron Gray,
Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford

Professor Paul D Griffiths,
Professor of Radiology, Academic Unit of Radiology, University of Sheffield

Mr Martin Hooper,
Public contributor

Professor Anthony Robert Kendrick,
Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton

Dr Nicola Lennard,
Senior Medical Officer, MHRA

Dr Anne Mackie,
Director of Programmes, UK National Screening Committee, London

Mr David Mathew,
Public contributor

Dr Michael Millar,
Consultant Senior Lecturer in Microbiology, Department of Pathology & Microbiology, Barts and The London NHS Trust, Royal London Hospital

Mrs Una Rennard,
Public contributor

Dr Stuart Smellie,
Consultant in Clinical Pathology, Bishop Auckland General Hospital

Ms Jane Smith,
Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds

Dr Allison Streetly,
Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine

Dr Matthew Thompson,
Senior Clinical Scientist and GP, Department of Primary Health Care, University of Oxford

Dr Alan J Williams,
Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital

Observers

Dr Tim Elliott,
Team Leader, Cancer Screening, Department of Health

Dr Joanna Jenkinson,
Board Secretary, Neurosciences and Mental Health Board (NMHB), Medical Research Council

Professor Julietta Patnick,
Director, NHS Cancer Screening Programme, Sheffield

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Disease Prevention Panel

Members

Chair,
Professor Margaret Thorogood,
Professor of Epidemiology, University of Warwick Medical School, Coventry

Dr Robert Cook,
Clinical Programmes Director, Bazian Ltd, London

Dr Colin Greaves,
Senior Research Fellow, Peninsula Medical School (Primary Care)

Mr Michael Head,
Public contributor

Professor Cathy Jackson,
Professor of Primary Care Medicine, Bute Medical School, University of St Andrews

Dr Russell Jago,
Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol

Dr Julie Mytton,
Consultant in Child Public Health, NHS Bristol

Professor Irwin Nazareth,
Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London

Dr Richard Richards,
Assistant Director of Public Health, Derbyshire County Primary Care Trust

Professor Ian Roberts,
Professor of Epidemiology and Public Health, London School of Hygiene & Tropical Medicine

Dr Kenneth Robertson,
Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow

Dr Catherine Swann,
Associate Director, Centre for Public Health Excellence, NICE

Mrs Jean Thurston,
Public contributor

Professor David Weller,
Head, School of Clinical Science and Community Health, University of Edinburgh

Observers

Ms Christine McGuire,
Research & Development, Department of Health

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

External Devices and Physical Therapies Panel

Members

Chair,
Dr John Pounsford,
Consultant Physician North Bristol NHS Trust

Deputy Chair,
Professor E Andrea Nelson,
Reader in Wound Healing and Director of Research, University of Leeds

Professor Bipin Bhakta,
Charterhouse Professor in Rehabilitation Medicine, University of Leeds

Mrs Penny Calder,
Public contributor

Dr Dawn Carnes,
Senior Research Fellow, Barts and the London School of Medicine and Dentistry

Dr Emma Clark,
Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol

Mrs Anthea De Barton-Watson,
Public contributor

Professor Nadine Foster,
Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University

Dr Shaheen Hamdy,
Clinical Senior Lecturer and Consultant Physician, University of Manchester

Professor Christine Norton,
Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust

Dr Lorraine Pinnington,
Associate Professor in Rehabilitation, University of Nottingham

Dr Kate Radford,
Senior Lecturer (Research), University of Central Lancashire

Mr Jim Reece,
Public contributor

Professor Maria Stokes,
Professor of Neuromusculoskeletal Rehabilitation, University of Southampton

Dr Pippa Tyrrell,
Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester

Dr Nefyn Williams,
Clinical Senior Lecturer, Cardiff University

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Interventional Procedures Panel

Members

Chair,
Professor Jonathan Michaels,
Professor of Vascular Surgery, University of Sheffield

Deputy Chair,
Mr Michael Thomas,
Consultant Colorectal Surgeon, Bristol Royal Infirmary

Mrs Isabel Boyer,
Public contributor

Mr Sankaran Chandra Sekharan,
Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust

Professor Nicholas Clarke,
Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust

Ms Leonie Cooke,
Public contributor

Mr Seumas Eckford,
Consultant in Obstetrics & Gynaecology, North Devon District Hospital

Professor Sam Eljamel,
Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee

Dr Adele Fielding,
Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School

Dr Matthew Hatton,
Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust

Dr John Holden,
General Practitioner, Garswood Surgery, Wigan

Dr Fiona Lecky,
Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust

Dr Nadim Malik,
Consultant Cardiologist/Honorary Lecturer, University of Manchester

