The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation


CRD summary
This well conducted and clearly reported review assessed the clinical and cost effectiveness of screening for open angle glaucoma. Much of the data, and the focus of this abstract, related to the accuracy of screening tests. The authors concluded that some tests showed promise, but data were insufficient to draw firm conclusions. These conclusions are supported by the data presented.

Authors' objectives
To assess the clinical and cost effectiveness of screening for open angle glaucoma (OAG) in the UK by evaluating: diagnostic performance of screening tests; patient acceptance of testing; effectiveness of screening; effectiveness of treatment; epidemiology; risk factors and progression of glaucoma; and economic evaluation. This abstract considers the first three of these elements only (data on the effectiveness of treatment for glaucoma were derived solely from another systematic review and data on the effectiveness of screening were derived from a Cochrane review).

Searching
Twenty databases were searched from inception to November/December 2005 for relevant English-language articles. Separate searches were conducted for each component of the review question. Search strategies were reported in full. Reference lists of all included studies were scanned. Web pages of relevant professional organisations were searched.

Study selection
Performance of screening tests: Eligible studies were RCTs in which participants were randomised to receive either index test(s) or comparator test(s). All received the reference standard. Observational studies (cohort or case-control) in which all participants received the index test(s) and reference standard were also eligible for inclusion. Studies were required to include only participants over 40 years of age and those in high risk groups due to family history of glaucoma, black ethnicity, myopia and diabetes.

Index tests eligible for inclusion: These assessed structure (ophthalmoscopy, optic disc photography, retinal nerve fibre layer (RNFL) photography, Heidelberg retina tomograph (HRT), scanning laser polarimetry, optical coherence tomography (OCT) and retinal thickness analyser (RTA)); function (frequency doubling technology (FDT), motion detection perimetry (MDP), oculokinetic perimetry (OKP), short-wavelength automated perimetry (SWAP) and white-on-white standard automated perimetry (SAP)); and intraocular pressure (Goldman applanation tonometry (GAT), non-contact tonometry (NCT) and tonopen).

Reference standards: The primary reference standard was clinically confirmed open angle glaucoma at follow-up. Studies that used ophthalmologist-diagnosed open angle glaucoma without follow-up confirmation as the reference standard were also accepted for inclusion. Reference standards could include one or more index tests. Studies using technology-based diagnostic tests alone as the reference standard were excluded.

Included studies were required to report at least one of the following outcomes: sufficient data to construct 2x2 contingency tables of test performance; adverse events; acceptability to patients; and reliability of tests.

Assessment of study quality
Two reviewers independently assessed the quality of included diagnostic accuracy studies using a topic-specific adaptation of the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool (reported in full). Any disagreements were resolved by consensus or consultation with a third party. Studies were deemed higher quality if they included a representative spectrum of participants and avoided verification and review biases.

Data extraction
One reviewer extracted data on index tests and reference standard and numbers of true positives, false negatives, false positives and true negatives using a piloted data extraction form. A second reviewer advised where there was uncertainty. Sensitivity, specificity and diagnostic odds ratio (DOR) were calculated and presented for each data set (some studies reported data for multiple thresholds) in all included diagnostic accuracy studies. Relative diagnostic odds ratio was calculated where studies compared two or more index tests in the same participants.

**Methods of synthesis**

**Performance of screening tests:**

Summary receiver operating characteristic (SROC) curves were generated for all tests that were evaluated by two or more studies, using the hierarchical SROC (HSROC) model fitted in WinBUGS software. A second set of curves was produced for a test-specific common cut off, where a cut off that was broadly similar across reporting studies was selected for each test (selection was made by discussion with two ophthalmologists). Summary sensitivity, specificity and diagnostic odds ratio, with 95% credible intervals, were produced for each model. A simplified model assuming a symmetrical SROC curve was used where limited data caused convergence problems.

Statistical investigation of potential sources of heterogeneity was not undertaken due to the limited number of studies in each model. Where two or more higher-quality studies were available for a test, an additional model using only these studies was fitted.

Comparisons of the performance of different tests between studies (indirect comparisons) were made by including all tests with two or more studies in a single HSROC model; pair-wise differences between the tests were assessed from median differences in sensitivity and specificity and 95% credible intervals.

**Results of the review**

Forty studies published in 46 reports met the inclusion criteria for studies assessing the performance of screening tests. Twenty were population-based studies representative of screening. Eight studies (six population-based and two cohort) met the criteria for higher quality studies.

