Accuracy of optical coherence tomography for diagnosing glaucoma: an overview of systematic reviews

Manuele Michelessi[1], **Tianjing Li**[2], **Alba Miele**[3], **Augusto Azuara-Blanco**[4], **Riaz Qureshi**[5], **Gianni Virgili**[3]

[1]IRCCS—Fondazione Bietti, Rome, Italy

[2]Department of Ophthalmology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

[3]Eye Clinic, Department of NEUROFARBA, University of Florence, Florence, Italy

[4]Centre for Public Health, Queen's University Belfast, Belfast, UK

[5]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

Aims—To assess the diagnostic accuracy (DTA) of optical coherence tomography (OCT) for detecting glaucoma by systematically searching and appraising systematic reviews (SRs) on this issue.

Methods—We searched a database of SRs in eyes and vision maintained by the Cochrane Eyes and Vision United States on the DTA of OCT for detecting glaucoma. Two authors working independently screened the records, abstracted data and assessed the risk of bias using the Risk of Bias in Systematic Reviews checklist. We extracted quantitative DTA estimates as well as qualitative statements on their relevance to practice.

Results—We included four SRs published between 2015 and 2018. These SRs included between 17 and 113 studies on OCT for glaucoma diagnosis. Two reviews were at low risk of bias and the other two had two to four domains at high or unclear risk of bias with concerns on applicability. The two reliable SRs reported the accuracy of average retinal nerve fibre layer (RNFL) thickness and found a sensitivity of 0.69 (0.63 to 0.73) and 0.78 (0.74 to 0.83) and a specificity of 0.94 (0.93 to 0.95) and 0.93 (0.92 to 0.95) in 57 and 50 studies, respectively. Only one review included a

clear specification of the clinical pathway. Both reviews highlighted the limitations of primary DTA studies on this topic.

Conclusions—The quality of published DTA reviews on OCT for diagnosing glaucoma was mixed. Two reliable SRs found moderate sensitivity at high specificity for average RNFL thickness in diagnosing manifest glaucoma. Our overview suggests that the methodological quality of both primary and secondary DTA research on glaucoma is in need of improvement.

INTRODUCTION

Optical coherence tomography (OCT) is a high-resolution imaging technology that allows quantitative and objective structural measurements of the optic nerve head (ONH) and retina.[1–3] Although the role of OCT in the clinical pathway has not yet been completely clarified, it is routinely used as a complementary test for diagnosing and monitoring glaucoma, and those at risk of glaucoma.[45] OCT could be also used to optimise referrals by primary eye care practitioners to ophthalmologists, or as a screening tool in a community setting.[67] Many primary studies as well as review articles have been published in the literature to explore the ability of OCT for detecting glaucoma at different stages of the disease.

Reliable systematic reviews (SRs) are necessary to synthetise current evidence and, as suggested by the Institute of Medicine in the USA (now called the National Academies of Sciences, Engineering, and Medicine), should be used in the development of clinical guidelines to inform the health decision-making and to improve the standard of care.[89]

With this goal in mind, in 2018, the European Glaucoma Society (EGS) and the Cochrane Eyes and Vision United States (CEV@US) collaborated on updating the EGS guidelines using reliable SRs. The partnership followed a model between CEV@US and the American Academy of Ophthalmology, in which CEV@US identifies potentially relevant reliable SRs for updating the Preferred Practice Patterns guidelines.[10–13]

In this report, we summarise the current evidence on DTA SRs addressing the accuracy of OCT for detecting glaucoma, as collected by CEV@US to be used in the EGS guidelines. Specifically, we examined the reliability of DTA SRs on OCT for glaucoma detection, how the diagnostic clinical pathways were described in these reviews, the diagnostic accuracy of OCT in diagnosing glaucoma and the overall impact of these findings on the role of OCT in glaucoma care.

METHODS

Search strategy

CEV@US maintains a database of SRs in vision research and eye care. The database was initially created in 2007 and subsequently updated in 2009, 2012, 2014, 2016, 2017, 2018 and 2019. A review article was considered eligible for this database when the title or abstract or full-text report claimed to be a 'systematic review' or 'meta-analysis' of studies on an eyes and vision condition, or when it met the definition of a SR or meta-analysis, as defined by the Institute of Medicine.[14] More details can be found elsewhere.[15]

For the EGS guidelines, EGS members and CEV@US investigators screened the records to identify SRs of various interventions and diagnostic tools for glaucoma. Two individuals independently examined the titles and abstracts, and then full text reports using Covidence. Any disagreement was resolved by a third senior member of the team.

