Comparative Effectiveness Review
Number xx

Comparative Effectiveness of Screening for Glaucoma

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

This information is distributed solely for the purposes of predissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services.

Prepared by:
[Redacted**]**

Investigators
[Redacted**]**

AHRQ Publication No. xx-EHCxxx-EF
August 2011

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested citation: Comparative Effectiveness of Screening for Glaucoma. Comparative Effectiveness Review No. xx. (Prepared by the XXXX Evidence-based Practice Center under Contract No. xxx-xx-xxxx.) AHRQ Publication No. xx-EHCxxx. Rockville, MD: Agency for Healthcare Research and Quality. August 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP). AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting Comparative Effectiveness Reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered. Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://effectivehealthcare.ahrq.gov/reference/purpose.cfm AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project.

Key Informants

Technical Expert Panel

Peer Reviewers

AHRQ Contacts
Shilpa Amin, MD, MBsc, FAAFP
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Structured Abstract

Objectives

Glaucoma is a progressive disease of the optic nerve characterized by a typical appearance of the optic nerve and by visual field defects that correspond to the areas of optic nerve structural damage. Open angle glaucoma (OAG) is the most common form of this progressive optic neuropathy which leads to variable degrees of visual impairment and is the second leading cause of irreversible blindness worldwide effecting over 60 million people worldwide with a prevalence in the United States of 2% at age 50 and 7% at age 70. The onset of OAG is symptomless. It is estimated that half of those who have glaucoma are undiagnosed, leading to the desire to have a population screening for this disease. The diagnosis of this disease can be challenging with frequent extended time periods for definitive diagnosis. This comparative effectiveness review includes both the diagnostic accuracy of screening examinations and diagnostic tests for OAG. We also searched for studies that address the effects of screening on visual impairment, patient reported outcomes, intraocular pressure, visual field loss, optic nerve damage, as well as possible adverse effects of screening.

Data Sources

We searched MEDLINE, EMBASE, LILACS, and CENTRAL from the beginning of each database through 29 October 2010. We also conducted a search in MEDLINE and CENTRAL for systematic reviews that address the key questions of interest (through 2 March 2011) and screened a previously published existing database of eye and vision systematic reviews to identify relevant open-angle glaucoma systematic reviews published prior to 2009.

Review Methods

We included studies of adult ("adult" as defined by included studies) asymptomatic participants in general or high risk populations. We included randomized controlled trials (RCTs), quasi randomized controlled trials, comparative and non-comparative cohort studies, and case control studies. For diagnostic test accuracy studies we included case control studies and study designs in which all tests (including the index, comparator, and reference standard) were performed on all participants, and designs in which participants were randomized to one test (among the index and potential comparator(s)) but all were evaluated with a reference standard. We included studies of ocular examinations (screening tests) that examine the structure or function of the optic nerve that were conducted alone or in any possible combination (including multi-component simultaneous or sequential testing).We excluded studies of tests/devices that are either not commercially available for screening or are not commonly used in the diagnosis of glaucoma. We excluded studies that examined technical aspects of included devices, e.g., usability, technician training. For diagnostic test accuracy studies, the reference standard had to include a clinical examination with measurement of intraocular pressure, assessment of the visual field, assessment of the optic nerve head and or retinal nerve fiber layer, or review of fundus photographs. We included the outcomes of visual impairment, intraocular

pressure, optic disc damage, visual field progression, the harms of treatment as well as the sensitivity and specificity of candidate screening or diagnostic tests.

Two reviewers independently assessed titles and abstracts resulting from the literature searches according to the inclusion criteria. Citations tagged as 'unsure' by both reviewers, 'unsure' by one reviewer and 'include' by the other, or 'include' by both reviewers, were promoted to full-text screening. Two reviewers independently applied the same inclusion criteria as used during abstract screening. Data abstraction forms were designed and pilot tested. One reviewer extracted descriptions of the study, including details about the population, intervention(s) and outcomes of interest and a second reviewer verified the data. Throughout the process, we resolved any disagreements regarding inclusion through discussion or, as needed, adjudicated unresolved conflicts during a team meeting.

Results

The electronic search of MEDLINE identified 64 systematic review titles and abstracts. The Li 2009 database included 105 systematic review titles and abstracts. We excluded 167 of the 169 titles and abstracts and identified two systematic reviews addressing questions considered in this comparative effectiveness review. One systematic review evaluated the diagnostic accuracy of candidate screening tests of OAG. A second review evaluated the effect of screening-based programs on the prevention of optic nerve damage.

The electronic searches that we conducted for concurrent comparative effectiveness reviews of screening and treatment for OAG, identified a total of 4680 primary study titles and abstracts. We retrieved the full text of 630 articles and assessed the studies for inclusion in the review. We included 72 primary studies addressing the diagnostic accuracy of candidate screening tests for the detection of OAG that were not included in the Burr 2007 systematic review. Newer imaging techniques reviewed include confocal scanning laser ophthalmoscopy, scanning laser polarimetry, and optical coherence tomography(OCT). The sensitivity of Standard automated perimetry (SAP) was higher than Goldmann tonometry, similar to Heidelberg retina tomography (HRT) and lower than disc photos or Frequency doubling technology (FDT). The specificity of SAP was higher than disc photos and FDT, similar to HRT, and lower than Goldmann tonometry. Some direct comparisons of screening tests could not be performed to variability in populations studied and reported thresholds. We identified no other studies that evaluated our listed objectives.

Conclusions

We did not find direct or indirect links between glaucoma screening and visual field loss, visual impairment, optic nerve damage, intraocular pressure, and patient-reported outcomes. There have been improvements in screening devices for assessing functional optic nerve deficits that are more rapid than the SAP such as the FDT and promising structural imaging tests such as OCT. Yet even with these improvements there is limited evidence that screening for OAG can reduce intermediate or final health outcomes including visual impairment, and optic nerve function.

Contents

Executive Summary ES-1-18
Introduction 1
Background 1
Purpose for Evidence Report 2
Key Questions 2
Methods 3
Topic Development 3
Analytic Framework 3
Study Selection 4
Search Strategy 7
Risk of Bias Assessment 8
Rating Body of Evidence 9
Data Synthesis 9
Results 10
Key Question 1 13
Key Question 2 13
Key Question 3 13
Summary 13
Evidence from Systematic Reviews 13
Analysis of Primary Studies 16
Grading of Evidence 20
Conclusion 20
Key Question 4 21
Key Question 5 21
Evidence from Systematic Reviews 21
Analysis from Primary Studies 21
Key Question 6 21
Summary and Discussion 21
Future Research Needs 23

References
References included 25
Abbreviations 33

Figures

Figure 1. Analytic Framework 4
Figure 2 Literature Search and Review Process 11

Tables

Table 1 Grading of Evidence 20

Appendices

Appendix A. Search Strategy
Appendix B. Screen and Data abstraction Forms
Appendix B. Screen and Data abstraction Forms
Appendix C. Evidence Tables
Appendix D. Excluded Articles
Appendix F. Included Devices

Effective Health Care

Comparative Effectiveness of Screening for Glaucoma

The Effective Health Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov. Ahrq.gov/reports/final.cfm

Executive Summary

Background

Glaucoma is a leading cause of visual impairment and blindness and is estimated to affect 60.5 million people worldwide by 2010.[2 3] Glaucoma may be classified by optic nerve damage, visual field loss and elevated intraocular pressure, but these mechanisms need not occur in tandem to confirm diagnosis.[4] Damage is irreversible, so early detection may prevent severe vision loss. Open-angle glaucoma (OAG), the most common subtype of the disease, impacts over 2.5 million people in the United States with a median age-adjusted prevalence of 4.6 percent and 1.6 percent among black and white persons respectively (based on year 2000 estimates).[5]

Unfortunately, only half of the prevalent cases of glaucoma have been identified in the United States (U.S.).[5] This is due to at least two factors. First, glaucoma is an asymptomatic disease that patients do not notice until the onset of advanced disease accompanied by severe vision loss. Second, there is no single test to identify persons with glaucoma. Due to the large burden of undiagnosed disease and the number of persons disabled by glaucoma, there is strong interest in developing effective screening strategies. Ocular examinations such as tonometry, perimetry, and direct ophthalmoscopy are performed alone or as a part of a multicomponent screening test for glaucoma. However, an inability to identify adequately sensitive and specific tests has hampered the establishment of screening-based programs.

The U.S. Preventive Services Task Force (USPSTF) reported on the safety and efficacy of screening for chronic open-angle glaucoma in March 2005 and concluded that intraocular pressure measurement and optic nerve head assessment alone have limited effectiveness as population-based screening tools.[6 7] The USPSTF also concluded that current methods used to assess visual field loss may be impractical for population-based screening due to the length of time required for testing and the challenge of equipment portability. Since the mid 2000s, there

have been advances in the devices used to test optic nerve structure and function.[6 7] Because of this new evidence, it is appropriate to re-evaluate the effectiveness and safety of population-based glaucoma screening. For example, studies of frequency-doubling technology, used to assess visual field loss, were underway at the time of the evidence synthesis for the 2005 guidelines.[7] This report includes evidence regarding the diagnostic accuracy of frequency-doubling technology and other screening examinations used alone or in combination to aid in the identification of persons with OAG.

Objectives

The objective of this review was to summarize the evidence regarding the effectiveness and safety of screening based programs for open-angle glaucoma with a specific focus on the effects of screening on visual impairment, patient reported outcomes, intraocular pressure, visual field loss, optic nerve damage, and adverse effects. This review also includes a summary of the diagnostic accuracy of screening examinations and tests for open-angle glaucoma.

Key Questions

Screening for a medical condition in asymptomatic individuals may be considered to be beneficial when 1) the condition has a significant individual or population burden; 2) the medical condition is associated with adverse impacts on the health of the individual; 3) there is an accurate test that detects the condition during the asymptomatic or early clinical stage; 4) treatment at this asymptomatic or early stage improves important health outcomes more than treatment once symptomatic; and 5) harms of screening and early intervention are limited. Following these requirements we considered and compared, where possible, whether patients were treated or not treated for early stages of glaucoma.

Key Question 1 (KQ1)

KQ1a: Does a screening-based program for open-angle glaucoma lead to less visual impairment when compared to no screening program?
KQ1b: How does visual impairment vary when comparing different screening-based programs for open-angle glaucoma?

Key Question 2 (KQ2)

KQ2a: Does a screening-based program for open-angle glaucoma lead to improvements in patient-reported outcomes when compared to no screening?
KQ2b: How do patient-reported outcomes vary when comparing different screening-based programs for open-angle glaucoma?

Key Question 3 (KQ3)

What is the predictive value of screening tests for open-angle glaucoma?

Key Question 4 (KQ4)

KQ4a: Does a screening-based program for open-angle glaucoma lead to reductions in intraocular pressure when compared to no screening program?

KQ4b: How does intraocular pressure vary when comparing different screening-based programs for open-angle glaucoma?

Key Question 5 (KQ5)

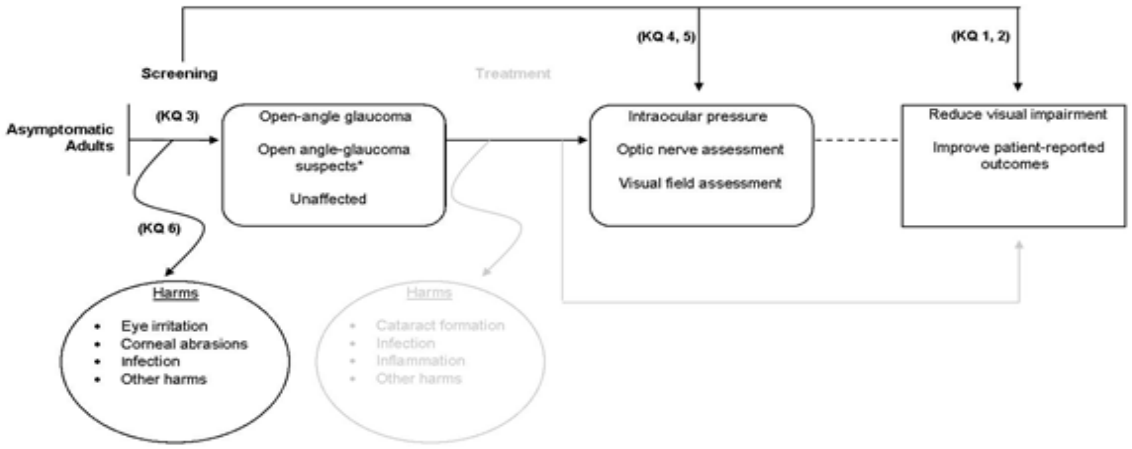
KQ5a: Does a screening-based program lead to a slowing of the progression of optic nerve damage and visual field loss when compared to no screening program?
KQ5b: How do optic nerve damage and visual field loss vary when comparing different screening-based programs for open-angle glaucoma?

Key Question 6 (KQ6)

What are the harms associated with screening for open-angle glaucoma?

Analytic Framework

Figure 1. Analytic Framework for Screening for Open-Angle Glaucoma



The analytic framework (Figure 1) is a modified version of a larger framework depicting the impact of both screening and treatment for open-angle glaucoma. This figure focuses on the screening aspects of the framework. It depicts the key questions within the context of the inclusion criteria described in the following sections. In general, the figure illustrates how screening-based programs (which may incorporate treatment when indicated) may reduce visual impairment (KQ1) and/or improve patient reported outcomes (KQ2), reduce intraocular pressure (KQ4) and possibly slow the progression of optic nerve damage and/or visual field loss (KQ5). The figure also illustrates the potential predictive value of candidate screening tests to detect open-angle glaucoma and open-angle glaucoma suspects (KQ3). Finally, the potential harms of screening (KQ6) are illustrated in the framework.

Methods

Input from stakeholders

The Agency for Healthcare Quality and Research (AHRQ) requested that the Evidence-based Practice Center (EPC) assist with the formulation and refinement of the Comparative Effectiveness Review (CER) topic "Effectiveness of screening and treatment for glaucoma."

In consultation with AHRQ, the EPC identified a small group of stakeholders to serve as members of a Key Informant Group. The Key Informant Group helped shape the key questions relevant to the topic by providing input regarding the populations and clinical subgroups, interventions, and outcomes of interest to clinicians, policy makers, payers, and consumers.

The EPC investigators incorporated the feedback of the Key Informants into a draft of the key questions, analytic framework, and inclusion criteria which was posted to the AHRQ website for public comment from 22 April to 20 May 2010. Inclusion criteria were finalized after consideration of the public comments received.

A Technical Expert Panel (TEP) was selected to provide expertise in and perspectives specific to the epidemiology of glaucoma, research methods, including systematic reviews and diagnostic studies, clinical practice, including screening persons for and treating those with glaucoma, and the development of clinical practice guidelines for screening and treatment of glaucoma.

The TEP reviewed a protocol outlining a proposed methodological approach for the completion of the comparative effectiveness review, provided information to the EPC to aid in the refinement of the inclusion criteria and literature search strategies and recommended approaches to specific issues as requested by the EPC. The final protocol entitled *Comparative Effectiveness of Screening For Open-Angle Glaucoma* was posted to the AHRQ website on 16 November 2010.

Data sources and selection

Study Selection

Types of studies

We included randomized controlled trials (RCTs), quasi randomized controlled trials, comparative and non-comparative cohort studies and case control studies, for Key Questions 1 through 6. For Key Question 3 we also included cross-sectional studies, study designs in which all tests (including the index, comparator, and reference standard) were performed on all participants, and designs in which participants were randomized to one test (among the index and potential comparator(s)) but all were evaluated with a reference standard.[1]

We also considered the inclusion of systematic reviews that address the key questions.

We excluded studies that addressed the following:

- Studies of the prevalence of glaucoma in a specific population unless studies also include tests of diagnostic accuracy
- Studies of disease progression that did not include participants previously screened for glaucoma
- Studies of risk factors for disease

Types of participants

We included studies of adult ("adult" as defined by included studies) asymptomatic participants in general or high risk populations. Asymptomatic general populations included populations not previously tested, those not diagnosed with glaucoma, and those not presenting with symptoms known to be related to a diagnosis of glaucoma. Asymptomatic high risk populations included those not previously tested, diagnosed or presenting with symptoms known to be related to glaucoma but also included those with a family history of glaucoma, specific racial/ethnic groups, older age, and specific ocular or other medical conditions as defined by included studies (e.g., diabetes).

