Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis

CRD summary
The review compared the diagnostic accuracy of candidate test for screening for open angle glaucoma and found that no test, or group of tests, was superior. Despite some limitations in the criteria for study selection and review process, these conclusions reflect the limited, poor quality data presented and are likely to be reliable.

Authors' objectives
To compare the diagnostic accuracy of potential screening tests for open-angle glaucoma.

Searching
MEDLINE, EMBASE, Science Citation Index, BIOSIS Previews and the Cochrane Register of Controlled Trials (CENTRAL) were searched from inception to November 2005. Search strategies were published in full elsewhere (see Other Publications of Related Interest field). Full-text searches of five journals were undertaken. Bibliographies of included articles were also screened to identify additional studies. Searches were restricted to English language publications.

Study selection
Studies that assessed the diagnostic accuracy of tests for detecting open-angle glaucoma, in screening populations aged over 40 years (i.e. no selection and no previous tests had been performed), or in patients with suspected glaucoma, were eligible for inclusion. Randomised controlled trials (RCTs) and observational studies (cohort studies and case-control studies with representative control groups) were included. The reference standard was required to be either open-angle glaucoma confirmed at follow-up or diagnosed by an ophthalmologist. Included studies were required to report sufficient data for the calculation of 2x2 contingency tables (numbers of true positives, false negatives, false positives, and true negatives).

Eligible index tests were: structural measures (ophthalmoscopy, optic disc photography, retinal nerve fibre layer photography, Heidelberg retinal tomography version II, GDx VCC retinal nerve fibre layer analyzer, optical coherence tomography and retinal thickness analyzer); functional measures (oculokinetic perimetry, white-on-white standard automated perimetry including suprathreshold and threshold, short wave-length automated perimetry, frequency-doubling technology, and motion-detection perimetry); and measures of intraocular pressure. Where reported, 51% of study participants were women and the median age across studies was 60.5 years (range 13 to 97 years).

Two reviewers independently selected studies for the review. Any disagreements were resolved by discussion with a third reviewer (details of study selection provided in the HTA report, see Other Publications of Related Interest).

Assessment of study quality
Two reviewers independently assessed the quality of the included studies using a version of the QUADAS (Quality Assessment of Diagnostic Accuracy Studies Assessment) tool adapted for this review. Disagreements were resolved by consensus or consultation of a third reviewer. A 'higher quality' study was considered one which met the QUADAS criteria for a representative patient spectrum, avoidance of verification biases, and avoidance of review biases.

Data extraction
Data were extracted to calculate sensitivity, specificity and diagnostic odds ratios (DOR), with 95% credible intervals (CrIs).

Data were extracted by one reviewer, with advice and validation provided by a second reviewer in the event of uncertainty.

Methods of synthesis
The most frequently reported diagnostic threshold, for each test, was selected by two ophthalmologists.
Summary receiver operating characteristic curves were estimated, using a hierarchical summary receiver operating characteristic model, for all tests where sensitivity and specificity at the selected threshold were reported by two or more studies.

Summary sensitivity, specificity and diagnostic odds ratio, with 95% credible intervals (CrI), were calculated, at the operating point, for each model.

Comparisons of performance between tests were made, either using direct comparison studies (where participants received all tests or were randomized to different tests), or by indirect comparison. For indirect comparisons, data for all tests assessed by two or more studies were included in a single hierarchical summary receiver operating characteristic model; pair wise differences in sensitivity and specificity were assessed from the median differences and corresponding 95% credible intervals.

Sensitivity analyses were conducted, where only higher quality studies were included.

Results of the review
Forty studies, published in 46 reports, and including more than 48,000 participants (more than 39,000 included in the meta-analyses), were included in the review. Twenty studies were conducted in screening populations and 20 in patients with suspected open angle glaucoma (of which eight were cohort studies and 12 were case-control). Eight studies met the criteria for 'higher quality' studies.