For this report, from the group of SRs on diagnostic tools, we selected SRs that evaluated the diagnostic accuracy of OCT for diagnosing glaucoma and reported relevant measures (eg, sensitivity, specificity, area under the receiver operating characteristic curve (AUC), likelihood ratio, positive and negative predictive value). We included SRs assessing OCT alone or in comparison with other imaging methods, and we accepted any diagnosis of glaucoma as given by reviews authors. We considered only SRs that evaluated the more recent version of OCT with spectral-domain (SD) technology. We excluded narrative reviews, reviews of method and those assessing only secondary forms of glaucoma such as pseudoexfoliative or pigmentary glaucoma. We also excluded reviews that included only population- based studies for screening purposes.

Data extraction and risk-of-bias assessment

Two investigators independently extracted the following information in duplicate from each included SR: search strategy (searched databases, date restrictions, grey literature); test included/OCT models; inclusion/exclusion criteria; number of studies and number of OCT studies included; effect measure and risk-of-bias assessment tool used for primary studies; qualitative conclusions and implication for practice and research; and a summary of clinical pathway.

We followed guidance in the Cochrane Handbook for DTA reviews in assessing the definition and description of the clinical pathway—the context in which a test might be used.[16] According to these criteria, review authors should put the collected evidence in the context by describing whether the test of interest (index test) is proposed as a triage test, an add- on test, or a replacement for an existing test, and describing the consequences of misclassification, by linking testing to management actions and downstream patient's outcome.

Two investigators independently assessed the risk of bias of the included SRs using the Risk of Bias in Systematic Reviews (ROBIS) tool.[17] We resolved any disagreements by discussion or by referring to a third senior author. The ROBIS tool adopts a domain- based structure, with four domains covering key aspects of the review process (eligibility criteria, identification and selection of studies, data collection and study appraisal, and synthesis/ findings). Each domain has several signalling questions; answers to the signalling questions inform a judgement about concerns related to that domain. An overall judgement on the risk of bias of the review can then be formed on the basis of all domain-level judgments.

To avoid potential for bias in reviewing the author's own systematic review paper,[18] MM was not involved in the risk-of-bias assessment and data extraction of its own work, and these tasks were conducted by two other review authors (AM and RQ).

We planned to extract sensitivity and specificity as the preferred measures of accuracy and when not available we reported AUCs. One of the included reviews presented overall accuracy estimates which included the Stratus OCT. We computed pooled findings for SD-OCTs only by transforming AUCs (95% CIs) to ORs.[19]

We planned to obtain data for retinal nerve fibre layer (RNFL) and ganglion cell complex (GCC)/ganglion cell +inner plexiform layer (GCIPL) and at least one sector parameter—preferably the inferior sector which was shown to have better accuracy.[18]

RESULTS

Characteristics of included SRS

We screened 4072 title/abstracts from the CEV@US database and identified 31 reviews as potentially relevant. After the full-text review, we excluded 27 reviews (see figure 1 and online supplementary appendix 1 available as supplementary material, for reasons for exclusion). We included four SRs and their characteristics are described in table 1.

The four included SRs were published between 2015 and 2018. One review was published in *The Cochrane Library*[18] and other three were published in peer- review journals.[20–22] Kansal *et al*[22] focused on OCT alone, whereas the other three SRs evaluated OCT in comparison with other imaging devices including confocal scanning laser ophthalmoscopy, scanning laser polarimetry and frequency doubling technology. The number of primary studies included in each review ranged from 81 to 357, and the number of primary studies specially investigating SD-OCT included in each review ranged from 17 to 113. All reviews included studies considering manifest glaucoma as the target disease. Three reviews excluded screening and population- based studies, except for Ahmed *et al*.[20] Three reviews considered studies using both functional and/or structural damages as reference standard for glaucoma, while Ahmed *et al*[20] used only the visual field damages for this purpose.

Risk-of-bias assessment of the included SRs

A summary of risk-of- bias assessment of the included SRs is presented in table 2. Michelessi *et al*[18] and Fallon *et al*[21] were judged at low risk of bias for all ROBIS domains and Ahmed *et al*[20] and Kansal *et al*[22] were judged at high risk for at least one domain (full details are reported in online supplementary appendix 2, available as supplementary material).