We also included studies of suspected open-angle glaucoma subpopulations which included participants identified from prior testing as possibly having glaucoma or as having a risk factor for glaucoma, e.g., high IOP, but with an unconfirmed diagnosis. We excluded studies of participants with known glaucoma at the time of screening (Key Questions 1,2,4 and 5) and those that included the healthy eye of a participant with known glaucoma (Key Question 3). We considered studies that enrolled healthy volunteers in addition to those with suspected open-angle glaucoma but excluded studies of healthy volunteers only.

Interventions

We included studies of the following ocular examinations (screening tests) conducted alone or in any possible combination (including multi-component simultaneous or sequential testing):

- Tonometry (contact and non-contact tonometry)
- Perimetry (including short-wavelength, high-pass, motion, flicker perimetry, yellow and blue perimetry)
- Direct and indirect ophthalmoscopy
- Fundus photography or computerized imaging of the posterior pole (optic disc or retinal nerve fiber layer assessments); Also includes optical coherence tomography (OCT), retinal tomography, scanning laser polarimetry.
- Pachymetry (corneal thickness measurement) when used in conjunction with another test to diagnose glaucoma. We excluded studies where pachymetry is used alone.

We excluded studies of the following tests/devices that are either not commercially available for screening or are not commonly used in the diagnosis of glaucoma:

- Electroretinography
- Tests of contrast sensitivity and visual acuity
- Color tests
- Water drinking tests
- OCT 1 and OCT 2 imaging systems
- Heidelberg Retina Tomograph 1 (HRT 1 - confocal scanning laser ophthalmoscope)
- GDx (Glaucoma Diagnosis, early version) scanning laser polarimeter

We excluded studies that examined technical aspects of included devices, e.g., usability, technician training.

Comparators/Reference standards

Key Questions 1, 2, 4, 5, and 6 explore comparisons of the interventions mentioned above (conducted alone or in any possible combination as a part of a screening-based program) to no screening program (including usual care, e.g., opportunistic case finding) and to different screening-based programs (abovementioned screening tests conducted alone or in any possible combination). Key Question 3 explores comparisons of screening/diagnostic tests to the reference standards of confirmed open-angle glaucoma at the time of follow-up or open–angle glaucoma requiring treatment (diagnosed by an ophthalmologist using objective assessments). The diagnosis should have included a clinical examination with measurement of intraocular pressure, assessment of the visual field, assessment of the optic nerve head and or retinal nerve fiber layer, or review of fundus photographs. We considered other methods to confirm diagnosis as defined by included studies if the examinations/tests were specified in the manuscript. We acknowledge that there is no consensus on the gold standard test (or combination of tests) for the identification of patients with open-angle glaucoma. We adapted the reference standards for Key Question 3 from a diagnostic test accuracy review conducted by Burr (2007).[1]

Outcomes

Key Question 1

Primary outcome: The proportion of participants with moderate, severe, and profound visual impairment (as defined in the International Classification of Diseases, Clinical Modification, 9th Revision[8]). We also considered other measurements of visual impairment as defined by included studies.

Secondary outcome: We also compared visual acuity outcomes among the groups of interest (Early Treatment Diabetic Retinopathy Study (ETDRS), Snellen, or any other valid chart that yields scores that can be converted to Snellen fractions or logMAR values) as reported in included studies (e.g., mean visual acuity or proportion of participants in prespecified visual acuity categories).

Key Question 2

We compared the participants' mean total or relevant item/subscale scores as measured by any validated questionnaire, e.g., National Eye Institute Visual Functioning Questionnaire (NEI-VFQ), for the following patient-reported outcomes among the treatment groups of interest:

- Vision-related quality of life (vision-related functional loss as well as the impact of functional loss on activities of daily living) (*Primary outcome*)
- Patient satisfaction (*Secondary outcome*)

Key Question 3

The primary outcome was the number of participants in the following categories: true positives, true negatives, false positives, and false negatives. We also considered studies that include sensitivity and specificity values only.

Key Question 4

The primary outcome of interest was the difference in the mean intraocular pressure among the groups of interest.

Key Question 5

The primary outcomes of interest were a) the proportion of participants with progressive optic nerve damage as defined by included studies and as observed via fundus photography or other imaging of the posterior pole and b) the proportion of participants with progression of visual field loss as defined by included studies.

Key Question 6

The primary outcome of interest was the proportion of participants experiencing the following adverse events among the treatment groups of interest (adapted from the USPSTF - http://www.ahrq.gov/clinic/uspstf05/glaucoma/glaucrs.htm):

- Eye irritation
- Corneal abrasions
- Infection (e.g., endophthalmitis)
- Distortion of sense of taste (due to anesthetic use)
- Exam apprehension
- Psychological effects related to a glaucoma diagnosis or misdiagnosis
- Harms related to overdiagnosis

We also planned to include other harms as reported in included studies

Timing of outcome

We assessed outcomes for Key Questions 1, 2, 4, and 5 at one year of follow-up and at annual intervals thereafter. There was no minimum length of follow-up for outcomes related to Key Questions 3 and 6.

Setting

Settings for this review included community screenings, non-eye care health provider settings, eye care provider clinical settings (ophthalmologists and optometrists), and telemedicine.

Search Strategy

We searched the following databases for primary studies: MEDLINE, EMBASE, LILACS (Latin American and Caribbean Literature on Health Sciences), CENTRAL (the Cochrane Central Register of Controlled Trials). We developed a search strategy for MEDLINE, accessed via PubMed, based on an analysis of the medical subject headings (MeSH) terms and text words of key articles identified a priori and adapted this search strategy for searches of EMBASE (using EMTREE terms) and CENTRAL (Appendix A). We searched the literature without imposed language, sample size or date restrictions, but excluded non-English language studies at the time of full text review. We searched relevant systematic reviews to identify any additional inclusions. The databases were searched to include all articles from the beginning date for each database to 29 October 2010.

We also conducted a search in MEDLINE and CENTRAL for systematic reviews that address the key questions of interest. The search included the topic strategy as noted in the Appendix combined with the term "AND systematic[sb]" and was limited to systematic reviews published from 2009 to 2011. We searched MEDION (www.mediondatabase.nl) for related diagnostic accuracy reviews (Key Question 3). The search for systematic reviews was conducted on 2 March 2011.

We screened an existing database of eye and vision systematic reviews prepared by Li (2010) to identify relevant open-angle glaucoma systematic reviews published prior to 2009.[9] The search strategy for the existing database is included in the full text of the referenced doctoral thesis by Li (2010).

Updated searches for systematic reviews and primary studies will be conducted in September 2011.

Abstract Screening

We screened potentially relevant citations (primary studies and systematic reviews) using the web-based systematic review software, DistillerSR (http://systematic-review.net/). All citations identified by the search strategies were uploaded to DistillerSR. Two reviewers independently assessed titles and abstracts resulting from the literature searches according to the inclusion criteria. We classified the titles and abstracts as 'include', 'exclude' or 'unsure'. Disagreements about eligibility were resolved through discussion among reviewers. A copy of the abstract screening form is included in Appendix B.

Full-Text Screening

Citations tagged as 'unsure' by both reviewers, 'unsure' by one reviewer and 'include' by the other, or 'include' by both reviewers, were promoted to full-text screening. Two reviewers independently applied the same inclusion criteria as used during abstract screening. Non-English language articles were also removed from further consideration at this stage. We resolved any disagreements regarding inclusion through discussion or, as needed, adjudicated unresolved conflicts during a team meeting. A copy of the full-text screening form is included in Appendix B.

Data Extraction and Quality Assessment

Data Abstraction

Data abstraction forms were designed and pilot tested. One reviewer extracted descriptions of the study, including details about the population, intervention(s) and outcomes of interest, using the systematic review software, DistillerSR. A second reviewer verified the data. We resolved disagreements through discussion. A copy of the data abstraction forms is included in Appendix B.

Risk of Bias Assessment

We used the Cochrane Collaboration's tool for assessing the risk of bias of randomized and quasi randomized trials. Two reviewers assessed the included studies for sources of systematic bias according to the guidelines in Chapter 8 of the Cochrane Handbook for Systematic Reviews of Interventions and evaluated the studies for the following criteria: sequence generation and allocation concealment (selection bias), masking of participants, study investigators, and outcome assessors (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias), and other sources of bias.[10] Masking of investigators and participants might not be possible with some of the interventions being examined, but was noted when mentioned. We reported judgments for each criterion as 'Low risk of bias', 'High risk of bias' or 'Unclear risk of bias (information is insufficient to assess)'. The two reviewers resolved disagreements through discussion.

Two reviewers assessed the methodological rigor of observational studies using a modified version of the Newcastle Ottawa Scale (NOS).[11] The NOS includes domains to assess the quality of study group selection (representativeness, selection, case definitions), comparability of cohorts/cases and controls on the basis of the design or analysis, and ascertainment of exposure(s) or outcome(s) adequacy of follow-up, non-response rate and financial or other conflicts of interest. Each item query required a yes, no, or unable to determine/not reported response. Additionally, reviewers provided an overall assessment of the quality of each study as "good" "fair" or "poor" using the reporting bias, selection bias, and confounding domains as a basis for the assessment.

For Key Question 3, we used the QUADAS checklist, which is specific to risk of bias assessment for diagnostic accuracy studies.[12] The QUADAS tool includes 14 items that evaluate numerous domains including representativeness, inclusion/exclusion criteria, choice of reference standard, masked interpretation of results of tests and reference standard, and study withdrawal. We reported judgments for each checklist item as 'Yes', 'No', or 'Unclear.'

We used a tool adapted by Li (2010) from the Critical Appraisal Skills Program, Assessment of Multiple Systematic Reviews (AMSTAR), and the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement to assess the methodological quality of systematic reviews.[9] We used the following criteria, adapted from Li, to determine which were of sufficient quality to be considered for inclusion in this review: comprehensive search for primary studies (searches of more than one bibliographic database); conduct of a risk of bias assessment of included studies; performed appropriate analysis methods (no pooled arm analysis). A copy of the tool is included in Appendix B.

Data Synthesis and Analysis

Rating Body of Evidence

We assessed the quantity, quality and consistency of the body of available evidence addressing Key Questions 1 through 6. We used an evidence grading scheme recommended by the GRADE Working Group, adapted by AHRQ in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews (http://www.effectivehealthcare.ahrq.gov/index.cfm/search-

for-guides-reviews-and-reports/?pageaction=displayproduct&productid=328) and recently published in the Journal of Clinical Epidemiology.[13,14]

We considered the strength of the study designs with randomized controlled trials as the highest level of evidence, followed by observational studies. If an outcome was evaluated by at least one randomized controlled trial as well as observational studies our evidence grade was based on the randomized controlled trials and was followed by the quality of the cohort studies. If an outcome was evaluated by one or no randomized controlled trials, our evidence grade was based on the single randomized controlled trial in addition to the best available observational study.

We assessed the quality and consistency of the best available evidence, including assessment of the risk of bias in relevant studies, as well as aspects of consistency, directness, and precision as described in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews and by Owens (2010).[14] For each outcome of interest, two reviewers graded the major outcomes for each Key Question and then the entire team discussed their recommendations and reached consensus.

Data Synthesis

When we identified existing high quality systematic reviews that addressed the key questions, we cited these reviews as evidence and did not abstract and synthesize data from primary studies.

We searched for and summarized evidence from primary studies that were published or identified after the date of last search conducted for the systematic review. We also abstracted evidence from primary studies for interventions (diagnostic tests), comparisons, and outcomes that were not addressed by existing systematic reviews.

We adapted the recommendations of Whitlock (2008) for incorporating systematic reviews in complex reviews to include providing a narrative summary of the review methods (i.e., inclusion/exclusion criteria, search strategy, statistical methodology) and findings (i.e., number of studies included, quantitative and qualitative results) and in the instance of multiple reviews, evaluating the consistency across reviews addressing the same key question.[15]

Results

The electronic search of MEDLINE identified 64 systematic review titles and abstracts. The Li 2009 database included 105 systematic review titles and abstracts. We excluded 167 of the 169 titles and abstracts for the following reasons: did not address any of the key questions, narrative summary only, could not retrieve full text to assess, similar inclusion criteria but date of search for studies older than another included systematic review, and duplicate reference of included systematic review. We identified two systematic reviews for inclusion[1 88]. One systematic review[1] addressed the diagnostic test accuracy of candidate screening tests for the detection of open-angle glaucoma (Key Question 3) and the second review addressed the question of whether screening-based programs prevent optic nerve damage due to open-angle glaucoma when compared to no screening (Key Question 5) (Evidence Tables 1 and 2 in Appendix C).

The electronic searches that we conducted for concurrent comparative effectiveness reviews of screening and treatment for OAG, identified a total of 4680 primary study titles and abstracts. After de-duplicating the database and removing conference abstracts and book chapters (N = 1083), we reviewed 3,597 titles and abstracts. We retrieved the full text of 630 articles and assessed the studies for inclusion in the review. We included 72 primary studies addressing the diagnostic accuracy of candidate screening tests for the detection of OAG that were not included in the Burr 2007 systematic review (Key Question 3 - Evidence Tables 3 to 7 in Appendix C). (See Figure 2). We did not identify any primary studies eligible for inclusion for any other key question. A listing of the 558 excluded studies, with reason(s) for exclusion, is included in Appendix D.

A listing of devices from the primary studies is included in Appendix F. In summary the following number of diagnostic studies included the devices summarized in this comparative effectiveness review:

Tests of optic nerve structure

- Heidelberg Retina Tomograph (HRT) II (17 studies)
- HRT III (11 studies)
- Optical coherence tomography (OCT) (38 studies)
- Optic disc photography (2 studies)
- Retinal nerve fiber layer photography (2 studies)
- Scanning laser polarimetry (Glaucoma diagnosis – GDx device) (24 studies)

Tests of optic nerve function

- Frequency doubling technology (FDT) 24-2 perimetry (5 studies)
- FDT 30-2 (2 studies)
- FDT C-20 (3 studies)
- FDT N30 (4 studies)
- Humphrey Visual Field Analyzer (HFA) (11 studies)
- Goldmann applanation tonometry (1 study)
- Non contact tonometry (1 study),
- Octopus 301 perimeter (1 study)

Because there was appreciable variability in devices, parameters, thresholds, and measurement of outcomes reported in the primary studies of interest, we did not combine the results using meta-analysis and instead presented a narrative summary. A meta analysis of the included studies would shift the focus of discussion to a limited number of relevant device parameters and common thresholds (cut offs) that were commonly reported for any one device. As we are unable to determine which parameters are most important for identifying persons with OAG, and as our reported results would have been limited to a few parameters in a subset of studies, we chose to include in Evidence Tables 5 to 7 and discuss as appropriate the full complement of device parameters and thresholds as reported in the included studies. We summarize, where possible, the magnitude of validity across all parameters of interest for devices considered in this report.

Risk of bias of included systematic reviews and primary studies

The two systematic reviews were both of sufficient quality based on pre-specified quality criteria as follows (Appendix C). The two reviews included a comprehensive search for primary studies with searches of five databases conducted for both included systematic reviews. A risk of bias assessment of included studies was performed, and appropriate statistical analyses, when included, were conducted.

Twenty-eight of the 72 primary studies (39 percent) included in this review, enrolled a spectrum of participants that were representative of those that would be encountered in a screening setting. Including participants who are not representative of those one reasonably expects to encounter in a screening setting may lead to biased estimates of diagnostic performance and limit the generalizability of findings. The remaining studies (N = 44) were at high risk of bias as the investigators included healthy volunteers as well as those known to have glaucoma at the time of screening. Sixteen studies included an inappropriate reference standard or did not describe the reference in enough detail to make a determination of its appropriateness. The remaining 12 domains as described in the QUADAS tool are summarized in Appendix C.

Summary of findings

Key Question 1

We did not identify any studies that addressed whether participation in a open-angle glaucoma screening-based program leads to less visual impairment when compared to no screening or another screening-based program.