Mr Hisham Mehanna,
Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust

Dr Jane Montgomery,
Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust

Professor Jon Moss,
Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust

Dr Simon Padley,
Consultant Radiologist, Chelsea & Westminster Hospital

Dr Ashish Paul,
Medical Director, Bedfordshire PCT

Dr Sarah Purdy,
Consultant Senior Lecturer, University of Bristol

Dr Matthew Wilson,
Consultant Anaesthetist, Sheffield Teaching Hospitals NHS Foundation Trust

Professor Yit Chiun Yang,
Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Pharmaceuticals Panel

Members

Chair,
Professor Imti Choonara,
Professor in Child Health,
University of Nottingham

Deputy Chair,
Dr Yoon K Loke,
Senior Lecturer in Clinical
Pharmacology, University of East
Anglia

Dr Martin Ashton-Key,
Medical Advisor, National
Commissioning Group, NHS
London

Dr Peter Elton,
Director of Public Health, Bury
Primary Care Trust

Dr Ben Goldacre,
Research Fellow, Epidemiology
London School of Hygiene and
Tropical Medicine

Dr James Gray,
Consultant Microbiologist,
Department of Microbiology,
Birmingham Children's Hospital
NHS Foundation Trust

Dr Jurjees Hasan,
Consultant in Medical Oncology,
The Christie, Manchester

Dr Carl Heneghan,
Deputy Director Centre for
Evidence-Based Medicine and
Clinical Lecturer, Department of
Primary Health Care, University
of Oxford

Dr Dyfrig Hughes,
Reader in Pharmacoeconomics
and Deputy Director, Centre for
Economics and Policy in Health,
IMSCaR, Bangor University

Dr Maria Kouimtzi,
Pharmacy and Informatics
Director, Global Clinical Solutions,
Wiley-Blackwell

Professor Femi Oyebode,
Consultant Psychiatrist and Head
of Department, University of
Birmingham

Dr Andrew Prentice,
Senior Lecturer and Consultant
Obstetrician and Gynaecologist,
The Rosie Hospital, University of
Cambridge

Ms Amanda Roberts,
Public contributor

Dr Gillian Shepherd,
Director, Health and Clinical
Excellence, Merck Serono Ltd

Mrs Katrina Simister,
Assistant Director New Medicines,
National Prescribing Centre,
Liverpool

Professor Donald Singer,
Professor of Clinical
Pharmacology and Therapeutics,
Clinical Sciences Research
Institute, CSB, University of
Warwick Medical School

Mr David Symes,
Public contributor

Dr Arnold Zermansky,
General Practitioner, Senior
Research Fellow, Pharmacy
Practice and Medicines
Management Group, Leeds
University

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Mr Simon Reeve,
Head of Clinical and Cost-
Effectiveness, Medicines,
Pharmacy and Industry Group,
Department of Health

Dr Heike Weber,
Programme Manager, Medical
Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

Psychological and Community Therapies Panel

Members

Chair,
Professor Scott Weich,
Professor of Psychiatry, University
of Warwick, Coventry

Deputy Chair,
Dr Howard Ring,
Consultant & University Lecturer
in Psychiatry, University of
Cambridge

Professor Jane Barlow,
Professor of Public Health in
the Early Years, Health Sciences
Research Institute, Warwick
Medical School

Dr Sabyasachi Bhaumik,
Consultant Psychiatrist,
Leicestershire Partnership NHS
Trust

Mrs Val Carlill,
Public contributor

Dr Steve Cunningham,
Consultant Respiratory
Paediatrician, Lothian Health
Board

Dr Anne Hesketh,
Senior Clinical Lecturer in Speech
and Language Therapy, University
of Manchester

Dr Peter Langdon,
Senior Clinical Lecturer, School
of Medicine, Health Policy and
Practice, University of East Anglia

Dr Yann Lefeuvre,
GP Partner, Burrage Road Surgery,
London

Dr Jeremy J Murphy,
Consultant Physician and
Cardiologist, County Durham and
Darlington Foundation Trust

Dr Richard Neal,
Clinical Senior Lecturer in General
Practice, Cardiff University

Mr John Needham,
Public contributor

Ms Mary Nettle,
Mental Health User Consultant

Professor John Potter,
Professor of Ageing and Stroke
Medicine, University of East
Anglia

Dr Greta Rait,
Senior Clinical Lecturer and
General Practitioner, University
College London

Dr Paul Ramchandani,
Senior Research Fellow/Cons.
Child Psychiatrist, University of
Oxford

Dr Karen Roberts,
Nurse/Consultant, Dunston Hill
Hospital, Tyne and Wear

Dr Karim Saad,
Consultant in Old Age Psychiatry,
Coventry and Warwickshire
Partnership Trust

Dr Lesley Stockton,
Lecturer, School of Health
Sciences, University of Liverpool

Dr Simon Wright,
GP Partner, Walkden Medical
Centre, Manchester

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.