Only Fallon *et al*[21] and Michelessi *et al*[18] had a protocol available. The lack of a predefined protocol increases the concern for bias in several aspects of the review process.[23] All SRs searched a range of databases but additional methods of searching, such as searching conference abstracts or citations in existing reviews, were lacking in Fallon *et al*[21] and Kansal *et al*.[22] Three SRs reported the full search strategy, while Ahmed *et al*[20] did not report any detail. Kansal *et al*[22] included only primary studies published in English, while Ahmed *et al*[20] did not report any details on the language accepted. All SRs used appropriate tools in conducting the risk-of-bias assessment of the included studies, such as Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[24] and QUADAS-2.[25] However, Ahmed *et al*[20] and Kansal *et al*[22] did not report the number of assessors involved in the risk-

of-bias assessment. Only Michelessi *et al*[18] published a guidance for QUADAS-2 assessment in the protocol. Ahmed *et al*[20] failed to report all results following the methods described. Finally, none of the SRs conducted any sensitivity analysis although Michelessi *et al*[18] addressed the reason for not conducting such analysis.

Clinical pathway assessment

Ahmed *et al*,[20] Fallon *et al*[21] and Kansal *et al*[22] limited the clinical pathway description in specifying the target condition to be considered, the current gold standard for glaucoma diagnosis, and how the new tests might benefit patients (eg, by anticipating the diagnosis at an early stage of the disease course). Michelessi *et al*[18] explicitly reported the diagnostic care pathway by describing how patients might present, how the index test is intended to be used and the downstream consequences of the index test results on the patient management care.

Methods used in meta-analyses

Michelessi *et al*[18] and Fallon *et al*[21] used appropriate meta-analysis methods, for example, bivariate or hierarchical summary receiver-operating characteristic methods. Ahmed *et al*[20] seemed to have applied univariate random-effect meta-analyses of all accuracy measures as they reported I^2 for sensitivity and specificity, despite a lack of an explicit description. Kansal *et al*[22] conducted univariate analyses of AUCs, which is statistically correct but not the standard method of collecting and presenting accuracy data to clinical users.

Quantitative estimates of accuracy and investigation of heterogeneity

No SRs reported the thickness threshold at which sensitivity and specificity pairs were calculated because they were not available in included studies, which mostly reported sensitivity at fixed specificity levels, such as 0.80 or 0.90. In fact, accuracy estimates cannot be used in clinical practice if a classification criterion, such as a thickness cut-off, is not reported.

Three SRs[18 20 21] reported at least one sensitivity and specificity pair as seen in table 3. For average RNFL thickness, average GCC/GCIPL thickness, or GGC thickness C/D vertical ratio, sensitivities were generally between 70% and 80% at high specificities slightly over 90%.

Regarding investigation of heterogeneity, Michelessi *et al*[18] found the sensitivity to be better for more severe glaucoma (average MD>−6 dB at study level) and Fallon *et al*[21] reached similar conclusions analysing variation in DORs. They also assessed between-study heterogeneity regarding sensitivity in ROC and forest plots and reported this was large.

Qualitative interpretation of results

Michelessi *et al*[18] concluded that 'the accuracy of imaging tests for detecting manifest glaucoma was variable across studies, but overall similar for different devices' and that 'accuracy may have been overestimated due to the case-control design, which is a serious limitation of the current evidence base'. They did not qualify the diagnostic performance of OCT but suggested that 'the implications of using our estimates for clinical decision-making

is highly dependent on the care pathway and the diagnostic alternatives available, which goes beyond the scope of this review'. Kansal et al[22] compared OCT devices and concluded that all OCTs 'demonstrated good diagnostic accuracy in their ability to differentiate patients with glaucoma from normal controls' and that 'this ability increased with the severity of the glaucoma'. Ahmed et al[20] and Fallon et al[21] focused their conclusions on the performance of OCT relative to other devices.

Only Michelessi et al[18] used Grading of Recommendations, Assessment, Development, and Evaluation methodology to present a summary of findings table—a succinct, transparent and informative summarisation, which presents both the magnitude of effects and the quality of evidence.

DISCUSSION

We identified four DTA SRs on OCT for detecting manifest glaucoma, all of which were published in the last 5 years.