Key Question 2

We did not identify any studies that addressed whether participation in a open-angle glaucoma screening based program leads to improvements in patient-reported outcomes when compared to no screening or another screening-based program.

Key Question 3

Evidence from Systematic Reviews

Burr (2007) conducted a diagnostic test accuracy review of candidate screening tests for open-angle glaucoma.[1] Study inclusion criteria and pooled outcomes of sensitivity, specificity, and diagnostic odds ratios are listed in the evidence table (Evidence Tables 1 and 2 in Appendix C), but in summary, the investigators included 40 studies of more than 48,000 participants 40 years of age and older and those at high risk for the development of open-angle glaucoma based on demographic characteristics or comorbidities. The focus was on studies of participants likely to be encountered in a routine screening setting. Tests of optic nerve structure, optic nerve function, and intraocular pressure were included and compared to other individual or combination tests. The primary reference standard was confirmation of open-angle glaucoma at follow-up. Also considered was diagnosis of open-angle glaucoma requiring treatment. Outcomes were measures related to sensitivity and specificity as well as harms, acceptability and reliability of the tests. The included tests were as follows:

- Heidelberg Retina Tomograph (HRT) II
- Ophthalmoscopy
- Optic disc photography
- Retinal Nerve Fiber Layer (RNFL) photography
- Frequency Doubling Technology (FDT) (C-20-1) Perimetry
- FDT (C-20-5) Perimetry
- Oculokinetic Perimetry (OKP)
- Standard Automated Perimetry (SAP) Suprathreshold Test
- SAP Threshold Test
- Goldmann Applanation Tonometry (GAT)
- Non-contact tonometry (NCT)

There was significant statistical heterogeneity among the included studies for the majority of the tests with the exception of optic disc photography (sensitivity), HRT II (sensitivity and specificity), and FDT C-20-1 (sensitivity). The full text of the comparative effectiveness review that we have summarize in this Executive Summary includes a synopsis of findings for the tests included in the Burr 2007 review with summary estimates from meta analyses as reported.

Based on analyses of the common cut offs for each test, accuracies for high quality studies only (data not shown in this evidence report), combination tests, tests for glaucoma at specific stages, and direct and indirect comparisons of tests, Burr (2007) determined that the following tests may be potentially incorporated into a screening-based program:

- Optic disc photography
- HRT II
- FDT
- SAP
- GAT

Six studies included comparisons of optic disc photography, HRT II, SAP, FDT, and GAT with all comparing SAP to at least one of the tests. Burr 2007 concluded that:

- The sensitivity results at the common cut off revealed that SAP was generally:
    - Better than GAT
    - Similar or worse than HRT II
    - Worse than optic disc photography, FDT C-20-5, and FDT C-20
- The specificity results at the common cut off revealed that SAP was generally:
    - Better than optic disc photography, FDT C-20-5, and FDT C-20
    - Similar or worse than HRT II
    - Worse than GAT

Burr (2007) also acknowledged that given the "imprecision in estimates from the pooled meta-analysis models for the diagnostic performance of each test it was not possible to identify a single test (or even a group of tests) as the most accurate."

Detailed Analysis of Primary Studies

We undertook a search for primary studies as described in the Methods section to address the diagnostic test accuracy of candidate screening tests and identified 72 studies that were not

included in Burr 2007. The results, including sensitivity and specificity by parameter, are summarized in the evidence tables (Appendix C) and a narrative summary follows with a particular emphasis on *studies that identify early disease and those that examine newer and more frequently reported technologies*. When there were three or fewer included studies for a device, we summarized all included studies.

Tests of optic nerve structure

HRT II

Seventeen studies include measures of diagnostic accuracy for HRT II. [16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32] Two of these studies specifically focused on detecting early (or moderate) glaucoma, Naithani (2007) [24] and Uysal (2007) [26], are discussed in this narrative section. The remaining 15 studies are summarized in the evidence tables included in Appendix C.

Naithani (2007) enrolled 60 participants with glaucoma (30 early defects and 30 moderate visual field defects) and 60 healthy volunteers.[24] AUC values were reported in the range of 0.474 (disc area) to 0.852 (vertical cup to disc ratio).

Uysal (2007) enrolled 70 participants with early or moderate glaucoma and 70 healthy volunteers.[26] The range of sensitivity across 12 parameters was 47.1 percent (RNFL cross-sectional area) to 74.3 percent (Linear cup/disc area ratio) and the range of specificity was 47.1 percent (mean RNFL thickness) to 71.4 percent (Cup shape measure). The investigators concluded that some parameters have better sensitivity and specificity than others and that subgroup analyses by disc size revealed more difficulty distinguishing glaucoma for participants with smaller discs.

HRT III

Eleven studies examined the diagnostic accuracy of HRT III and two identified participants with early disease. [33 34 35 36 37 22 38 39 23 40 27]

Reddy (2009) identified in addition to 142 healthy volunteers, a subgroup of participants with early visual field loss defined as a mean deviation < 5dB (81 of 247 participants with glaucoma).[35] The sensitivity of the Glaucoma Probability Score (GPS) for distinguishing eyes with early field loss from healthy eyes was 67.9 percent and the Moorfields Regression Analysis (MRA) was 71.9 (at a fixed specificity of 92 percent). The investigators concluded that "MRA and GPS have similar ability to detect glaucomatous changes, and typically agree. The relative ease and sensitivity of the operator-independent GPS function of the HRT III may facilitate glaucoma screening."

Badala (2007) compared four imaging methods for their ability to distinguish early glaucoma from healthy eyes.[39] Forty six eyes of 46 participants with early OAG and 46 healthy volunteers were enrolled. Sensitivity ranged from 4 percent (parameter: reference height) to 70 percent (Frederick S. Mikelberg discriminant function and Reinhard O. W. Burk discriminant function) at a fixed specificity of 95 percent.

Optical Coherence Tomography (OCT)

Of the 38 included studies that investigated the diagnostic accuracy of OCT, [41 42 43 44 45 46 17 47 48 49 33 50 51 52 53 54 55 56 57 39 24 25 58 59 60 28 61 62 29 63 64 65 66 67 30 68 31] 33 considered the Stratus OCT, five included the Cirrus OCT, four considered the RTVue OCT, and one examined the OTI OCT. Across the 33 studies that examined the Stratus OCT (Stratus OCT considered a

more frequently reported technology and thus we have provided a summary of Stratus OCT studies), the sample size ranged from 26 to 95 participants with glaucoma and 37 to 128 healthy volunteers. One study also enrolled 130 participants with OHT. For the parameter average RNFL thickness, the range of sensitivity was 24 percent to 96 percent suggesting appreciable heterogeneity among the studies. The range of the specificity was 66 percent to 100 percent. The evidence table for this report (Appendix C) includes diagnostic test accuracy outcomes for more than 25 additional parameters.

Optic disc photography

We included two studies of the diagnostic accuracy of optic disc photography.[32 69] Danesh-Meyer (2006) included participants with open-angle glaucoma (n=42), as well as glaucoma suspects (n=23) and healthy volunteers (n=45).[32] Optic disc photographs were taken using the Canon CF60U camera, 30°setting, with Kodak Ektachrome EPR 150 film and graded by two investigators. The Disc Damage Likelhood Score (DDLS) was determined by two investigators using a Nikon 60 diopter fundus lens with a slit-lamp. The AUC (glaucoma and borderline disease versus normal) was 0.84 (95 percent CI, 0.74 to 0.92) for the cup to disc ratio (C/D ratio) and 0.95 (95 percent CI, 0.80 to 0.98) for DDLS suggesting that the DDLS is a more effective means of discriminating persons with and without disease.

RNFL photography

Two studies examined the accuracy of Retinal Nerve Fiber Layer (RNFL) photography.[59 70] Hong (2007b) analyzed RNFL photographs of 72 glaucoma and 48 healthy participants taken with the Heidelberg Retina Angiograph 1 (HRA1).[59] Two investigators, masked to the participant's diagnosis, reviewed the photographs with a third investigator serving as an adjudicator to resolve any disagreements. The RNFL defect score II, with a AUC of 0.75 ($p < 0.001$), was the best parameter for discriminating early glaucoma and healthy eyes (sensitivity 58.3 percent and specificity 95.8 percent).

Medeiros (2004a) compared RNFL photography to the GDx-VCC in 42 participants with open-angle glaucoma, 32 open-angle glaucoma suspects, and 40 healthy volunteers.[70] One eye of each participant was photographed using the Topcon TRC-50VT camera and Kodak Kodalith high contrast film (red-free filter). Two investigators used a set of 25 reference photos to score photographs with a third investigator adjudicating disagreements. The sensitivities of the global RNFL score were 36 percent and 81 percent respectively for fixed specificities of 95 percent and 80 percent. At a fixed specificity of 95 percent, the sensitivity of the Nerve Fiber Indicator (NFI) was 71 percent versus the 36 percent reported above for red free photos. Overall the global RNFL score determined from red free photos did not perform as well as scanning laser polarimetry. The area under the ROC curve was 0.91 for tht GDx-VCC (NFI) versus 0.84 for the global RNFL score.

Scanning laser polarimetry - GDx

Twenty four studies included an investigation of scanning laser polarimetry (SLP) with variable corneal compensation (GDx VCC).[44 17 69 19 71 57 21 72 39 73 25 58 74 59 28 60 29 63 75 76 77 30 70 31] Discriminating early disease from no disease was the aim of seven studies.[44 17 19 72 39 59 75] In the studies that focused on early OAG, the range of sensitivity across all comparisons and cut offs for the most frequently reported parameter Temporal, Superior, Nasal, Inferior, Temporal

(TSNIT) average was 29.8 percent to 81.63 percent. Specificity was fixed at 80 percent, 90 percent, or 95 percent in three studies, and the lowest reported specificity was 66.36 percent. The range in sensitivity for the nerve fiber indicator (NFI) parameter across all comparisons and cut offs was 28.3 percent to 93.3 percent. Specificities ranged from 52.9 percent to fixed at 80 percent, 90 percent, and 95 percent.

Three studies examined the GDx with enhanced corneal compensation (GDx ECC).[73 58 74] The sample sizes of the included studies ranged from 63 to 92 glaucoma participants and 41 to 95 healthy volunteers. Medeiros (2007) compared the AUCs for GDx VCC and GDx ECC and reported that GDx ECC performed significantly better than GDx VCC for the parameters TSNIT average, Superior average, and Inferior average ($p =<0.01$).[73] Sehi (2007) [58] and Mai (2007) [74] concurred with Medeiros (2007) that imaging with ECC appears to improve the ability to diagnose OAG.

Tests of optic nerve function

FDT 24-2 Perimetry

Five studies examined the diagnostic accuracy of FDT 24-2 threshold test using the Humphrey Matrix Perimeter.[63 78-81] All studies included participants with known glaucoma and healthy volunteers. The range of sample size was 25 to 174 glaucomatous eyes and 15 to 164 healthy eyes. Sensitivities and specificities were reported for the parameters mean deviation, pattern standard deviation and glaucoma hemifield test (GHT) outside of normal limits. There was appreciable heterogeneity in the estimates of sensitivity at 80 percent, 90 percent and 95 percent specificity that may be attributed to a number of factors including different patient populations and variations in cut off points. The sensitivity for the mean deviation was 55 percent and 94 percent at 80 percent fixed specificity. [79 80] Tafreshi (2009) and Leeprechanon (2007) reported 39 percent and 87 percent at 90 percent fixed specificity, and 32 percent and 82 percent at fixed 95 percent specificity respectively.[78 80] Sensitivity and specificity for PSD and GHT are reported with respective cut off points in the evidence tables in Appendix C.

Bagga (2006) [63] and Burgansky-Eliash (2007) [81] reported AUC for the mean deviation (0.69 for both studies with $p < 0.04$ and 95 percent CI, 0.564 to 0.815 respectively). The AUCs for PSD were 0.66 ($p = 0.09$) [63] and 0.733 (95 percent CI, 0.618 to 0.848).[81]

FDT 30-2 Perimetry

Two studies discuss the detection of early glaucoma using the FDT 30-2 threshold test with the Humphrey Matrix Perimeter. [59 82] Both Hong (2007a) [82] and Hong (2007b) [59] enrolled open-angle glaucoma participants with early visual field loss and healthy controls. The MD and PSD were judged to be good parameters for distinguishing between eyes with early disease and eyes with no known defects. The MD was 0.795 and 0.750 and the PSD 0.808 and 0.934 for Hong (2007a) and Hong (2007b) respectively. Both study groups, however, determined that the best parameter for distinguishing eyes with early glaucoma from healthy eyes was the number of points that have $p < 5$ percent in the pattern deviation plot with AUCs of 0.985 (95 percent CI, 0.943 to 0.998) and 0.990 ($p < 0.001$) in Hong (2007a) and Hong (2007b) respectively.

FDT C-20 Perimetry

Three studies discussed the accuracy of FDT C-20 perimetry.[17 83 84] Pueyo (2009) enrolled 130 participants with OHT and 48 healthy volunteers.[17] With a cut off of a cluster of at least four

points with a sensitivity outside 95 percent normal limits, or three points outside 98 percent, or at least one point outside 99 percent, the sensitivity of FDT was determined to be 31.25 percent and specificity 72.9 percent among the subset of 32 participants with glaucomatous optic neuropathy (of the 130 with OHT). The investigators concluded that FDT may not be an ideal test for participants with early defects.

Salim (2009) enrolled 35 participants with known OAG and 35 age- and sex- matched controls with no evidence of glaucoma. FDT, non-contact tonometry, and a questionnaire were used individually and in all possible combinations to determine the accuracy of single and combination tests.[83] Sensitivity of FDT was 58.1 percent and specificity 98.6 percent. Overall, FDT was determined to be the best among the candidate single and combination tests in the study despite fair sensitivity for detecting OAG.

Pierre-Filho (2006) enrolled glaucoma patients who had never experienced perimetry prior to the study.[84] The investigators reported that 21 (32.8 percent) of the 64 participants with glaucoma were identified as having early disease but data are not provided for this subgroup. Sensitivity and specificity were 85.9 percent and 73.6 percent for the presence of at least one abnormal location and 82.8 percent and 83 percent respectively for two or more abnormal locations regardless of severity.

FDT N30 Perimetry

Four studies examined the accuracy of the FDT N30 threshold test.[19 79 85 86] Zeppieri (2010) focused on the detection of early glaucoma among a sample of 75 participants with OAG, 87 with OHT, 67 with GON and 90 healthy volunteers.[19] At the best cut off of < -0.78, the sensitivity of the mean deviation (MD) parameter was 61.3 percent and the specificity was 73.7 percent for distinguishing early OAG from healthy eyes. At the best cut off of > 3.89, the sensitivity of the patterm standard deviation (PSD) was 76.0 percent and the specificity was 87.8 percent. The investigators concluded that "FDT can potentially detect eyes with very early functional defects that do not show structural changes in patients at risk of developing glaucoma." Salvetat (2010) focused on the detection of early disease among a sample of 52 participants with early open-angle glaucoma and 53 healthy volunteers.[86] The sensitivity of MD for distinguishing early OAG from healthy eyes at the best cut off (< -1.12) was 67 percent and the specificity was 74 percent. At the best cut off of > 3.97, the sensitivity of the parameter PSD was 96 percent and the specificity was 85 percent.

Humphrey [Visual] Field Analyzer (HFA)

Eleven included studies examined the diagnostic accuracy of the HFA. Of these, six examined HFA Short Wavelength Automated Perimetry (SWAP),[17 43 63 78 85 87] two tested HFA Standard Automated Perimetry (SAP)-SITA and HFA SAP-Full Threshold (FT), [78 85] four examined HFA-SITA-Standard, [32 77 81 84] and one tested the HFA SITA-Fast protocol.[84] The HFA SWAP testing protocol (the most frequently reported) included 25 to 286 participants with glaucoma and 22 to 289 healthy volunteers across the six included studies. Sensitivity across all comparisons and cut offs for the mean deviation was 25.9 percent to 83 percent, specificity was 80 percent to 95.2 percent. Cut offs ranged from -5.42 to -11.06 dB.

