

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 Deeks² and J Cook¹

¹Health Services Research Unit, University of Aberdeen, Aberdeen, UK

²Public Health, Epidemiology and Biostatistics Unit, University of
Birmingham, Birmingham, UK

³Health Economics Research Unit, University of Aberdeen, Aberdeen, UK

⁴Department of Primary Health Care, University of Oxford, Oxford, UK

⁵Glaucoma Research Unit, Moorfields Eye Hospital NHS Foundation Trust,
London, UK

⁶Department of Optometry and Visual Science, City University, London, UK

⁷Glaucoma Services, Rotterdam Eye Hospital, Rotterdam, the Netherlands

⁸Queen's Medical Centre, University Hospital, Nottingham, UK

⁹Queen Margaret Hospital, Dunfermline, UK

¹⁰Eye Centre, Tampere University Hospital, Tampere, Finland

*Corresponding author



Executive summary

Health Technology Assessment 2012; Vol. 16: No. 29

DOI: 10.3310/hta16290

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk



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*.

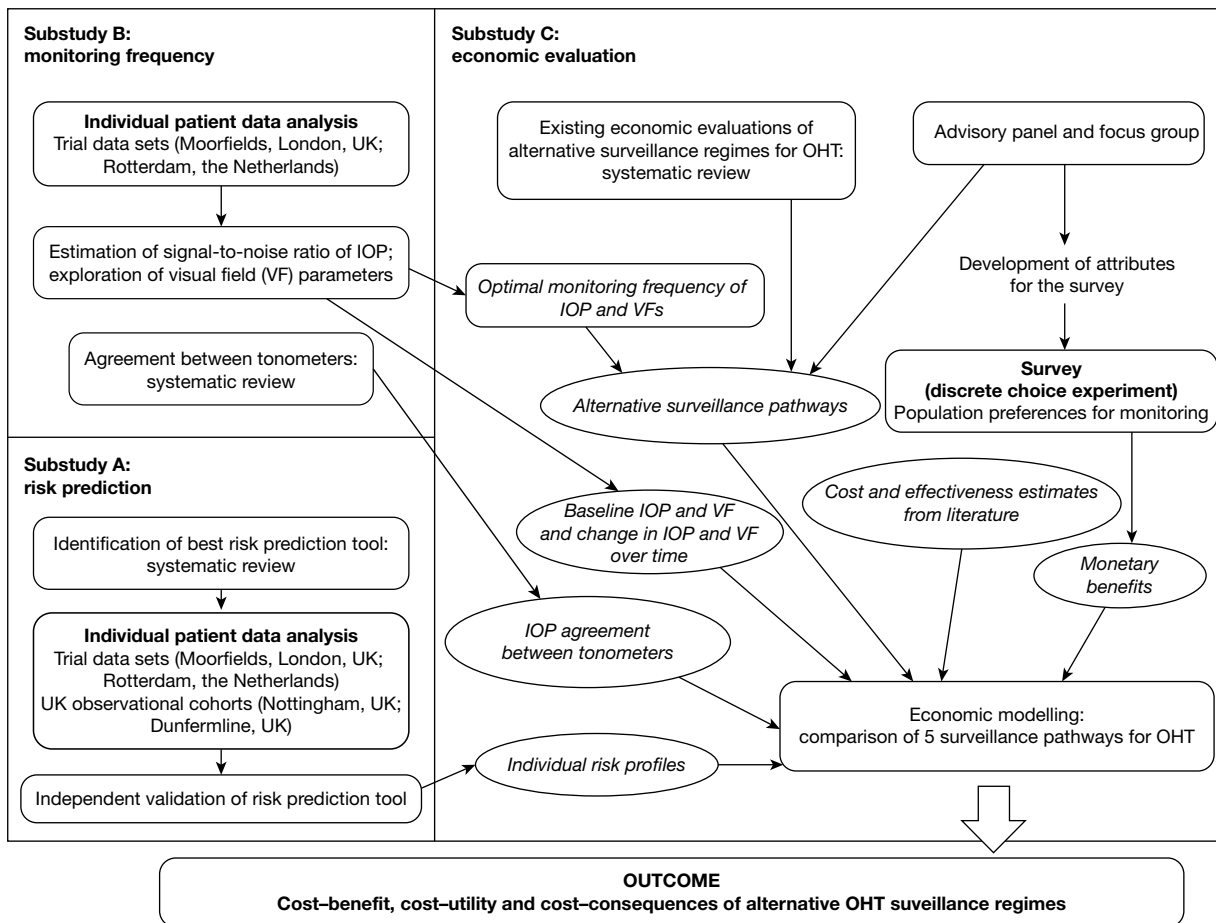


FIGURE 1 Overview of the project.

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 & Präzisionsbau, 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 $\geq 6\%$ 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 \leq £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

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Publication

Burr JM, Botello-Pinzon P, Takwoingi Y, Hernández R, Vazquez-Montes M, Elders A, *et al.* Surveillance for ocular hypertension: an evidence synthesis and economic evaluation. *Health Technol Assess* 2012;**16**(29).

NIHR Health Technology Assessment programme

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.

The research reported in this issue of the journal was commissioned by the HTA programme as project number 07/46/02. The contractual start date was in February 2009. The draft report began editorial review in June 2011 and was accepted for publication in November 2011. As the funder, by devising a commissioning brief, the HTA programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

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

Editor-in-Chief: Professor Tom Walley CBE
Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Peter Davidson, Dr Tom Marshall, Professor John Powell, Dr Rob Riemsma and Professor Ken Stein
Editorial Contact: edit@southampton.ac.uk

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

ISSN 2046-4932 (DVD)

© Queen's Printer and Controller of HMSO 2012. This work was produced by Burr *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>).

This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by Charlesworth Press.