With the increasing number of published SRs and the adoption of SRs to support evidence-based guidelines, the need for rigorous SRs has become critical. In fact, as for individual studies, SRs are also prone to bias and errors which can lead to misleading conclusions.[26] We found that the methods of two reviews had at least one domain at high risk of bias, and were judged to be at high risk overall, whereas two other reviews were at low risk of bias for all domains. The areas of particular concerns were study eligibility criteria and data collection/study appraisal methods. In addition, two reviews did not report having a protocol.

Different inclusion criteria among the four reviews regarding index test (eg, type of OCT or OCT parameter used) and reference standard (perimetry and/or ONH assessment), resulted in a very different number of included studies and also in different parameters' accuracy estimated. With regard to the reference standard, while Ahmed et al[20] included only studies using functional criteria, other three reviews included both functional and/or structural reference tests, which were variably mixed. For example, the included studies using a combination of functional and structural criteria were 30.9%, 73.3% and 68.8%, respectively, for Fallon et al,[21] Kansal et al[22] and Michelessi et al.[18] With such discrepancies, it is difficult to draw overall conclusions on the effect of different reference standard on OCT accuracy and it further highlights the need for consensus strategies aiming to standardise diagnostic testing studies in glaucoma.

Our findings that DTA reviews on OCT for diagnosing glaucoma have variable design and quality are common in other health areas[27] and reflect that of primary DTA studies in this field. Methodological studies have found that studies on imaging tests for diagnosing glaucoma are poorly reported and lack a clear specification of the clinical pathway (eg, by specifying how the patient might present or be referred, any prior testing, the intended use of the test and the potential downstream consequences of different test results).[28] This information allows readers to put the effect of estimated accuracy in the appropriate context,

and envision how the test results will impact the treatment strategies and patient's outcome.[29]

We found that the accuracy of several OCT parameters was highly variable among the included SRs, but ultimately suboptimal. Overall, sensitivity was only moderate at high specificity levels. The optimal balance between sensitivity and specificity is highly dependent on the setting. High specificity is usually desirable for a screening/case detection, in order to minimise the burden of false positive referrals in a low-prevalence population. On the contrary, an enriched population is typical of studies on referrals to secondary care, meaning that a high sensitivity is desirable to avoid missing many patients with glaucoma. The thickness threshold at which sensitivity and specificity pairs were calculated was not reported in the included SRs, as it was not available in primary DTA studies. In fact, without an explicit classification criterion (eg, thickness cut-off), accuracy estimates cannot be used in clinical practice.

In addition, accuracy may have been overestimated in primary DTA studies. A perfect reference standard for glaucoma diagnosis is still lacking and as a consequence, glaucoma suspect is commonly not involved in the DTA studies evaluating OCT for glaucoma detection, which often adopt a case-control design. Such design includes healthy controls separately from glaucoma cases; thus, it can magnify accuracy and reduce the applicability of the results to daily practice.[30]

A large study on consecutively referred glaucoma suspects, which was not included in any of our SRs because of its recent publication date, found that sensitivity was 0.77 and specificity was 0.79 for diagnosing manifest glaucoma using the RNFL classification of outside normal limits.[31] A secondary publication of the same study found that the best parameter was inferior RNFL thickness and sensitivity was 0.55 at high specificity (0.95) and specificity was 0.39 at high sensitivity (0.95).[32] This study, which avoided a case-control design, suggests that the accuracy of OCT for diagnosing manifest glaucoma may be actually worse than that found in our included SRs.

Our focus was on SRs that investigated the overall accuracy of several OCT parameters compared with a reference standard, and we did not include the review published by Oddone *et al*[3] in 2016 which found that accuracy of macular parameters was similar, but slightly lower, to that of RNFL parameters. This review used the same search strategies developed by the Cochrane Eyes and Vision Group for Michelessi 2015,[18] included all studies already included in that review and restricted the analysis to a direct comparison between macular and RNFL parameters, thus not fitting the aim of this overview.

Quantitative imaging tests, such as OCT, are currently recommended by international guidelines as an add-on test to provide supplemental information to assist clinicians in glaucoma diagnosis.[45] Both glaucoma suspects and patients with glaucoma with early and moderate disease are suggested to be examined with OCT at baseline, but it is mandatory that the OCT results must be interpreted in the context of all available clinical data.[5]

Our overview confirms that current evidence does not support the use of OCT as a stand-alone test for diagnosing manifest glaucoma. Ultimately, the consequences of any estimate

of OCT accuracy on clinical decision-making depend on the specific pathway of care in which the test is to be used. Framing the clinical question in a specific setting and pathway is particularly important in DTA research, since the impact of test-guided decisions on patient's outcome is in essence indirect in such studies.