Goldmann Applanation Tonometry (GAT)

Baggam (2006) compared the ability of various tests of structure and function to discriminate eyes with known glaucomatous optic neuropathy (25 eyes) from healthy eyes (22 eyes).[63] The AUC for IOP as measured by GAT was 0.66 ($p = 0.05$).

Non-Contact Tonometry (NCT)

Salim (2009) included non-contact tonometry individually and in all possible combinations with other measures of structure and function to determine the accuracy of single and combination tests.[83] Intraocular pressure, as measured by NCT, was found not to be a very sensitive test for detecting glaucoma (sensitivity 22.1 percent). The investigators acknowledge that use of topical medications by the glaucoma participants could limit the ability to identify those with disease.

Key Question 4

We did not identify any studies that addressed whether participation in an open-angle glaucoma screening based program leads to reductions in intraocular pressure when compared to no screening or other screening-based program.

Key Question 5

Evidence from Systematic Reviews

Hatt (2006) undertook a systematic review of screening modalities for open-angle glaucoma compared to no screening (including opportunistic case finding and referral). There were no restrictions on included populations.[88] The primary outcome of interest was the prevalence of visual field loss defined as the proportion of participants with a prespecified severity of visual field loss diagnosed by either manual or automated field assessment. Other primary outcomes included the prevalence of optic nerve damage and visual impairment. The focus was on identifying randomized controlled trials (RCTs) as these are the optimal study design for addressing lead time bias and length bias. Electronic searches of five databases including MEDLINE and CENTRAL were conducted in 2006 and again in January 2009, but none of the studies that were identified were eligible for inclusion. The review authors acknowledge that RCTs require lengthy follow-up and are predicated on identifying appropriate candidate tests that may be incorporated into a screening-based program.

Detailed Analysis of Primary Studies

We did not identify any studies that addressed whether participation in an open-angle glaucoma screening based program leads to reductions in visual field loss or optic nerve damage when compared to no screening or another screening-based program.

Key Question 6

We did not identify any studies addressing the harms associated with screening for open-angle glaucoma.

Introduction

The Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program requested a comparative effectiveness review of glaucoma screening. The topic was selected through the Effective Health Care Program nomination process and a review of the scientific medical literature.

Background

Glaucoma is a leading cause of visual impairment and blindness and is estimated to affect 60.5 million people worldwide by 2010.[2 3] Glaucoma may be classified by optic nerve damage, visual field loss and elevated intraocular pressure, but these mechanisms need not occur in tandem to confirm diagnosis.[4] Damage is irreversible, so early detection may prevent severe vision loss. Open-angle glaucoma (OAG), the most common subtype of the disease, impacts over 2.5 million people in the United States, with a median age-adjusted prevalence of 4.6 percent, and 1.6 percent, among black and white people respectively (based on year 2000 estimates).[5]

Unfortunately, it has been shown that only half of the prevalent cases of glaucoma have been identified in the United States (U.S.).[5] This is due to at least two factors. First, glaucoma is an asymptomatic disease that patients doesn't notice until the onset of advanced disease, accompanied by severe vision loss. Second, there's no single test to identify people with glaucoma, which has severely hampered the establishment of screening-based programs to detect the disease.

However, since the last evaluations of glaucoma screening in the mid 2000s, there have been significant advances in the devices used to test optic nerve structure and function.[6 7] Because of this new evidence, we believe a re-evaluation of the safety and efficacy of population-based glaucoma screening is warranted.

In the most recent review of glaucoma screening--the U.S. Preventive Services Task Force report on the safety and efficacy of screening for chronic OAG, March 2005--ocular examinations such as tonometry, perimetry, and direct ophthalmoscopy were performed alone or as a part of a multicomponent screening test for glaucoma. The results of that report stated that intraocular pressure measurement and optic nerve head assessment alone have limited effectiveness as population-based screening tools.[6 7] The U.S. Preventive Services Task Force also concluded that current methods used to assess visual field loss may be impractical for population-based screening due to the length of time required for testing and the challenge of equipment portability. Since that report, researchers have published studies on new diagnostic tests, such as frequency-doubling technology, used to assess visual field loss.[7] This report includes new evidence regarding the diagnostic accuracy of frequency-doubling technology as well as data on other screening examinations used alone or in combination to aid in the diagnosis of OAG.

Purpose for Evidence Report

The objective of this review was to summarize the evidence regarding the safety and efficacy of screening-based programs for OAG with a specific focus on the effects of screening on visual impairment, patient reported outcomes, intraocular pressure, visual field loss, optic nerve damage, and adverse effects. This review also includes a summary of the diagnostic accuracy of screening examinations and tests for OAG.

Key Questions

Screening for a medical condition in asymptomatic individuals may be considered to be beneficial when 1) the condition has a significant individual or population burden; 2) the condition is associated with adverse impacts on the health of the individual; 3) there's an accurate test that detects the condition during its asymptomatic or early clinical stage; 4) treatment of the condition at an asymptomatic or early stage is significantly more effective at improving important health outcomes than treatment once it's symptomatic; and 5) the potential harm to the individual due to screening and early intervention is limited. Following these requirements, we considered and compared, where possible, the safety and efficacy of screening-based programs for OAG as a tool for preventing or greatly reducing loss of sight due to the disease.

Key Question 1a: Does a screening-based program for OAG lead to less visual impairment when compared to no screening program?

Key Question 1b: How does visual impairment vary when comparing different screening-based programs for OAG?

Key Question 2a: Does a screening-based program for OAG lead to improvements in patient-reported outcomes when compared to no screening?

Key Question 2b: How do patient-reported outcomes vary when comparing different screening-based programs for OAG?

Key Question 3: What is the predictive value of screening tests for OAG?

Key Question 4a: Does a screening-based program for OAG lead to reductions in intraocular pressure when compared to no screening program?

Key Question 4b: How does intraocular pressure vary when comparing different screening-based programs for OAG?

Key question 5a: Does a screening-based program lead to a slowing of the progression of optic nerve damage and visual field loss when compared to no screening program?

Key Question 5b: How do optic nerve damage and visual field loss vary when comparing different screening-based programs for OAG?

Key Question 6: What are the harms associated with screening for OAG?

Methods

Topic Development

AHRQ requested that the EPC assist with the formulation and refinement of the Comparative Effectiveness Review (CER) topic "Effectiveness of screening and treatment for glaucoma."

In consultation with AHRQ, the EPC investigators identified a small group of stakeholders to serve as members of a Key Informant Group. The Key Informant Group helped shape the Key questions (KQs) relevant to the topic by providing input regarding the populations and clinical subgroups; interventions; and outcomes of interest to clinicians, policy makers, payers, and consumers.

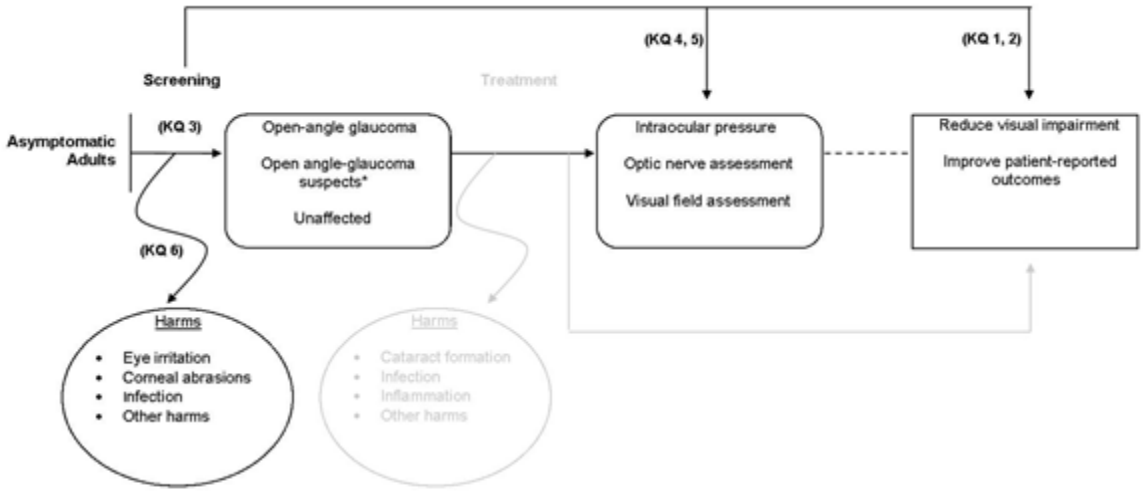
The EPC investigators incorporated the feedback of the Key Informants into a draft of the Key Questions, analytic framework, and inclusion criteria. And then they posted the draft on the AHRQ Web site for public comment from April 22, to May 20, 2010. The investigators finalized the inclusion criteria after considering these public comments.

A Technical Expert Panel (TEP) was selected to provide broad expertise and perspectives specific to the topic. The TEP reviewed the proposed methodological approach for completing the comparative effectiveness review and provided information to the EPC to aid in the refinement of the inclusion criteria and literature search strategies. The final protocol entitled *Comparative Effectiveness of Screening For Open-Angle Glaucoma* was posted to the AHRQ Web site on November 16, 2010.

Analytic Framework

The analytic framework (Figure 1) is a modified version of a larger framework depicting the impact of both screening and treatment for OAG. This figure focuses on the screening aspects of the framework. It depicts the KQs within the context of the inclusion criteria described in the following sections. In general, the figure illustrates how screening-based programs (which may incorporate treatment when indicated) may reduce visual impairment (KQ1) and/or improve patient reported outcomes (KQ2), reduce intraocular pressure (KQ4) and possibly slow the progression of optic nerve damage and/or visual field loss (KQ5). The figure also illustrates the potential predictive value of screening-based programs to detect OAG and OAG suspects (KQ3). Finally, the potential harms of screening (KQ6) are illustrated in the framework.

Figure 1. Analytic Framework for Screening for Open-Angle Glaucoma



Study Selection

Types of studies

We included randomized controlled trials (RCTs), quasi randomized controlled trials, and observational study designs, including cohort and case control studies, for Key Questions 1 through 6. For KQ3 we also included cross-sectional studies, study designs in which all tests (including the index, comparator, and reference standard) were performed on all participants, and designs in which participants were randomized to one test (among the index and potential comparator(s)) but all were evaluated with the reference standard.[1]

We also considered and sought systematic reviews that addressed the KQs.

We excluded studies that addressed the following:

- Studies of the prevalence of glaucoma in a specific population unless they also include tests of diagnostic accuracy
- Studies of disease progression that did not include participants previously screened for glaucoma
- Studies of risk factors for disease

Types of participants

We included studies of adult (as defined by included studies) asymptomatic participants in general or high-risk populations. Asymptomatic general populations included populations not previously tested, those not diagnosed with glaucoma, and those not presenting with symptoms known to be related to a diagnosis of glaucoma. Asymptomatic high-risk populations included those not previously tested, diagnosed or presenting with symptoms known to be related to

glaucoma; those with a family history of glaucoma; those from specific racial/ethnic groups; those with specific ocular or other medical conditions as defined by included studies (e.g., diabetes); and those over a certain age.

We also included studies of suspected OAG subpopulations, which included participants identified from prior testing as possibly having glaucoma or as having a risk factor for glaucoma, e.g., high inner ocular pressure, but with an unconfirmed diagnosis. We excluded studies of participants with known glaucoma at the time of screening (KQ 1, 2, 4 and 5) and those that included the healthy eye of a participant with known glaucoma (KQ 3). We considered studies that enrolled healthy volunteers in addition to those with suspected OAG but excluded studies of healthy volunteers only.

Interventions

We included studies of the following screening tests conducted alone or in any possible combination (including multi-component simultaneous or sequential testing):

- Tonometry (contact and non-contact tonometry)
- Perimetry (including short-wavelength, high-pass, motion, flicker perimetry, yellow and blue perimetry)
- Direct and indirect ophthalmoscopy
- Fundus photography or computerized imaging of the posterior pole, optic disc or retinal nerve (includes optical coherence tomography, retinal tomography, scanning laser polarimetry)
- Pachymetry (corneal thickness measurement) when used in conjunction with another test to diagnose glaucoma (We excluded studies where pachymetry is used alone.)

We excluded studies of the following screening tests that are either not commercially available for screening or are not commonly used in the diagnosis of glaucoma:

- Electroretinography
- Contrast sensitivity and visual acuity
- Color tests
- Water drinking tests
- Optical coherence tomography (OCT) 1 and OCT 2 imaging
- Heidelberg Retina Tomograph 1 (confocal scanning laser ophthalmoscope)
  Glaucoma Diagnosis, early version (scanning laser polarimeter)

We also excluded studies that examined technical aspects of included devices, e.g., usability, technician-training.

Comparators/Reference standards

KQ1, 2, 4, 5, and 6 explore comparisons of the interventions mentioned above (conducted alone or in any possible combination as a part of a screening-based program) to no screening program (including usual care) and to different screening-based programs (abovementioned screening tests conducted alone or in any possible combination). KQ3 explores comparisons of screening/diagnostic tests to the gold/reference standards of confirmed OAG at the time of follow-up or OAG requiring treatment (diagnosed by an ophthalmologist using objective assessments). The diagnosis should have included a clinical examination with measurement of intraocular pressure, assessment of the visual field, assessment of the optic nerve head and or retinal nerve fiber layer, or review of fundus photographs. We considered other methods to

confirm diagnosis as defined by included studies if the examinations/tests were specified in the manuscript. We acknowledge that there is no consensus on the gold standard test (or combination of tests) for the identification of patients with OAG. We adapted the reference standards for KQ3 from a diagnostic test accuracy review conducted by Burr (2007).[1]

Outcomes

Key Question 1

Primary outcome

We looked at the proportion of participants with moderate, severe, and profound visual impairment (as defined in the International Classification of Diseases, Clinical Modification, 9th Revision[8]). We also considered other measurements of visual impairment as defined by included studies.

Secondary outcome

We also looked at visual acuity outcomes (e.g., mean visual acuity or proportion of participants in pre-specified visual acuity categories) as reported in studies with Snellen, or any other valid chart that yields scores that can be converted to Snellen fractions or Logarithm of the Minimum Angle of Resolution values.

Key Question 2

We looked at the participants' mean total or relevant item/subscale scores as measured by any validated questionnaire, e.g., National Eye Institute Visual Functioning Questionnaire, for the following patient-reported outcomes among the treatment groups of interest:

- Vision-related quality of life (vision-related functional loss as well as the impact of functional loss on activities of daily living) (Primary outcome)
- Patient satisfaction (Secondary outcome)

Key Question 3

We looked at the number of participants in the following categories: true positives, true negatives, false positives, and false negatives. We also considered studies that include sensitivity and specificity values only.

Key Question 4

We looked at the difference in the mean intraocular pressure among the groups of interest.

Key Question 5

We looked at a) the proportion of participants with progressive optic nerve damage as defined by included studies and as observed via fundus photography or other imaging of the posterior pole and b) the proportion of participants with progression of visual field loss as defined by included studies.

Key Question 6

We looked at the proportion of participants experiencing the following adverse events (adapted from the U.S. Preventive Services Task Force -- http://www.ahrq.gov/clinic/uspstf05/glaucoma/glaucrs.htm):

- Eye irritation
- Corneal abrasions
- Infection (e.g., endophthalmitis)
- Distortion of sense of taste (due to anesthetic use)
- Exam apprehension
- Psychological effects related to a glaucoma diagnosis or misdiagnosis
- Harms related to overdiagnosis

We also planned to include other harms as reported in included studies

Timing of outcome

We assessed outcomes for Key Questions 1, 2, 4, and 5 at one year of follow-up and at annual intervals thereafter. There was no minimum length of follow-up for outcomes related to Key Questions 3 and 6.

Setting

Settings for this review included community screenings, non-eye care health provider settings, eye care provider clinical settings (ophthalmologists and optometrists), and telemedicine.

Search Strategy

We searched the following databases for primary studies: MEDLINE, EMBASE, LILACS (Latin American and Caribbean Literature on Health Sciences), and CENTRAL (the Cochrane Central Register of Controlled Trials). We developed a search strategy for MEDLINE, accessed via PubMed, based on an analysis of the medical subject headings, (MeSH) terms, and text words of key articles identified a priori. We adapted this search strategy for searches of EMBASE (using EMTREE terms) and CENTRAL (Appendix A). We searched the literature without imposed language, sample size or date restrictions, but excluded non-English language studies at the time of full text review. We searched relevant systematic reviews to identify any additional inclusions. We searched from the beginning of each database through 29 October 2010.