Tests included in the hierarchical summary receiver operating characteristic model were ophthalmoscopy (seven studies); optic disc photography (six studies); retinal nerve fibre layer photography (four studies); Heidelberg retinal tomography II (three studies); oculokinetic perimetry (four studies); standard automated perimetry (14 studies); frequency-doubling technology (eight studies); Goldmann applanation tonometry (nine studies); and non-contact tonometry (one study). Summary sensitivities ranged from 46% (95% CrI 22 to 71) for Goldmann applanation tonometry with a threshold of 21mmHg intraocular pressure, to 92% (95% CrI 65 to 99) for frequency-doubling technology C-20-1 with a threshold of one abnormal point. Summary specificities ranged from 75% (95% CrI 57 to 87) for frequency-doubling technology C-20-5 with a threshold of one abnormal point, to 95% (95% CrI 89 to 97) for Goldmann applanation tonometry with a threshold of 21mmHg intraocular pressure.

Six studies directly compared two or more tests. Standard automated perimetry (either suprathreshold or threshold) was included as a comparator in all six. Comparing diagnostic odds ratios, standard automated perimetry performed better than Goldmann applanation tonometry in one study but worse in another, better than Heidelberg retinal tomography II in one study but worse in another, worse than frequency-doubling technology C-20-5 and frequency-doubling technology C-20-1 in one study each, and similarly to optical disc photography in one study.

Indirect comparisons indicated that frequency-doubling technology C-20-1 was significantly more sensitive than both ophthalmoscopy and Goldmann applanation tonometry, and that both standard automated perimetry threshold and Heidelberg retinal tomography II were significantly more sensitive than Goldmann applanation tonometry. Goldmann applanation tonometry was significantly more specific than both frequency-doubling technology C-20-5 and standard automated perimetry threshold. Because of the imprecision in estimates (wide credible intervals) no test, or group of tests, was clearly most accurate at the 5% significance level.

For both standard automated perimetry and frequency-doubling technology C-20-5, analysis of higher quality studies produced lower estimates of sensitivity and specificity. For optical disc photography, higher quality studies reported similar sensitivity, but lower specificity. For Heidelberg retinal tomography II, higher quality studies reported higher sensitivity and slightly lower specificity.

Authors' conclusions
No test, or group of tests, was clearly superior for glaucoma screening.

CRD commentary
This review addressed a clearly stated research question, defined by relevant inclusion criteria. The restriction of the search to published English language studies left open the possibility of language and publication biases. With the exception of quality assessment, review processes did not appear to have included measures to reduce error and bias.
The methodological quality of included studies was assessed using an appropriate tool, reported in full and incorporated in the meta-analyses. Appropriate, robust meta-analytic methods were used to generate pooled estimates of diagnostic performance metrics and to compare performance between tests. The authors’ conclusions reflect the data presented and are likely to be broadly reliable, but should be considered in the light of potential for omission of some relevant data.

Implications of the review for practice and research

Practice: The authors stated that no one test, or group of tests, was clearly most accurate. From the limited, poor quality data available, ophthalmoscopy, standard automated perimetry, retinal photography, and Goldmann applanation tonometry performed relatively poorly, whilst frequency doubling technology, (C-20-1), Heidelberg retinal tomography II, and oculokinetic perimetry were appeared to have better diagnostic performance than other candidate tests.

Research: The authors stated that future studies should directly compare the performance of the most promising tests in relevant populations.

Funding
UK National Institute for Health Research Health Technology Assessment programme, project number 04/08/02.

Bibliographic details

PubMedID
18614810

DOI
10.1167/iovs.07-1501

Original Paper URL
http://www.iovs.org/content/49/12/5373

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Diagnostic Techniques, Ophthalmological; False Positive Reactions; Glaucoma, Open-Angle /diagnosis; Humans; Intraocular Pressure; Ophthalmoscopy; Predictive Value of Tests; Reproducibility of Results; Sensitivity and Specificity; Tomography; Tonometry, Ocular; Visual Field Tests

AccessionNumber
12009103789

Date bibliographic record published
05/08/2009

Date abstract record published
25/11/2009
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.