Previous methodological studies have investigated the agreement between different SRs on the same interventions in similar populations.[33][34] Siontis *et al* suggested that while some independent replication of meta-analyses by different teams is possibly useful, the overall picture suggests that there is a waste of efforts with many topics covered by multiple overlapping SRs and meta-analyses.[35] This overlap and redundancy apply to more recent review methods, including network meta-analyses.[36] Our findings also reinforce the need that primary DTA study authors and DTA SR authors should adhere to reporting recommendations such as Standards for Reporting of Diagnostic Accuracy Studies and Preferred Reporting Items for Systematic Reviews andMeta-Analyses-DTA, respectively.[37][38] Finally, overinterpretation, or spin, of accuracy estimates should be avoided.[39]

In conclusion, we have provided an overview of SRs on OCT for diagnosing manifest glaucoma. The diagnostic accuracy was suboptimal and potentially overestimated due to the case-control design of primary studies, partially supporting the role of OCT as add-on test in detecting glaucoma. Two reviews were at high risk of bias, used inappropriate methods and did not clearly report sensitivity and specificity for key OCT parameters. Our overview suggests that the methodological quality of both primary and secondary DTA research on glaucoma is in need of improvement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

1. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. Science 1991;254:1178–81. [PubMed: 1957169]

2. Sung KR, Kim JS, Wollstein G, et al. Imaging of the retinal nerve fibre layer with spectral domain optical coherence tomography for glaucoma diagnosis. Br J Ophthalmol 2011;95:909–14. [PubMed: 21030413]

3. Oddone F, Lucenteforte E, Michelessi M, et al. Macular versus retinal nerve fiber layer parameters for diagnosing manifest glaucoma: a systematic review of diagnostic accuracy studies. Ophthalmology 2016;123:939–49. [PubMed: 26891880]

4. Prum BE, Rosenberg LF, Gedde SJ, et al. Primary Open-Angle Glaucoma Preferred Practice Pattern($^{®}$) Guidelines. Ophthalmology 2016;123:P41–111. [PubMed: 26581556]

5. European Glaucoma Society. European glaucoma Society terminology and guidelines for glaucoma. 4th Edn. Br J Ophthalmol, 2017: 1–72.

6. Banister K, Boachie C, Bourne R, et al. Can automated imaging for optic disc and retinal nerve fiber layer analysis aid glaucoma detection? Ophthalmology 2016;123:930–8. [PubMed: 27016459]

7. Dabasia PL, Fidalgo BR, Edgar DF, et al. Diagnostic accuracy of technologies for glaucoma case-finding in a community setting. Ophthalmology 2015;122:2407–15. [PubMed: 26411836]

8. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–2. [PubMed: 8555924]

9. Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Board on Health Care Services, Institute of Medicine. Clinical practice guidelines we can trust. Washington, DC: National Academies Press, 2011.

10. Golozar A, Chen Y, Lindsley K, et al. Identification and description of reliable evidence for 2016 American Academy of ophthalmology preferred practice pattern guidelines for cataract in the adult eye. JAMA Ophthalmol 2018;136:514–23. [PubMed: 29800249]

11. Saldanha IJ, Lindsley KB, Lum F, et al. Reliability of the evidence addressing treatment of corneal diseases: a summary of systematic reviews. JAMA Ophthalmol 2019;137:775–85. [PubMed: 31070698]

12. Le JT, Qureshi R, Twose C, et al. Evaluation of systematic reviews of interventions for retina and vitreous conditions. JAMA Ophthalmol 2019. doi:10.1001/jamaophthalmol.2019.4016. [Epub ahead of print: 10 Oct 2019].

13. Mayo- Wilson E, Ng SM, Chuck RS, et al. The quality of systematic reviews about interventions for refractive error can be improved: a review of systematic reviews. BMC Ophthalmol 2017;17:164. [PubMed: 28870179]

14. Institute of Medicine. Finding what works in health care: standards for systematic reviews. Washington, DC: National Academies Press, 2011.