We also conducted a search in MEDLINE and CENTRAL for systematic reviews that address the key questions of interest. The search included the topic strategy as noted in the Appendix combined with the term "AND systematic[sb]" and was limited to systematic reviews published from 2009 to 2011. We searched MEDION (www.mediondatabase.nl) for related diagnostic accuracy reviews (KQ3). The search for systematic reviews was conducted on March 2, 2011.

We screened an existing database of eye and vision systematic reviews prepared by Li (2010) to identify relevant OAG systematic reviews published prior to 2009.[9]

Abstract Screening

We screened potentially relevant citations (primary studies and systematic reviews) using the Web-based systematic review software, DistillerSR (http://systematic-review.net/). All citations identified by the search strategies were uploaded to DistillerSR. Two reviewers independently assessed titles and abstracts resulting from the literature searches according to the inclusion criteria. We classified the titles and abstracts as 'include', 'exclude' or 'unsure'. We resolved disagreements about eligibility through discussion among reviewers. A copy of the abstract screening form is included in Appendix B.

Full-Text Screening

Two reviewers independently applied the same inclusion criteria as used during abstract screening. Citations tagged as 'unsure' by both reviewers, 'unsure' by one reviewer and 'include' by the other, or 'include' by both reviewers, were promoted to full-text screening. They also removed non-English language articles from further consideration at this stage. We resolved any disagreements regarding inclusion through discussion between reviewers, or, as needed, during a team meeting. A copy of the full-text screening form is included in Appendix B.

Data Abstraction

Data abstraction forms were designed and pilot tested. One reviewer extracted descriptions of the study, including details about the population, intervention(s) and outcomes of interest, using the systematic review software, DistillerSR. A second reviewer verified the data. We resolved disagreements through discussion. A copy of the data abstraction forms is included in Appendix B.

Risk of Bias Assessment

We used the Cochrane Collaboration's tool for assessing the risk of bias of randomized and quasi-randomized trials. Two reviewers assessed the included studies for sources of systematic bias according to the guidelines in Chapter 8 of the Cochrane Handbook for Systematic Reviews of Interventions and evaluated the studies for the following criteria: sequence generation and allocation concealment (selection bias), masking of participants, study investigators, and outcome assessors (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias), and other sources of bias.[10] Masking of investigators and participants might not be possible with some of the interventions being examined, but was noted when mentioned. We reported judgments for each criterion as 'Low risk of bias', 'High risk of bias' or 'Unclear risk of bias (information is insufficient to assess)'. The two reviewers resolved disagreements through discussion.

Two reviewers assessed the methodological rigor of observational studies using a modified version of the Newcastle Ottawa Scale.[11] The Newcastle Ottawa Scale includes domains to assess the quality of study group selection (representativeness, selection, case definitions); comparability of cohorts/cases and controls on the basis of the design or analysis; and ascertainment of exposure(s) or outcome(s) adequacy of follow-up, non-response rate and financial or other conflicts of interest. Each item query required a yes, no, or unable to determine/not reported response. Additionally, reviewers provided an overall assessment of the quality of each study as "good" "fair" or "poor" using the reporting bias, selection bias, and confounding domains as a basis for the assessment.

For Key Question 3, we used the QUADAS checklist, which is specific to risk of bias assessment for diagnostic accuracy studies.[12] The QUADAS tool includes 14 items that evaluate

numerous domains including representativeness, inclusion/exclusion criteria, choice of reference standard, masked interpretation of results of tests and reference standard, and study withdrawal. We reported judgments for each checklist item as 'Yes', 'No', or 'Unclear.'

We used a tool adapted by Li (2010) from the Critical Appraisal Skills Program, Assessment of Multiple Systematic Reviews, and the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement to assess the methodological quality of systematic reviews.[9] We used the following criteria, adapted from Li, to determine which were of sufficient quality to be considered for inclusion in this review: comprehensive search for primary studies (searches of more than one bibliographic database); risk of bias assessment; and appropriate analysis methods (no pooled arm analysis).

Rating Body of Evidence

We assessed the quantity, quality and consistency of the body of available evidence addressing KQ1 through KQ6. We used an evidence grading scheme recommended by the Grading of Recommendation Assessment, Development and Evaluation Working Group, adapted by AHRQ in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews (http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=328) and recently published in the Journal of Clinical Epidemiology.[13,14]

We considered the strength of the study designs with randomized controlled trials as the highest level of evidence, followed by observational studies. If an outcome was evaluated by at least one randomized controlled trial, as well as observational studies, our evidence grade was based on the randomized controlled trials and was followed by the quality of the cohort studies. If an outcome was evaluated by one or no randomized controlled trials, our evidence grade was based on the single randomized controlled trial in addition to the best available observational study.

We assessed the quality and consistency of the best available evidence, including assessment of the risk of bias in relevant studies, as well as aspects of consistency, directness, and precision as described in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews and by Owens (2010).[14] For each outcome of interest, two reviewers graded the major outcomes for each Key Question and then the entire team discussed their recommendations and reached consensus.

Data Synthesis

When we identified existing high-quality systematic reviews that addressed the Key questions, we cited these reviews as evidence and did not abstract and synthesize data from primary studies. For interventions (diagnostic tests), comparisons, and outcomes that were not identified in systematic reviews, we abstracted evidence from primary studies, including those that had been published or identified after the date of last search conducted for the systematic review. We followed the recommendations of Whitlock (2008) for incorporating systematic reviews in complex reviews and provided a narrative summary of the review methods (i.e., inclusion/exclusion criteria, search strategy, statistical methodology) and findings (i.e., number of studies included, quantitative and qualitative results). Similarly, in the instance of multiple reviews, we evaluated the consistency across reviews addressing the same key question.[15]

Results

The electronic search of MEDLINE identified 64 systematic review titles and abstracts. The Li 2010 database included 105 systematic review titles and abstracts. We excluded 167 of the 169 titles and abstracts for the following reasons: did not address any of the key questions, narrative summary only, could not retrieve full text to assess, similar inclusion criteria but date of search for studies older than another included systematic review, and duplicate reference of included systematic review. We identified two systematic reviews for inclusion [need to add references here]. One systematic review (Burr 2007) addressed the diagnostic test accuracy of candidate screening tests for the detection of open-angle glaucoma (Key Question 3) and the second review (Hatt 2006) addressed the question of whether screening-based programs prevent optic nerve damage due to open-angle glaucoma when compared to no screening (Key Question 5) (Evidence Tables 1 and 2 in Appendix C).

The electronic searches that we conducted for concurrent comparative effectiveness reviews of screening and treatment for OAG, identified a total of 4680 primary study titles and abstracts. After de-duplicating the database and removing conference abstracts and book chapters (N = 1083), we reviewed 3,597 titles and abstracts. We retrieved the full text of 630 articles and assessed the studies for inclusion in the review. We included 72 primary studies addressing the diagnostic accuracy of candidate screening tests for the detection of OAG that were not included in the Burr 2007 systematic review (Key Question 3 - Evidence Tables 3 to 7 in Appendix C). (See Figure 2). We did not identify any primary studies eligible for inclusion for any other key question. A listing of the 558 excluded studies, with reason(s) for exclusion, is included in Appendix D.

A listing of devices from the primary studies is included in Appendix F. In summary the following number of diagnostic studies included the devices summarized in this comparative effectiveness review:

Tests of optic nerve structure

- Heidelberg Retina Tomograph (HRT) II (17 studies)
- HRT III (11 studies)
- Optical coherence tomography (OCT) (38 studies)
- Optic disc photography (2 studies)
- Retinal nerve fiber layer photography (2 studies)
- Scanning laser polarimetry (Glaucoma diagnosis – GDx device) (24 studies)

Tests of optic nerve function

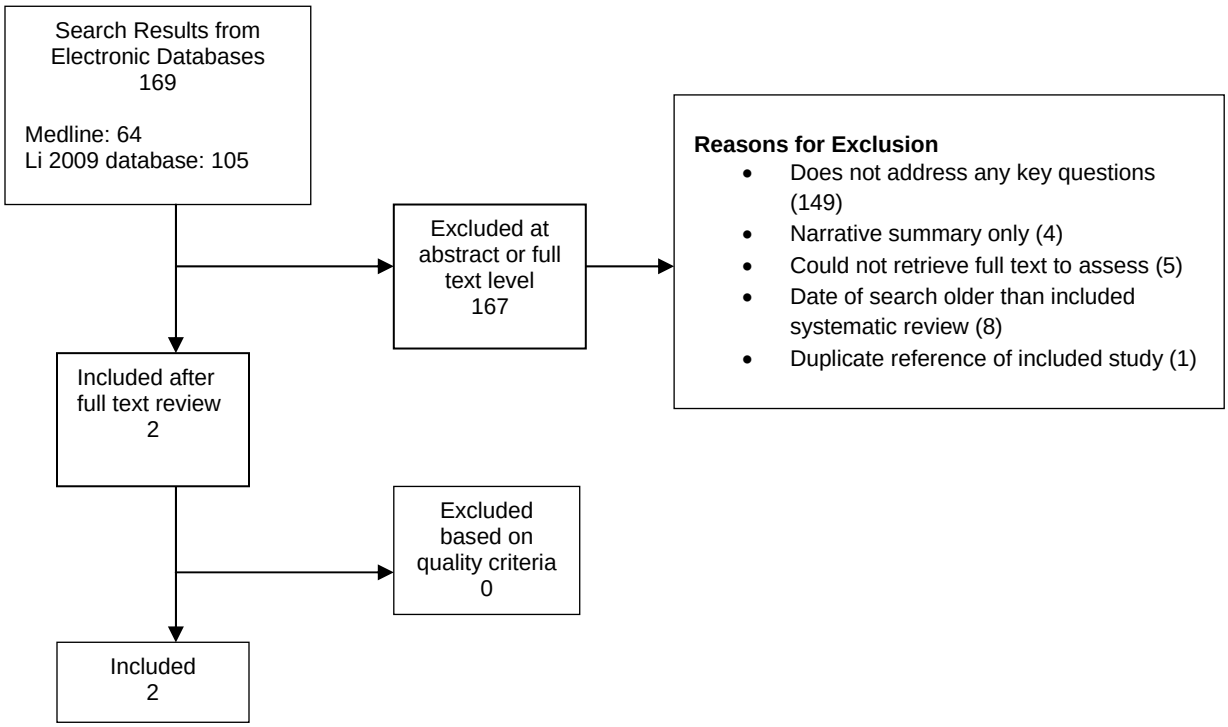
- Frequency doubling technology (FDT) 24-2 perimetry (5 studies)
- FDT 30-2 (2 studies)
- FDT C-20 (3 studies)
- FDT N30 (4 studies)
- Humphrey Visual Field Analyzer (HFA) (11 studies)
- Goldmann applanation tonometry (1 study)
- Non contact tonometry (1 study),
- Octopus 301 perimeter (1 study)

Because there was appreciable variability in devices, parameters, thresholds, and measurement of outcomes reported in the primary studies of interest, we did not combine the

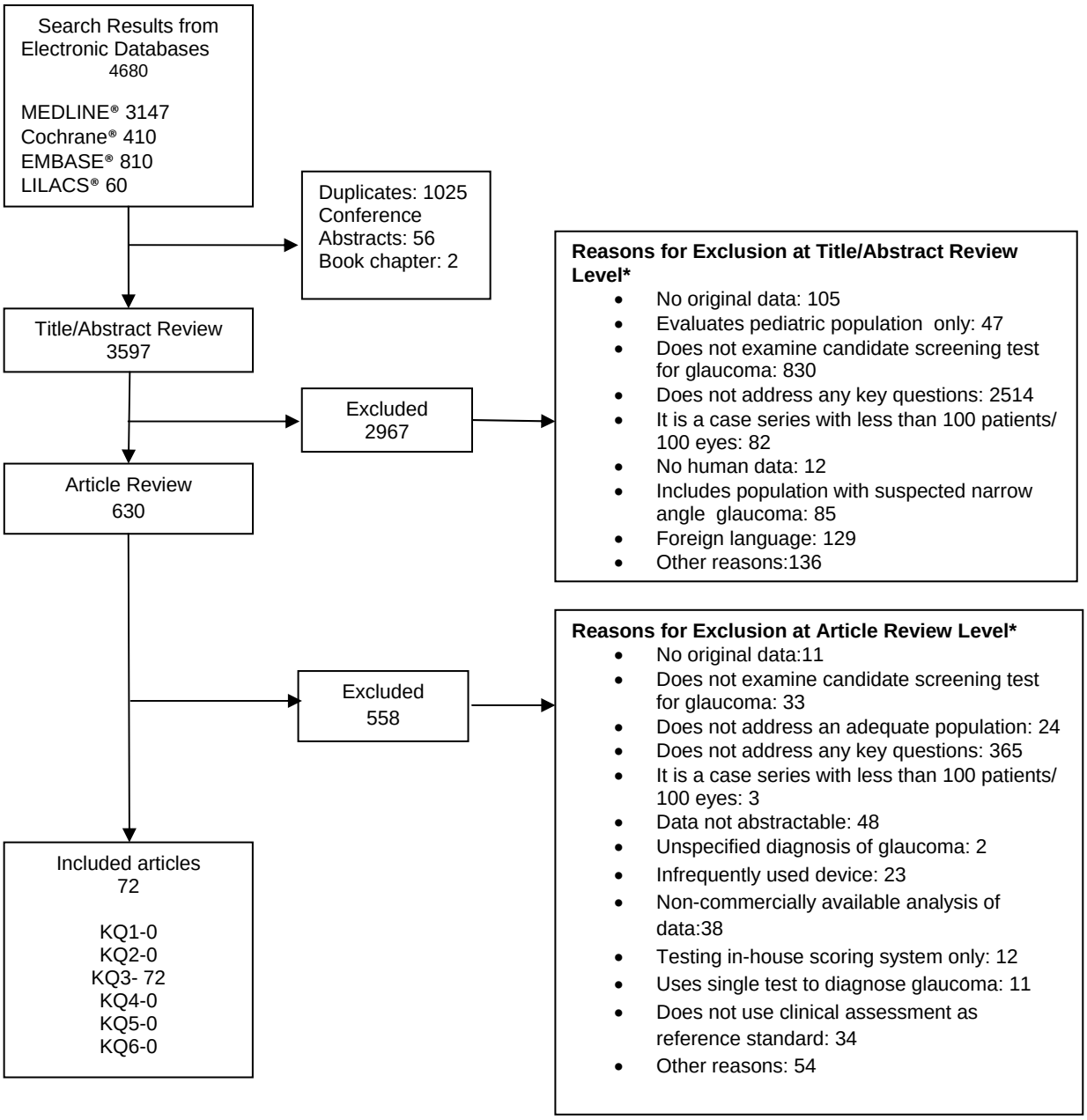
results using meta-analysis and instead presented a narrative summary. A meta analysis of the included studies would shift the focus of discussion to a limited number of relevant device parameters and common thresholds (cut offs) that were commonly reported for any one device. As we are unable to determine which parameters are most important for identifying persons with OAG, and as our reported results would have been limited to a few parameters in a subset of studies, we chose to include in Evidence Tables 5 to 7 and discuss as appropriate the full complement of device parameters and thresholds as reported in the included studies. We summarize, where possible, the magnitude of validity across all parameters of interest for devices considered in this report.

Figure 1. Summary of the Literature Search

A. Systematic Reviews Literature Search



B. Primary Studies Literature search



* Total may exceed number in corresponding box, as articles excluded by two reviewers at this level.

Key Question 1

We did not identify any studies that addressed whether participation in an OAG screening-based program leads to less visual impairment when compared to another screening-based program or no screening.

Key Question 2

We did not identify any studies that addressed whether participation in an OAG screening-based program leads to improvements in patient-reported outcomes when compared to another screening-based program or no screening.

Key Question 3

Summary

SAP was compared with several candidate tests at the time of a systematic review by Burr 2007.[1] Results indicated that the sensitivity of SAP was higher than Goldmann tonometry, similar to Heidelberg retina tomography (HRT), and lower than disc photos or frequency doubling technology (FDT). Results also indicated that the specificity of SAP was higher than disc photos and FDT, similar to HRT, and lower than Goldmann tonometry.