Pooled estimates of sensitivity for individual tests ranged from 46% (95% credible interval: 22% to 71%) for Goldman applanation tonometry to 92% (95% credible interval: 65% to 95%) for frequency doubling technology C-20-1.

Pooled estimates of specificity ranged from 75% (95% credible interval: 57% to 87%) for frequency doubling technology C-20-5 to 95% (95% credible interval: 80% to 97%) for Goldman applanation tonometry.

Frequency doubling technology C-20-1 had the highest sensitivity and one of the highest specificities (94%, 95% credible interval: 73% to 99%). Full accuracy data for all included tests were presented in the report. Some accuracy data for different stages of glaucoma were also presented.

Differences in sensitivity and specificity between higher quality and other studies were inconsistent across different tests. There was no clear single cause for heterogeneity, but the authors listed the potential contributing factors as: differences in population, study design, setting, prevalence and severity of glaucoma; differences in the reference standard and tests included in the same data set; and the extent to which studies were affected by potential sources of bias.

For direct comparisons of test performance, standard automated perimetry suprathreshold had higher sensitivity and lower specificity than Goldman applanation tonometry (two studies), higher sensitivity and similar specificity to Heidelberg retina tomograph (one study) and lower sensitivity and higher specificity than optic disc photography (one study). Standard automated perimetry threshold had higher sensitivity and lower specificity than frequency doubling technology (two studies), but lower sensitivity and specificity than Heidelberg retina tomograph.

For indirect comparisons, four sensitivity comparisons and two specificity comparisons showed statistically significant differences: sensitivity of frequency doubling technology C-20-1 was higher than ophthalmoscopy and Goldman applanation tonometry; standard automated perimetry threshold and Heidelberg retina tomograph had higher sensitivity than Goldman applanation tonometry; and Goldman applanation tonometry had higher specificity than standard...
automated perimetry threshold or frequency doubling technology C-20-5.

Imprecision in the pooled estimates derived from meta-analytic models meant that it was not possible to identify a single test or group of tests as the most accurate.

Reported uptake levels for screening ranged from 28.3% to 99.5%.

**Cost information**
The reported cost of an invitation to attend screening was £10.45. The average cost for an ophthalmologist outpatient visit was £65 (range £38 to £195).

**Authors’ conclusions**
For a low-prevalence disease, a screening test needed high specificity. Ophthalmoscopy, optic disc photography, retinal nerve fibre layer photography, Heidelberg retina tomograph, frequency doubling technology C-20-1, oculokinetic perimetry, standard automated perimetry suprathreshold and Goldman applanation tonometry were all found to have specificity of at least 85%. However, paucity and heterogeneity of data precluded determination of whether any one test was more accurate than others.

**CRD commentary**
The review addressed a wide ranging research question with the overall aim of assessing whether or not screening for open angle glaucoma met National Screening Committee Criteria. Inclusion criteria were clearly defined and appropriate and review methodology was clearly reported, with appropriate measures taken to minimise error and bias. The restriction of searches to English-language studies may have resulted in loss of some data, although this may have had limited relevance as the review specified an evaluation for the UK setting. Most of the reported data were from diagnostic accuracy studies and related to a large number of tests for open angle glaucoma. Details of the studies were fully reported, an appropriate quality assessment tool was used and the results of methodological quality assessment were fully reported. The methods used to generate summary measures of accuracy and to draw comparisons between studies were appropriate and clearly described. The authors concluded that a number of tests appeared to meet the required levels of specificity, but data were insufficient to draw firm conclusions on relative accuracy. They further concluded that screening for open angle glaucoma did not meet NSC criteria. The conclusions were appropriate to the data presented.

**Implications of the review for practice and research**
**Practice**: The authors did not make any recommendations for practice.

**Research**: The authors specified further research to generate data to populate the economic model presented. In order of priority the research was: a feasibility study of interventions to improve detection (the study should consider optimal testing strategy, acceptability of testing and interventions to improve uptake and associated benefits and harms); an RCT of interventions to improve screening uptake based upon the findings of the feasibility study.

**Bibliographic details**

**PubMedID**
17927922

**Original Paper URL**
http://www.hta.ac.uk/execsumm/summ1141.htm

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Age Factors; Cost-Benefit Analysis; Disease Progression; Glaucoma, Open-Angle /diagnosis /epidemiology /prevention & control; Humans; Sensitivity and Specificity; Technology Assessment, Biomedical /economics; Time Factors; Treatment Outcome; Vision Screening /economics /standards

**AccessionNumber**
12008008018

**Date bibliographic record published**
09/08/2008

**Date abstract record published**
19/08/2009

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.