15. Li T, Ervin A-M, Scherer R, et al. Setting priorities for comparative effectiveness research: a case study using primary open- angle glaucoma. Ophthalmology 2010;117:1937–45. [PubMed: 20800896]

16. Deeks JJ, Wisniewski S, Davenport C. Guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for systematic reviews of diagnostic test accuracy version 1.0.0, 2013. http://srdta.cochrane.org/

17. Whiting P, Savovi J, Higgins JPT, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225–34. [PubMed: 26092286]

18. Michelessi M, Lucenteforte E, Oddone F, et al. Optic nerve head and fibre layer imaging for diagnosing glaucoma. Cochrane Database Syst Rev 2015;11:CD008803.

19. Salgado JF. Transforming the Area under the Normal Curve (AUC) into Cohen's d, Pearson's r pb, Odds- Ratio, and Natural Log Odds-Ratio: Two Conversion Tables. Eur J Psychol Appl Leg Context 2018;10:35–47.

20. Ahmed S, Khan Z, Si F, et al. Summary of glaucoma diagnostic testing accuracy: an evidence-based meta-analysis. J Clin Med Res 2016;8:641–9. [PubMed: 27540437]

21. Fallon M, Valero O, Pazos M, et al. Diagnostic accuracy of imaging devices in glaucoma: a meta-analysis. Surv Ophthalmol 2017;62:446–61. [PubMed: 28093287]

22. Kansal V, Armstrong JJ, Pintwala R, et al. Optical coherence tomography for glaucoma diagnosis: an evidence based meta-analysis. PLoS One 2018;13:e0190621.

23. Straus S, Moher D. Registering systematic reviews. CMAJ 2010;182:13–14. [PubMed: 19620270]

24. Whiting P, Rutjes AWS, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25. [PubMed: 14606960]

25. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36. [PubMed: 22007046]

26. Ioannidis JPA. The mass production of redundant, misleading, and Conflicted systematic reviews and meta- analyses. Milbank Q 2016;94:485–514. [PubMed: 27620683]

27. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86–9. [PubMed: 19525005]

28. Michelessi M, Lucenteforte E, Miele A, et al. Diagnostic accuracy research in glaucoma is still incompletely reported: an application of standards for reporting of diagnostic accuracy studies (STARD) 2015. PLoS One 2017;12:e0189716.

29. Gopalakrishna G, Langendam MW, Scholten RJPM, et al. Defining the clinical pathway in Cochrane diagnostic test accuracy reviews. BMC Med Res Methodol 2016;16:153. [PubMed: 27832765]

30. Rutjes AWS, Reitsma JB, Vandenbroucke JP, et al. Case-Control and two-gate designs in diagnostic accuracy studies. Clin Chem 2005;51:1335–41. [PubMed: 15961549]

31. Azuara- Blanco A, Banister K, Boachie C, et al. Automated imaging technologies for the diagnosis of glaucoma: a comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost- effectiveness (gate study). Health Technol Assess 2016;20:1–168.

32. Virgili G, Michelessi M, Cook J, et al. Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the gate study. Br J Ophthalmol 2018;102:604–10. [PubMed: 28855198]

33. Lucenteforte E, Moja L, Pecoraro V, et al. Discordances originated by multiple meta- analyses on interventions for myocardial infarction: a systematic review. J Clin Epidemiol 2015;68:246–56. [PubMed: 25533151]

34. Moja L, Fernandez del Rio MP, Banzi R, et al. Multiple systematic reviews: methods for assessing discordances of results. Intern Emerg Med 2012;7:563–8. [PubMed: 22941412]

35. Siontis KC, Ioannidis JPA, Replication IJPA. Replication, duplication, and waste in a quarter million systematic reviews and meta- analyses. Circ Cardiovasc Qual Outcomes 2018;11:e005212.

36. Naudet F, Schuit E, Ioannidis JPA. Overlapping network meta- analyses on the same topic: survey of published studies. Int J Epidemiol 2017;46:1999–2008. [PubMed: 29040566]

37. Johnson ZK, Siddiqui MAR, Azuara-Blanco A. The quality of reporting of diagnostic accuracy studies of optical coherence tomography in glaucoma. Ophthalmology 2007;114:1607–12. [PubMed: 17434589]

38. Salameh J-P, McInnes MDF, Moher D, et al. Completeness of reporting of systematic reviews of diagnostic test accuracy based on the PRISMA- DTA reporting guideline. Clin Chem 2019;65:291–301. [PubMed: 30237150]