Despite improvements in technology, including newer imaging and functional technologies, it's still not clear whether any one test or combination of tests is suitable for use in glaucoma screening.

The lack of a definitive diagnostic reference standard for glaucoma and the need for more heterogeneity in the design and conduct of diagnostic test accuracy studies, prevents a coherent synthesis of data and therefore limits conclusive statements regarding these tests.

Evidence from Systematic Reviews

Burr (2007) conducted a diagnostic test accuracy review of candidate screening tests for OAG.[1] Study inclusion criteria and pooled outcomes of sensitivity, specificity, and diagnostic odds ratios are listed in the evidence table (Evidence Tables 1 and 2 in Appendix C). In summary, the investigators included 40 studies of more than 48,000 participants 40 years of age and older and those at high risk for the development of OAG based on demographic characteristics or comorbidities. The focus was on studies of participants likely to be encountered in a routine screening setting. Tests of optic nerve structure, optic nerve function, and intraocular pressure were included and compared to other individual or combination tests. The primary reference standard was confirmation of OAG at follow-up. Also considered was diagnosis of OAG requiring treatment. Outcomes were measures of the test related to sensitivity, specificity, potential harm, acceptability and reliability. There was significant statistical heterogeneity among the included studies for the majority of the tests, with the exception of optic disc photography (sensitivity), HRT II (sensitivity and specificity), and FDT C-20-1 (sensitivity). We include a summary of findings for the tests in this review. (A listing of devices is included in Appendix F.)

Tests of optic nerve structure

Heidelberg Retina Tomograph (HRT) II

HRT II was a diagnostic test of interest in three studies, all with a common cut-off value and two of which were judged to be of higher quality than the third. One study specifically recruited high-risk populations (family history of OAG, African or Caribbean descent, aged 50 years or older). Using the common cut-off of one or more results that are borderline or outside normal limits, the pooled sensitivity was 86 percent and the pooled specificity was 89 percent.

Ophthalmoscopy

Burr (2007) included seven studies addressing the diagnostic accuracy of ophthalmoscopy including slit-lamp biomicroscopy (two studies) and direct ophthalmoscopy (five studies).[1] Using a common cut-off of a vertical cup-to-disc ratio greater than or equal to 0.7 (also defined as gradings of "normal" and "suspicious" or other subjective criteria as defined by consultant ophthalmologists), pooled sensitivity and specificity for the five studies with common cut-offs were 60 percent (95 percent CI, 34 to 82 percent) and 94 percent (95 percent CI, 76 to 99 percent), respectively. The diagnostic odds ratio was 25.70 (95 percent CI, 5.79 to 109.50) suggesting a 25-fold higher odds of a positive test among those with glaucoma when compared to those without glaucoma.

Optic disc photography

There were six studies of optic disc photography with five using a common cut-off of between vertical cup-to-disc ratio greater than 0.59 to greater than or equal to 0.7. The range of sensitivity was 65 to 77 percent and the range of specificity was 59 to 98 percent. The authors noted that some photographs were taken with pupils dilated (3/6 studies) while the remaining did not specify whether dilation was used or not.

Retinal Nerve Fiber Layer (RNFL) photography

The common cut-off for the four included studies was diffuse and/or localized defect. Among these studies, two were described as including participants "representative of a screening or diagnostic setting." The pooled diagnostic odds ratio was 23.10 (95 percent CI, 4.41 to 123.50), and the pooled sensitivity and specificity were 75 and 88 percent respectively.

Tests of optic nerve function

FDT (C-20-1) Perimetry

Three studies of FDT (C-20-1) were considered, all of which used the common cut-off of one abnormal test point. The pooled sensitivity and specificity results for this test were high (92 and 94 percent respectively).

FDT (C-20-5) Perimetry

Five studies with statistically significant heterogeneity were included using the common cut-off of one abnormal test point. The range of sensitivity was 7 to 100 percent and the specificity range was 55 to 89 percent.

Oculokinetic Perimetry

Four studies were included for the diagnostic accuracy of oculokinetic perimetry. Three were studies of participants who may be encountered in a screening setting, and one had a lower risk of bias (based on QUADAS domains). The common cut-off varied in description, but is best described as one or more points missing. The odds of a positive test were 57 times higher for those with glaucoma when compared to those without glaucoma. The pooled sensitivity and specificity were 86 and 90 percent respectively.

SAP Suprathreshold Test

Nine studies, including the Baltimore Eye Survey and the Blue Mountains Eye Study, were included in the analysis. Although the sensitivity and specificity were similar among the Baltimore and Blue Mountains studies, there was significant heterogeneity among the included studies. The range in sensitivity was 25 to 90 percent and the range in specificity was 67 to 96 percent.

SAP Threshold Test

Among the five studies analyzed, tests for SAP threshold included Humphrey 30-2, 24-2 threshold, and Octopus 500. The pooled sensitivity was 88 percent and specificity was 80 percent for the common cut-off (the definition of the common cut-off differed by included study, but is defined in Burr (2007)).

Goldmann Applanation Tonometry (GAT)

At the common cut-off of inner ocular pressure greater than 20.5-22 mm Hg, nine studies with significant hetereogeneity reported sensitivity in the range of 10 to 90 percent and specificity in the range of 81 to 99 percent.

Non-contact tonometry

One study with an inappropriate reference standard reported a sensitivity of 92 percent and specificity of 92 percent using a cut-off of inner ocular pressure greater than 21 mmHg.

Direct comparisons of candidate tests

Six studies included comparisons of SAP to optic disc photography, HRT II, SAP, FDT, and/or GAT. Burr 2007 concluded that sensitivity results at the common cut-off revealed that SAP was generally better than GAT, similar or worse than HRT II, and worse than optic disc photography, FDT C-20-5, and FDT C-20.

Burr 2007 also concluded that the specificity results at the common cut off revealed that SAP was generally better than optic disc photography (FDT C-20-5, and FDT C-20), similar or worse than HRT II, and worse than GAT.

Based on analysis of the common cut-offs for each test, accuracy (high-quality studies only-data not shown in this evidence report), combination tests, tests for glaucoma at specific stages, and direct and indirect comparisons of tests, Burr (2007) determined that optic disc photography, HRT II, FDT, SAP and GAT tests may be potentially incorporated into a screening-based program.

Burr (2007) also acknowledged that given the “imprecision in estimates from the pooled meta-analysis models for the diagnostic performance of each test it was not possible to identify a single test (or even a group of tests) as the most accurate.”

Detailed Analysis of Primary Studies

We undertook a search for additional primary studies as described in the Methods section to address the diagnostic accuracy of candidate screening tests and identified 72 studies. Overall, almost all of these studies have potential spectrum bias, since they enrolled healthy volunteers and participants with known disease. For relatively unbiased estimates of sensitivity and specificity, diagnostic accuracy studies should include participants one would expect to encounter in practice, i.e., in a screening or diagnostic test setting. The results, including sensitivity and specificity by parameter, are summarized in the evidence tables (Appendix C). A narrative summary follows with a particular emphasis on studies that identify early disease, and/or examine newer and more frequently reported technologies.

Tests of optic nerve structure

HRT II

Seventeen studies include measures of diagnostic accuracy for HRT II. [16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32] Two of these studies specifically focused on detecting early (or moderate) glaucoma, Naithani (2007) [24] and Uysal (2007) [26], and are discussed in this narrative section. The remaining 15 studies are summarized in the evidence tables included in Appendix C.

Naithani (2007) enrolled 60 participants with glaucoma (30 early defects and 30 moderate visual field defects) and 60 healthy volunteers.[24] Area under the receiver operating characteristic curve (AUC) values were reported in the range of 0.474 (disc area) to 0.852 (vertical cup-to-disc ratio).

Uysal (2007) enrolled 70 participants with early or moderate glaucoma and 70 healthy volunteers.[26] The range of sensitivity across 12 parameters was 47.1 percent (RNFL cross-sectional area) to 74.3 percent (Linear cup/disc area ratio) and the range of specificity was 47.1 percent (mean RNFL thickness) to 71.4 percent (cup shape measure). The investigators concluded that some parameters have better sensitivity and specificity than others. In addition, a subgroup analyses by disc size revealed that is was more difficult to distinguish glaucoma in participants with smaller discs.

HRT III

Eleven studies examined the diagnostic accuracy of HRT III. [33 34 35 36 37 22 38 39 23 40 27]

Reddy (2009) identified 81 participants with early visual field loss (out of 247 participants with glaucoma) and 142 healthy volunteers. Early visual field loss was defined as a mean deviation less than 5dB[35]. The sensitivity of the Glaucoma Probability Score for distinguishing eyes with early field loss from healthy eyes was 67.9 percent and the Moorfields Regression Analysis was 71.9 (at a fixed specificity of 92 percent). The investigators concluded that, "Moorfields Regression Analysis and Glaucoma Probability Score have similar ability to detect glaucomatous changes, and typically agree. The relative ease and sensitivity of the operator-independent Glaucoma Probability Score function of the HRT III may facilitate glaucoma screening."

Badala (2007) compared four imaging methods for their ability to distinguish early glaucoma from healthy eyes.[39] Forty-six eyes from 46 participants with early OAG and 46 eyes from healthy volunteers were enrolled. Sensitivity (parameter: reference height) ranged from 4 to 70

percent (Frederick S. Mikelberg discriminant function and Reinhard O. W. Burk discriminant function) at a fixed specificity of 95 percent.

Optical Coherence Tomography (OCT)

Of the 38 included studies that investigated the diagnostic accuracy of OCT, [41 42 43 44 45 46 17 47 48 49 33 50 51 52 53 54 55 56, 57 39 24 25 58 59 60 28 61 62 29 63 64 65 66 67 30 68 31] 33 considered the Stratus OCT, five included the Cirrus OCT, four considered the RTVue OCT, and one examined the OTI OCT. Across the 33 studies that examined the Stratus OCT, the sample size ranged from 26 to 95 participants with glaucoma and 37 to 128 healthy volunteers. One study also enrolled 130 participants with ocular hypertension. For the parameter average RNFL thickness, the range of sensitivity was 24 to 96 percent, suggesting appreciable heterogeneity among the studies. The range of the specificity was 66 to 100 percent. The evidence table for this report (Appendix C) includes diagnostic test accuracy outcomes for more than 25 additional parameters.

Optic disc photography.

We included two studies of the diagnostic accuracy of optic disc photography.[32 69] Danesh-Meyer (2006) included participants with OAG (n=42), as well as glaucoma suspects (n=23) and healthy volunteers (n=45).[32] Investigators took optic disc photographs using the Canon CF60U camera, 30-degree setting, with Kodak Ektachrome EPR 150 film and graded by two investigators. Two investigators determined the Disc Damage Likelhood Score by using a Nikon 60 diopter fundus lens with a slit-lamp. The AUC (glaucoma and borderline disease versus normal) was 0.84 (95 percent CI, 0.74 to 0.92) for the cup-to-disc ratio and 0.95 (95 percent CI, 0.80 to 0.98) for Disc Damage Likelhood Score suggesting that the Disc Damage Likelhood Score is a more effective means of discriminating people with and without disease.

Scanning laser polarimetry (GDx).

Twenty-four studies included an investigation of GDx with variable corneal compensation. [44 17 69 19 71 57 21 72 39 73 25 58 74 59 28 60 29 63 75 76 77 30 70 31] The aim of seven studies was to discriminate early disease from no disease .[44 17 19 72 39 59 75] In the studies that focused on early OAG, the range of sensitivity across all comparisons and cut-offs for the most frequently reported parameter, Temporal, Superior, Nasal, Inferior, Temporal average, was 29.8 to 81.63 percent. Specificity was fixed at 80, 90, or 95 percent in three studies, and the lowest reported specificity was 66.36 percent. The range in sensitivity for the nerve fiber indicator parameter across all comparisons and cut-offs was 28.3 to 93.3 percent. Specificities ranged from 52.9 percent to fixed at 80, 90, and 95 percent.

Three studies examined the GDx with enhanced corneal compensation [73 58 74] The sample sizes of the included studies ranged from 63 to 92 glaucoma participants and 41 to 95 healthy volunteers. Medeiros (2007) compared the AUCs for GDx with variable corneal compensation and GDx with enhanced corneal compensation and reported that GDx with enhanced corneal compensation performed significantly better than GDx with variable corneal compensation for the parameters Temporal, Superior, Nasal, Inferior, Temporal average, Superior average, and Inferior average ($p =<0.01$).[73] Sehi (2007) [58] and Mai (2007) [74] concurred with Medeiros (2007) that imaging with enhanced corneal compensation appears to improve the ability to diagnose OAG.

RNFL photography

Two studies examined the accuracy of Retinal Nerve Fiber Layer (RNFL) photography.[59 70] Hong (2007b) analyzed RNFL photographs of 72 glaucoma and 48 healthy participants taken with the Heidelberg Retina Angiograph 1.[59] Two investigators, unaware of the participant's diagnosis, reviewed the photographs. A third investigator served as an adjudicator to resolve any disagreements. Results showed the RNFL defect score II, with an AUC of 0.75 ($p < 0.001$), was the best parameter for discriminating early glaucoma and healthy eyes (sensitivity 58.3 percent and specificity 95.8 percent).

Medeiros (2004a) compared RNFL photography to the GDx with variable corneal compensation in 42 participants with OAG, 32 OAG suspects, and 40 healthy volunteers.[70] Investigators photographed one eye of each participant using the Topcon TRC-50VT camera and Kodak Kodalith high-contrast film (red-free filter). Two investigators used a set of 25 reference photos to score photographs. A third investigator adjudicated disagreements. The sensitivities of the global RNFL score were 36 and 81 percent respectively for fixed specificities of 95 and 80 percent. At a fixed specificity of 95 percent, the sensitivity of the Nerve Fiber Indicator was 71 percent versus the 36 percent reported above for red-free photos. Overall, the global RNFL score determined from red-free photos did not perform as well as scanning laser polarimetry. The area under the ROC curve was 0.91 for the GDx with variable corneal compensation Nerve Fiber Indicator versus 0.84 for the global RNFL score.

Tests of optic nerve function

FDT 24-2 Perimetry

Five studies examined the diagnostic accuracy of FDT 24-2 threshold tests using the Humphrey Matrix Perimeter.[63 78-81] All studies included participants with known glaucoma and healthy volunteers. The range of sample size was 25 to 174 glaucomatous eyes and 15 to 164 healthy eyes. Sensitivities and specificities were reported for the parameters mean deviation, pattern standard deviation and glaucoma hemifield test outside of normal limits. There was appreciable heterogeneity in the estimates of sensitivity at 80, 90 and 95 percent specificity that may be attributed to a number of factors including different patient populations and variations in cut-off points. The sensitivity for the mean deviation was 55 and 94 percent at 80 percent fixed specificity. [79 80] Tafreshi (2009) and Leeprechanon (2007) reported 39 and 87 percent at 90 percent fixed specificity, and 32 and 82 percent at fixed 95 percent specificity respectively.[78 80] Sensitivity and specificity for pattern standard of deviation (PSD) and glaucoma hemifield test are reported with respective cut-off points in the evidence tables in Appendix C.

Bagga (2006) [63] and Burgansky-Eliash (2007) [81] reported AUC for the mean deviation (0.69 for both studies with $p < 0.04$ and 95 percent CI, 0.564 to 0.815 respectively). The AUCs for PSD were 0.66 ($p = 0.09$) [63] and 0.733 (95 percent CI, 0.618 to 0.848).[81]

FDT 30-2 Perimetry

Two studies discuss the detection of early glaucoma using the FDT 30-2 threshold test with the Humphrey Matrix Perimeter. [59 82] Both Hong (2007a) [82] and Hong (2007b) [59] enrolled OAG participants with early visual field loss and healthy controls. The mean deviation and PSD were judged to be good parameters for distinguishing between eyes with early disease and eyes with no known defects. The mean deviation was 0.795 and 0.750 and the PSD 0.808 and 0.934 for Hong (2007a) and Hong (2007b) respectively. Both study groups, however, determined that the

best parameter for distinguishing eyes with early glaucoma from healthy eyes was the number of points that have p less than 5 percent in the pattern deviation plot with AUCs of 0.985 (95 percent CI, 0.943 to 0.998) and 0.990 ($p < 0.001$) in Hong (2007a) and Hong (2007b) respectively.

FDT C-20 Perimetry

Three studies discussed the accuracy of FDT C-20 perimetry.[17 83 84] Pueyo (2009) enrolled 130 participants with ocular hypertension and 48 healthy volunteers.[17] Using a cut-off of a cluster of at least four points with a sensitivity outside 95 percent normal limits, or three points outside 98 percent, or at least one point outside 99 percent, investigators determined the sensitivity of FDT to be 31.25 percent and specificity 72.9 percent among the subset of 32 participants with glaucomatous optic neuropathy (of the 130 with ocular hypertension). The investigators concluded that FDT might not be an ideal test for participants with early defects.

Salim (2009) enrolled 35 participants with known OAG and 35 age- and sex-matched controls with no evidence of glaucoma. Investigators used FDT, non-contact tonometry, and a questionnaire individually and in all possible combinations to determine the accuracy of single and combination tests.[83] Sensitivity of FDT was 58.1 percent and specificity 98.6 percent. Overall, FDT was determined to be the best among the candidate single and combination tests in the study, despite fair sensitivity for detecting OAG.

Pierre-Filho (2006) enrolled glaucoma patients who had never experienced perimetry prior to the study.[84] The investigators reported that 21 (32.8 percent) of the 64 participants with glaucoma were identified as having early disease, but data were not provided for this subgroup. Sensitivity and specificity were 85.9 and 73.6 percent for the presence of at least one abnormal location and 82.8 and 83 percent respectively for two or more abnormal locations regardless of severity.

FDT N30 Perimetry

Four studies examined the accuracy of the FDT N30 threshold test.[19 79 85 86] Zeppieri (2010) focused on the detection of early glaucoma among a sample of 75 participants with OAG, 87 with ocular hypertension, 67 with glaucomatous optic neuropathy and 90 healthy volunteers.[19] At the best cut-off of less than -0.78, the sensitivity of the mean deviation parameter was 61.3 percent and the specificity was 73.7 percent for distinguishing early OAG from healthy eyes. At the best cut-off of greater than 3.89, the sensitivity of the PSD was 76.0 percent and the specificity was 87.8 percent. The investigators concluded that, "FDT can potentially detect eyes with very early functional defects that do not show structural changes in patients at risk of developing glaucoma." Salvetat (2010) focused on the detection of early disease among a sample of 52 participants with early OAG and 53 healthy volunteers.[86] The sensitivity of mean deviation for distinguishing early OAG from healthy eyes at the best cut-off (less than -1.12) was 67 percent and the specificity was 74 percent. At the best cut-off of greater than 3.97, the sensitivity of the parameter PSD was 96 percent and the specificity was 85 percent.

Humphrey [Visual] Field Analyzer (HFA)

Eleven studies examined the diagnostic accuracy of the HFA. Of these, six examined HFA Short Wavelength Automated Perimetry,[17 43 63 78 85 87] two tested HFA-SAP, (SAP)-SITA and HFA SAP-Full Threshold (FT), [78 85] four examined HFA-SITA-Standard, [32 77 81 84] and one tested the HFA SITA-Fast protocol.[84] The HFA Short Wavelength Automated Perimetry testing

protocol (the most frequently reported) included 25 to 286 participants with glaucoma and 22 to 289 healthy volunteers across the six included studies. Sensitivity across all comparisons and cut-offs for the mean deviation was 25.9 to 83 percent. Specificity was 80 to 95.2 percent. Cut-offs ranged from -5.42 to -11.06 dB.

Goldmann Applanation Tonometry (GAT)

Baggam (2006) compared the ability of various tests of structure and function to discriminate healthy eyes (22) from eyes with known glaucomatous optic neuropathy (25)[63] The AUC for inner ocular pressure, as measured by GAT, was 0.66 (p = 0.05).

Non-Contact Tonometry

Salim (2009) included non-contact tonometry, individually and in all possible combinations, with other measures of structure and function to determine the accuracy of single and combination tests.[83] Intraocular pressure, as measured by non-contact tonometry, was found not to be a very sensitive test for detecting glaucoma (sensitivity 22.1 percent). The investigators acknowledge that use of topical medications by the glaucoma participants could limit the ability to identify those with disease.

Other tests of optic nerve function

One study included the Octopus 301 perimeter84.

Table 1 Grading of Evidence

Number of studies; participants	Risk of bias	Consistency	Directness	Precision	Strength of evidence
Sensitivity and specificity of candidate screening tests					
72; 15,000+	Observational /High	Inconsistent	Direct	Imprecise	Low

Conclusion

Based on the Burr (2007) analysis,[1] standard automated perimetry was compared with other tests available at that time. With regard to sensitivity, SAP was better than Goldmann tonometry, similar to HRT, and worse than disc photos or FDT. In terms of specificity, SAP was better than disc photos and FDT, similar to HRT, and worse than Goldmann tonometry.

We identified several additional studies assessing the performance of glaucoma screening tests not included in the Burr review. The studies included newer imaging (GDx, OCT, HRT-3) and functional (Short Wavelength Automated Perimetry, new FDT patterns) technologies. However, despite improvements in the technology, it's still not clear that there is any one test or combination of tests suitable for use in glaucoma screening. Two significant barriers that remain in terms of identifying and characterizing potential glaucoma screening tests are the lack of a definitive diagnostic reference standard for glaucoma and the heterogeneity in the design of the

studies. Because of these, the ranges of sensitivities, specificities, and areas under the ROC curve are large and prevent a coherent synthesis.

Key Question 4

We did not identify any studies that addressed whether participation in an OAG screening-based program leads to reductions in intraocular pressure when compared to another screening-based program or no screening.

Key Question 5

Evidence from Systematic Reviews

Hatt (2006) undertook a systematic review of screening modalities for OAG compared to no screening (including opportunistic case finding and referral). There were no restrictions on included populations.[88] The primary outcome of interest was the prevalence of visual field loss, defined as the proportion of participants with a prespecified severity of visual field loss diagnosed by either manual or automated field assessment. Other primary outcomes included the prevalence of optic nerve damage and visual impairment. The focus was on identifying randomized controlled trials, as these are the optimal study design for addressing lead-time bias and length bias. Electronic searches of five databases including MEDLINE and CENTRAL were conducted in 2006 and again in January 2009, but none of the studies that were identified were eligible for inclusion. The review authors acknowledge that randomized controlled trials require lengthy follow-up and are predicated on identifying appropriate candidate tests that may be incorporated into a screening-based program.

Detailed Analysis of Primary Studies

We did not identify any primary studies that addressed whether participation in an OAG screening-based program leads to reductions in visual field loss or optic nerve damage when compared to another screening-based program or no screening.

Key Question 6

We did not identify any studies addressing the harms associated with screening for OAG.

Discussion

Multiple prior reviews of the evidence about glaucoma screening have failed to identify appropriate studies to address the question. The evidence report commissioned by the Agency for Healthcare Research and Quality in 2005 found inadequate data to draw conclusions about the safety and efficacy of screening for glaucoma.[7] Our reconsideration of the question did not identify any new studies to address five of the six Key questions in our analytic framework (Figure 1).

Given the results of prior work, we attempted to enhance the ability of this project to inform decisions about screening by including intermediate outcomes in our analytic framework. Key Questions 4 and 5 connected glaucoma screening programs with either optic nerve damage or

visual field loss. These outcomes could then potentially be linked by other studies to the final outcomes of visual impairment and patient reported outcomes. Despite accommodating the potential for evidence that could lead stepwise from screening to final outcomes, we were also unable to find evidence that provided support for or against glaucoma screening.

In contrast to the prior evidence report, we also included a question seeking to identify evidence regarding the relative effectiveness of the various tests that might be used in a glaucoma-screening program (Key Question 3). Building on the comprehensive evaluation by Burr (2007)[1], we did identify a number of new studies evaluating such devices. While there is now more evidence regarding optical coherence tomography, the Heidelberg retina tomography-3 (HRT-3), and the glaucoma diagnosis scanning laser polarimeters, the ability of these devices to identify glaucoma in a screening setting is not clear for at least two reasons: the lack of a single diagnostic standard for glaucoma and the high degree of variability in the largely cross-sectional studies of diagnostic accuracy.

The lack of diagnostic standards continues to complicate all studies of glaucoma, including those of screening. The lack of standard definitions results in studies that attempt to address the same questions using different definitions, thereby preventing synthesis of study results. Use of a standard, such as that proposed by Foster (2002)[4] in studies of glaucoma screening or of devices potentially used in screening, would help overcome this problem.

Many of the diagnostic studies include healthy volunteers as well as those known to have glaucoma at the time of screening. Including participants who are not representative of those one reasonably expects to encounter in a screening setting may lead to biased estimates of diagnostic performance and limit the generalizability of findings.

More uniform reporting of participant characteristics would also enhance diagnostic studies. Since inclusion criteria are highly variable and the important characteristics of the resulting populations are not uniformly described, synthesis across studies is difficult. Better characterization of participants would also help address the question of whom to screen for glaucoma. It's clear that discriminating healthy participants from those with early glaucoma is more difficult than discriminating healthy participants from those with moderate or advanced glaucoma. If participants were described in enough detail to distinguish those with mild, moderate, or severe disease, it would facilitate secondary questions regarding which groups should undergo screening and which stages of disease should be of primary interest. It may be the case, for instance, that identifying people with severe disease is a reasonable goal of a screening program since such people are likely at the highest risk of visual impairment. Screening for glaucoma is a difficult problem due to the fact that it is asymptomatic, has low prevalence, is typically only slowly progressive, and has no agreed upon standard for diagnosis. These issues, while challenging, might be overcome by with a combination of creative thinking with regard to populations amenable to screening and hard work on the needed studies and diagnostic standards.

Future Research Needs

Given the ongoing lack of evidence regarding screening for open-angle glaucoma (OAG), there's a clear need for appropriate research to fill the multiple gaps that exist. It has repeatedly been suggested that a standard for diagnosis be adopted to allow for synthesis of evidence and comparison of outcomes across studies. Glaucoma professional societies would be well suited to help address this issue and should be encouraged to do so, since it fundamentally limits research in a number of areas, including screening. Some of the important questions are outlined below using the standard Population-Intervention-Comparison-Outcomes format.

Does screening for glaucoma alter either intermediate or final outcomes?

Population

People at risk of glaucoma

- Age
- Race

Intervention

- Randomization to one or more screening tests for glaucoma
- Appropriate follow-up or treatment based on screening outcome

Outcome

Measurements of Intermediate outcomes

- Visual field.
- Optic nerve damage

Measurement of final outcomes

- Visual impairment
- Patient reported outcomes

What tests or devices are best able to identify people with glaucoma in a screening setting?

Population

- People with glaucoma of various stages, determined using standard criteria
- People without glaucoma, matched to the glaucoma group

Intervention

- Multiple candidate glaucoma screening tests applied to both groups

Outcome

- Measures of diagnostic accuracy (sensitivity, specificity, area under the ROC curve) for each device
- Assessment of the comparative effectiveness of devices

Are there groups in which screening might be more effective?

Population

- People at varying degrees of glaucoma risk based on population studies

Intervention

- One or more glaucoma screening programs or tests
- Definitive diagnosis based on standard definition of disease

Outcome

- The comparative effectiveness of various tests or programs in terms of identifying glaucoma

Are there stages of disease at which screening is more appropriate?

Population

- People with glaucoma of various stages, determined using standard criteria
- People without glaucoma, matched to the glaucoma group

Intervention

- One or more glaucoma screening programs or tests
- Definitive diagnosis based on standard definition of disease

Outcome

- Assessment of the relative benefit of screening for various stages of glaucoma

Reference List

1. Burr JM, Mowatt G, Hernındez R *et al*. The clinical effectiveness and cost-effectiveness of screening for open-angle glaucoma: A systematic review and economic evaluation. 11. 2007.

2. Quigley HA, Browman AT. The number of people with glaucoma worldwide in 2010 and 2020. 90. 2006:262-7.

3. The Eye Diseases Prevalence Research Group. Prevalence of OAG among adults in the United States. 122. 2004:532-8.

4. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. 86. 2002:238-42.

5. Quigley HA, Vitale S. Models of OAG prevalence and incidence in the United States. 38. 1997:83-91.

6. US Preventive Services Task Force. Screening for glaucoma: Recommendation statement. 2005.

7. Fleming C., Whitlock E., Beil T, Smit B., Harris RP. Screening for primary OAG in the primary care setting: An update for the US Preventive Services Task Force. 3. 2005:167-70.

8. International Classification of Diseases, Ninth Revision (ICD-9). Centers for Disease Control and Prevention, 2009.

9. Li T. Bridging the gap between evidence generation and clinical decision making [Ph.D. dissertation] United States -- Maryland: The Johns Hopkins University, 2010.

10. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 (updated September 2009). The Cochrane Collaboration, 2009.

11. Wells GA, Shea B, O'Connell D *et al*. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses.

12. 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. 3. 2003.

13. Atkins D, Best D, Briss PA *et al*. Grading quality of evidence and strength of recommendations. 328. 2004:1490.

14. Owens DK, Lohr KN, Atkins D *et al*. AHRQ series paper 5: Grading the strength of a body of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Health-care Program. 63. 2010:513-23.

15. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. 148. 2008:776-82.

16. Healey PR, Lee AJ, Aung T, Wong TY, Mitchell P. Diagnostic accuracy of the Heidelberg Retina Tomograph for glaucoma a population-based assessment. Ophthalmology 2010; 117(9):1667-73.

17. Pueyo V, Polo V, Larrosa JM, Pablo LE, Ferreras A, Honrubia FM. Ability of optical imaging devices to detect early structural damage in ocular hypertension. Ann Ophthalmol (Skokie) 2009; 41(3-4):150-6.

18. Zheng Y, Wong TY, Lamoureux E *et al*. Diagnostic ability of Heidelberg Retina Tomography in detecting glaucoma in a population setting: The Singapore Malay Eye Study. Ophthalmology 2010; 117(2):290-7.

19. Zeppieri M, Brusini P, Parisi L, Johnson CA, Sampaolesi R, Salvetat ML. Pulsar perimetry in the diagnosis of early glaucoma. Am J Ophthalmol 2010; 149(1):102-12.

20. Saito H, Tsutsumi T, Araie M, Tomidokoro A, Iwase A. Sensitivity and specificity of the Heidelberg Retina Tomograph II Version 3.0 in a population-based study: The Tajimi Study. Ophthalmology 2009; 116(10):1854-61.

21. Medeiros FA, Vizzeri G, Zangwill LM, Alencar LM, Sample PA, Weinreb RN. Comparison of retinal nerve fiber layer and optic disc imaging for diagnosing glaucoma in patients suspected of having the disease. Ophthalmology 2008; 115(8):1340-6.

22. Ferreras A, Pablo LE, Pajarin AB, Larrosa JM, Polo V, Pueyo V. Diagnostic ability of the Heidelberg Retina Tomograph 3 for glaucoma. Am J Ophthalmol 2008; 145(2):354-9.

23. De Leon-Ortega JE, Sakata LM, Monheit BE, McGwin G Jr, Arthur SN, Girkin CA. Comparison of diagnostic accuracy of Heidelberg Retina Tomograph II and Heidelberg Retina Tomograph 3 to discriminate glaucomatous and nonglaucomatous eyes. Am J Ophthalmol 2007; 144(4):525-32.

24. Naithani P, Sihota R, Sony P *et al*. Evaluation of optical coherence tomography and heidelberg retinal tomography parameters in detecting early and moderate glaucoma. Invest Ophthalmol Vis Sci 2007; 48(7):3138-45.

25. Pueyo V, Polo V, Larrosa JM, Ferreras A, Pablo LE, Honrubia FM. Diagnostic ability of the Heidelberg retina tomograph, optical coherence tomograph, and scanning laser polarimeter in OAG. J Glaucoma 2007; 16(2):173-7.