39. McGrath TA, McInnes MDF, van Es N, et al. Overinterpretation of Research Findings: Evidence of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies. Clin Chem 2017;63:1353–62. [PubMed: 28606911]
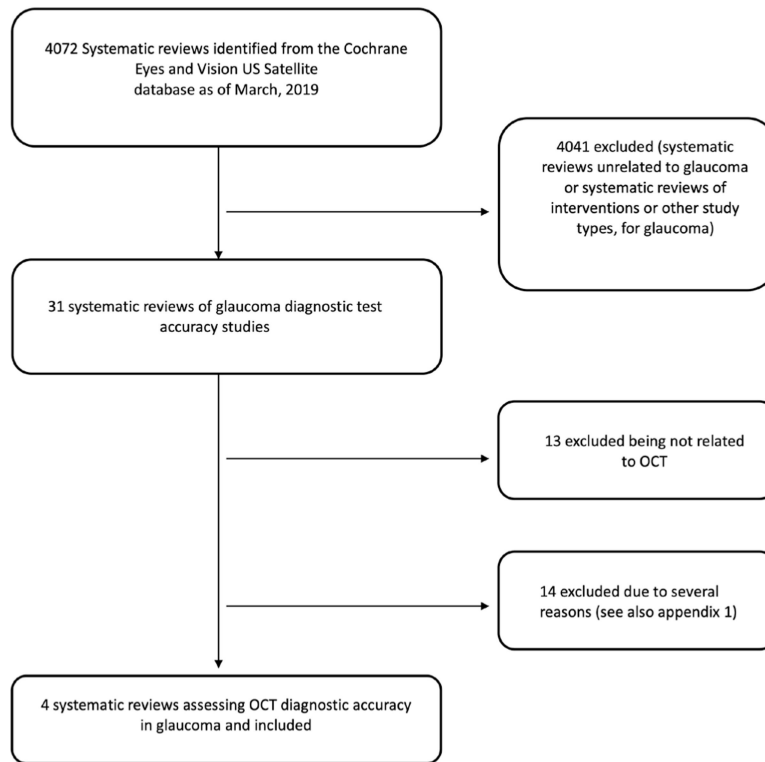
Figure 1.
Flow chart showing the identification of systematic reviews of optical coherence tomography (OCT) diagnostic accuracy for detecting glaucoma.

Table 1

Characteristics of the included systematic reviews

	Ahmed et al[20]	Fallon et al[21]	Kansal et al[22]	Michelessi et al[28]
Clinical pathway discussion	Target condition and the current gold standard for glaucoma diagnosis were described. Five glaucoma technologies compared with the gold standard of white on white perimetry for glaucoma detection. No further info	Target condition defined in term of frequency, severity and prognosis. The authors reported how glaucoma is currently identified and how the imaging devices may play a role in glaucoma diagnosis, with the aim to identify patients with glaucoma early in the disease course. No further info	Target condition defined in terms of frequency and severity. The currently accepted gold standard for glaucoma diagnosis is discussed. Test variations in terms of different manufacturers, which are playing an increasing role in glaucoma diagnosis. No further info	Evaluated the accuracy of OCT, HRT or the GDx for diagnosing glaucoma by using these devices as add-on test, in people who have already been tested by means of clinical examination at primary care level, including ONH clinical assessment, IOP measurement and even visual field testing. Target condition, index test and clinical pathway categories, almost completely described
Databases searched and date restrictions	5 (from 1993 to February 2015)	4 (from January 2004 to February 2015)	6 (up to February 2017)	8 (up to February 2015)
Grey literature	Websites of HTA agencies, professional associations. Google search for web based, hand searching the bibliographies and abstracts of key papers and conference proceedings, contacts with appropriate experts and agencies	None	None	Reference lists of the included studies
Tests included and OCT models	OCT (time domain (TD), spectral domain (SD)), confocal scanning laser ophthalmoscopy, scanning laser polarimetry or frequency doubling technology, blue/yellow perimetry	OCT SD, confocal scanning laser ophthalmoscopy, scanning laser polarimetry	OCT (SD,TD)	OCT (SD), confocal scanning laser ophthalmoscopy, scanning laser polarimetiy
Inclusion/exclusion criteria	Inclusion criteria: any study evaluating testing outcomes for open-angle glaucoma. Any geographical study type (single vs multicentre); any sampling strategy; all ethnic background and studies conducted in any country. Any study bases (population or clinic-based samples). Age >18 years old Exclusion criteria:<20 participants	Inclusion criteria: any study evaluating the ability of image devices (HRT3, GDx, and SD-OCT) to differentiate between normal and glaucoma in which index and reference test were used in the same population. Prospective design. Observational, descriptive (base population, cross sectional), analytical studies (case control, cohort), and experimental studies (clinical trials), >50 participants. Study reporting sensitivity and specificity of the test(s) and quality criteria. Study published in English, Spanish or French. Screening studies were included only in the qualitative analysis Exclusion criteria: study investigating specific subgroups of subjects, such as children, young individuals or subjects with ophtha l mo logical or neurological diseases	Inclusion criteria: study evaluating the diagnostic accuracy of OCT (Stratus OCT, Cirrus OCT, Spectralis OCT, RTVue, and 3D-OCT Topcon) for glaucoma detection. Human, clinical studies published in English-language. Age>18 years. Any patient ethnicity, or county. Studies reporting area under the ROC curve. Exclusion criteria: not reporting SE or CIs, duplicate manuscripts, non-diagnostic studies, paediatric patients, no control group	Inclusion criteria: prospective and retrospective cohort studies and case-control studies evaluating the accuracy of OCT, HRT or the GDx for diagnosing glaucoma. Both single studies assessing each imaging method and comparative studies assessing more than one imaging method in the same patient population. Studies providing data to allow calculation of sensitivity and specificity estimates Exclusion criteria: screening study