26. Uysal Y, Bayer A, Erdurman C, Kilic S. Sensitivity and specificity of Heidelberg Retinal Tomography II parameters in detecting early and moderate glaucomatous damage: Effect of disc size. Clin Experiment Ophthalmol 2007; 35(2):113-8.

27. Burgansky-Eliash Z, Wollstein G, Bilonick RA, Ishikawa H, Kagemann L, Schuman JS. Glaucoma detection with the Heidelberg retina tomograph 3. Ophthalmology 2007; 114(3):466-71.

28. Shah NN, Bowd C, Medeiros FA *et al*. Combining structural and functional testing for detection of glaucoma. Ophthalmology 2006; 113(9):1593-602.

29. Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Influence of disease severity and optic disc size on the diagnostic performance of imaging instruments in glaucoma. Invest Ophthalmol Vis Sci 2006; 47(3):1008-15.

30. Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. Comparison of the GDx-variable corneal compensation scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and stratus OCT optical coherence tomograph for the detection of glaucoma. Arch Ophthalmol 2004; 122(6):827-37.

31. Ferreras A, Polo V, Larrosa JM *et al*. Can frequency-doubling technology and short-wavelength automated perimetries detect visual field defects before standard automated perimetry in patients with preperimetric glaucoma? J. Glaucoma 2007; 16(4):372-83.

32. Danesh-Meyer HV, Gaskin BJ, Jayusundera T, Donaldson M, Gamble GD. Comparison of disc damage likelihood scale, cup-to-disc ratio, and Heidelberg retina tomograph in the diagnosis of glaucoma. Br. J. Ophthalmol. 2006; 90(4):437-41.

33. Moreno-Montanes J, Anton A, Garcia N, Olmo N, Morilla A, Fallon M. Comparison of retinal nerve fiber layer thickness values using Stratus Optical Coherence Tomography and Heidelberg Retina Tomograph-III. J Glaucoma 2009; 18(7):528-34.

34. Bozkurt B, Irkec M, Arslan U. Diagnostic accuracy of Heidelberg Retina Tomograph III classifications in a Turkish primary OAG population. Acta Ophthalmol 2010; 88(1):125-30.

35. Reddy S, Xing D, Arthur SN *et al*. HRT III glaucoma probability score and Moorfields regression across the glaucoma spectrum. J Glaucoma 2009; 18(5):368-72.

36. Oddone F, Centofanti M, Iester M *et al*. Sector-based analysis with the Heidelberg Retinal Tomograph 3 across disc sizes and glaucoma stages: A multicenter study. Ophthalmology 2009; 116(6):1106-11.e1-3.

37. Takmaz T, Can I. Comparison of glaucoma probability score and Moorfields regression analysis to discriminate glaucomatous and healthy eyes. Eur J Ophthalmol 2009; 19(2):207-13.

38. Moreno-Montanes J, Anton A, Garcia N, Mendiluce L, Ayala E, Sebastian A. Glaucoma probability score vs Moorfields classification in normal, ocular hypertensive, and glaucomatous eyes. Am J Ophthalmol 2008; 145(2):360-8.

39. Badala F, Nouri-Mahdavi K, Raoof DA, Leeprechanon N, Law SK, Caprioli J. Optic disk and nerve fiber layer imaging to detect glaucoma. Am J Ophthalmol 2007; 144(5):724-32.

40. Ferreras A, Pajarin AB, Polo V, Larrosa JM, Pablo LE, Honrubia FM. Diagnostic ability of Heidelberg Retina Tomograph 3 classifications: Glaucoma probability score versus Moorfields regression analysis. Ophthalmology 2007; 114(11):1981-7.

41. Mansoori T, Viswanath K, Balakrishna N. Quantification of Retinal Nerve Fiber Layer Thickness in Normal Eyes, Eyes with Ocular Hypertension, and Glaucomatous Eyes with standard deviation-OCT. Ophthalmic Surg Lasers Imaging 2010; 41(6):S50-7.

42. Fang Y, Pan YZ, Li M, Qiao RH, Cai Y. Diagnostic capability of Fourier-Domain optical coherence tomography in early primary open-angle glaucoma. Chin Med J (Engl) 2010; 123(15):2045-50.

43. Zhong Y, Zhou X, Cheng Y, Xie L. Relation between blue-on-yellow perimetry and optical coherence tomography in normal tension glaucoma. Can J Ophthalmol 2010; 45(5):494-500.

44. Pablo LE, Ferreras A, Schlottmann PG. Retinal nerve fibre layer evaluation in ocular hypertensive eyes using optical coherence tomography and scanning laser polarimetry in the diagnosis of early glaucomatous defects. Br J Ophthalmol 2010.

45. Rao HL, Zangwill LM, Weinreb RN, Sample PA, Alencar LM, Medeiros FA. Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. Ophthalmology 2010; 117(9):1692-9, 1699.e1.

46. Leite MT, Zangwill LM, Weinreb RN *et al*. Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis. Invest Ophthalmol Vis Sci 2010; 51(8):4104-9.

47. Li G, Fansi AK, Boivin JF, Joseph L, Harasymowycz P. Screening for glaucoma in high-risk populations using optical coherence tomography. Ophthalmology 2010; 117(3):453-61.

48. Park SB, Sung KR, Kang SY, Kim KR, Kook MS. Comparison of glaucoma diagnostic Capabilities of Cirrus HD and Stratus optical coherence tomography. Arch Ophthalmol 2009; 127(12):1603-9.

49. Chang RT, Knight OJ, Feuer WJ, Budenz DL. Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma. Ophthalmology 2009; 116(12):2294-9.

50. Moreno-Montanes J, Olmo N, Alvarez A, Garcia N, Zarranz-Ventura J. Cirrus high-definition optical coherence tomography compared with Stratus optical coherence tomography in glaucoma diagnosis. Invest Ophthalmol Vis Sci 2010; 51(1):335-43.

51. Sehi M, Grewal DS, Sheets CW, Greenfield DS. Diagnostic ability of Fourier-domain vs. time-domain optical coherence tomography for glaucoma detection. Am J Ophthalmol 2009; 148(4):597-605.

52. Yuksel N, Altintas O, Ozkan B, Karadag S, Caglar Y. Discriminating ability of optical coherence tomography data in staging glaucomatous damage. Can J Ophthalmol 2009; 44(3):297-307.

53. Sung KR, Kim DY, Park SB, Kook MS. Comparison of retinal nerve fiber layer thickness measured by Cirrus HD and Stratus optical coherence tomography. Ophthalmology 2009; 116(7):1264-70, 1270.e1.

54. Polo V, Larrosa JM, Ferreras A, Mayoral F, Pueyo V, Honrubia FM. Retinal nerve fiber layer evaluation in OA. Optimum criteria for optical coherence tomography. Ophthalmologica 2009; 223(1):2-6.

55. Lu AT, Wang M, Varma R *et al*. Combining nerve fiber layer parameters to optimize glaucoma diagnosis with optical coherence tomography. Ophthalmology 2008; 115(8):1352-7, 1357.e1-2.

56. Nouri-Mahdavi K, Nikkhou K, Hoffman DC, Law SK, Caprioli J. Detection of early glaucoma with optical coherence tomography (Stratus OCT). J Glaucoma 2008; 17(3):183-8.

57. Takahashi H, Chihara E. Impact of diabetic retinopathy on quantitative retinal nerve fiber layer measurement and glaucoma screening. Invest Ophthalmol Vis Sci 2008; 49(2):687-92.

58. Sehi M, Ume S, Greenfield DS. Scanning laser polarimetry with enhanced corneal compensation and optical coherence tomography in normal and glaucomatous eyes. Invest Ophthalmol Vis Sci 2007; 48(5):2099-104.

59. Hong S, Ahn H, Ha SJ, Yeom HY, Seong GJ, Hong YJ. Early glaucoma detection using the Humphrey Matrix Perimeter, GDx-variable corneal compensation Stratus OCT, and retinal nerve fiber layer photography. Ophthalmology 2007; 114(2):210-5.

60. Brusini P, Salvetat ML, Zeppieri M, Tosoni C, Parisi L, Felletti M. Comparison between GDx-variable corneal compensation scanning laser polarimetry and Stratus OCT optical coherence tomography in the diagnosis of chronic glaucoma. Acta Ophthalmol Scand 2006; 84(5):650-5.

61. Sihota R, Sony P, Gupta V, Dada T, Singh R. Diagnostic capability of optical coherence tomography in evaluating the degree of glaucomatous retinal nerve fiber damage. Invest Ophthalmol Vis Sci 2006; 47(5):2006-10.

62. Chen HY, Wang TH, Lee YM, Hung TJ. Retinal nerve fiber layer thickness measured by optical coherence tomography and its correlation with visual field defects in early glaucoma. J Formos Med Assoc 2005; 104(12):927-34.

63. Bagga H, Feuer WJ, Greenfield DS. Detection of psychophysical and structural injury in eyes with glaucomatous optic neuropathy and normal standard automated perimetry. Arch Ophthalmol 2006; 124(2):169-76.

64. Leung CK, Chan WM, Yung WH *et al*. Comparison of macular and peripapillary measurements for the detection of glaucoma: An optical coherence tomography study. Ophthalmology 2005; 112(3):391-400.

65. Medeiros FA, Zangwill LM, Bowd C, Vessani RM, Susanna R Jr., Weinreb RN. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. Am J Ophthalmol 2005; 139(1):44-55.

66. Wollstein G, Ishikawa H, Wang J, Beaton SA, Schuman JS. Comparison of three optical coherence tomography scanning areas for detection of glaucomatous damage. Am J Ophthalmol 2005; 139(1):39-43.

67. Leung CK, Yung WH, Ng AC, Woo J, Tsang MK, Tse KK. Evaluation of scanning resolution on retinal nerve fiber layer measurement using optical coherence tomography in normal and glaucomatous eyes. J Glaucoma 2004; 13(6):479-85.

68. Mori S, Hangai M, Sakamoto A, Yoshimura N. Spectral-domain optical coherence tomography measurement of macular volume for diagnosing glaucoma. J. Glaucoma 2010; 19(8):528-34.

69. Reus NJ, Lemij HG, Garway-Heath DF *et al*. Clinical assessment of stereoscopic optic disc photographs for glaucoma: The European Optic Disc Assessment Trial. Ophthalmology 2010; 117(4):717-23.

70. Medeiros FA, Zangwill LM, Bowd C, Mohammadi K, Weinreb RN. Comparison of scanning laser polarimetry using variable corneal compensation and retinal nerve fiber layer photography for detection of glaucoma. Arch Ophthalmol 2004; 122(5):698-704.

71. Chen HY, Huang ML, Tsai YY, Hung PT, Lin EJ. Comparing glaucomatous optic neuropathy in primary open angle and primary angle closure glaucoma eyes by scanning laser polarimetry-variable corneal compensation. J Glaucoma 2008; 17(2):105-10.

72. Parikh RS, Parikh SR, Kumar RS, Prabakaran S, Babu JG, Thomas R. Diagnostic capability of scanning laser polarimetry with variable cornea compensator in Indian patients with early primary OAG. Ophthalmology 2008; 115(7):1167-72.e1.

73. Medeiros FA, Bowd C, Zangwill LM, Patel C, Weinreb RN. Detection of glaucoma using scanning laser polarimetry with enhanced corneal compensation. Invest Ophthalmol Vis Sci 2007; 48(7):3146-53.

74. Mai TA, Reus NJ, Lemij HG. Diagnostic accuracy of scanning laser polarimetry with

enhanced versus variable corneal compensation. Ophthalmology 2007; 114(11):1988-93.

75. Kanamori A, Nagai-Kusuhara A, Escano MF, Maeda H, Nakamura M, Negi A. Comparison of confocal scanning laser ophthalmoscopy, scanning laser polarimetry and optical coherence tomography to discriminate ocular hypertension and glaucoma at an early stage. Graefes Arch Clin Exp Ophthalmol 2006; 244(1):58-68.

76. Da Pozzo S, Iacono P, Marchesan R, Fantin A, Ravalico G. Scanning laser polarimetry with variable corneal compensation and detection of glaucomatous optic neuropathy. Graefes Arch Clin Exp Ophthalmol 2005; 243(8):774-9.

77. Reus NJ, Lemij HG. Diagnostic accuracy of the GDx- variable corneal compensation for glaucoma. Ophthalmology 2004; 111(10):1860-5.

78. Tafreshi A, Sample PA, Liebmann JM *et al*. Visual function-specific perimetry to identify glaucomatous visual loss using three different definitions of visual field abnormality. Invest Ophthalmol Vis Sci 2009; 50(3):1234-40.

79. Racette L, Medeiros FA, Zangwill LM, Ng D, Weinreb RN, Sample PA. Diagnostic accuracy of the Matrix 24-2 and original N-30 frequency-doubling technology tests compared with standard automated perimetry. Invest Ophthalmol Vis Sci 2008; 49(3):954-60.

80. Leeprechanon N, Giangiacomo A, Fontana H, Hoffman D, Caprioli J. Frequency-doubling perimetry: Comparison with standard automated perimetry to detect glaucoma. Am J Ophthalmol 2007; 143(2):263-71.

81. Burgansky-Eliash Z, Wollstein G, Patel A *et al*. Glaucoma detection with matrix and standard achromatic perimetry. Br. J. Ophthalmol. 2007; 91(7):933-8.

82. Hong S, Chung W, Hong YJ, Seong GJ. Discriminating ability of Humphrey matrix perimetry in early glaucoma patients. Ophthalmologica 2007; 221(3):195-9.

83. Salim S, Netland PA, Fung KH, Smith ME, Aldridge A. Assessment of the Student Sight Savers Program methods for glaucoma screening. Ophthalmic Epidemiol 2009; 16(4):238-42.

84. Pierre-Filho Pde T, Schimiti RB, de Vasconcellos JP, Costa VP. Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, SITA Standard and SITA Fast perimetry in perimetrically inexperienced individuals. Acta Ophthalmol Scand 2006; 84(3):345-50.

85. Sample PA, Medeiros FA, Racette L *et al*. Identifying glaucomatous vision loss with visual-function-specific perimetry in the diagnostic innovations in glaucoma study. Invest Ophthalmol Vis Sci 2006; 47(8):3381-9.

86. Salvetat ML, Zeppieri M, Tosoni C, Parisi L, Brusini P. Non-conventional perimetric

methods in the detection of early glaucomatous functional damage. Eye 2010; 24(5):835-42.

87. Ng M, Racette L, Pascual JP *et al*. Comparing the full-threshold and Swedish interactive thresholding algorithms for short-wavelength automated perimetry. Invest Ophthalmol Vis Sci 2009; 50(4):1726-33.

88. Hatt S, Wormald R, Burr J. Screening for prevention of optic nerve damage due to chronic open-angle glaucoma. Cochrane Database Syst Rev 2006; (4):CD006129.

ABBREVIATIONS

AHRQ	Agency for Healthcare Research and Quality
AUC	Area under the receiver operating characteristic curve
COAG	Chronic OAG
EPC	Evidence-based Practice Center
FDT	Frequency doubling technology
GAT	Goldmann applanation tonometry
GDx	Glaucoma diagnosis
HRT	Heidelberg retina tomography
IOP	Intraocular pressure
ITT	Intention to treat
logMAR	Logarithm of the minimum angle of resolution
MRA	Moorfields Regression Analysis
MD	Mean deviation
NEI-VQF	National Eye Institute Visual Functioning Questionnaire
NFI	Nerve fiber indicator
NCT	Non-contact tonometry
NOS	Newcastle-Ottawa Scale
OAG	Open-Angle Glaucoma
OCT	Optical coherence tomography
OHT	Ocular hypertension
PSD	Pattern standard deviation
QOL	Quality of life
RCT	Randomized controlled trial
RNFL	Retinal Nerve Fiber Layer
RR	Relative risk
SAP	Standard automated perimetry
SD	Standard deviation
SITA	Swedish interactive threshold algorithm
SLP	Scanning laser polarimetry
SLT	Selective laser trabeculoplasty
SR	Systematic review
SWAP	Short wavelength automated perimetry
TSNIT	Temporal, Superior, Nasal, Inferior, Temporal
USPSTF	U.S. Preventive Services Task Force
VCC	Variable corneal compensation
VCDR	Vertical cup to disc ratio
VAS	Visual analogue scale
VF	Visual field