	Ahmed et al[20]	Fallon et al[21]	Kansal et al[22]	Michelessi et al[28]
Reference standard	White on white automated perimetry	The assessment of the structure and/or visual function	White on white automated perimetry and/or optic disc appearance (clinically or by photograph)	Any diagnosis of glaucoma given by the study investigators by using both optic disc and/or visual field damage
No studies (eligible/included)	2474/357	8028/81	2264/150	9332/106
No OCT studies (SD,TD)	84 (17 SD, 67TD)	54(54SD)	150 (113 SD, 42 TD)	63 (63 SD)
Effect measure	Sensitivity, specificity, DOR, AUC	Sensitivity, specificity, DOR	AUC	Sensitivity, specificity, relative DOR
Risk-of-bias assessment	QUADAS	QUADAS-2	QUADAS-2	QUADAS-2
Qualitative conclusions	OCT has the highest glaucoma diagnostic accuracy followed by GDx and then HRT	All three instruments showed good diagnostic accuracy, and OCT obtained the best value of DORs	The currently available OCT devices (Zeiss Cirrus, Zeiss Stratus, Heidelberg Spectralis, Optovue RTVue,Topcon 3D-OCT) demonstrated good diagnostic ability in differentiating normal from glaucoma. This ability increased with the severity of the glaucoma and no device-related differences in diagnostic accuracy emerged	The accuracy of imaging tests for detecting manifest glaucoma was variable across studies, but overall similar for different devices. Accuracy may have been overestimated due to the case-control design, which is a serious limitation of the current evidence base
Impact on practice and research	Work needs to be done by glaucoma content experts to create a more homogenous consensus regarding how to use the new technologies and to agree on cutoffs	The best performing algorithm/parametes were MRA for HRT, NFI for GDx and RNFL and GCC thickness for OCT	The diagnostic capacity of RNFL is similar to segmented macular regions (GCIPL, GCC), and better than total macular thickness. As OCT technology continues to evolve at a faster pace than functional assessments of optic nerve health, future studies will be needed to fully understand its role in glaucoma management	The findings of this review could be used in an add-on setting which could be a primary care, or a triage setting when somebody has already been referred from primary care to secondary care as suspect glaucoma and needs triage by a non-glaucoma specialist. Further case-control studies are not useful in this research fie Id. There is room for improving diagnostic accuracy studies evaluating OCT devices in glaucoma

AUC, area under the ROC; DOR, diagnostic OR; GCC, ganglion cell complex; GCIPL, ganglion cell layer +inner plexiform layer; GDx, Scanning laser polarimetry; HRT, Heidelberg Retina Tomograph; HTA, health technology assessment; IOP, intraocular pressure; MRA, Moorfield's regression analysis; NFI, nerve fibre indicator; OCT, optical coherence tomography; ONH, optic nerve head; QUADAS, Quality Assessment of Diagnostic Accuracy Studies; RNFL, retinal nerve fibre layer; ROC, receiver operating characteristic curve.

Table 2

Summary of risk-of-bias assessment of included systematic reviews

Review	Domain				
	#1: study eligibility criteria	#2: identification and selection of studies	#3: data collection and study appraisal	#4: synthesis and findings	overall risk of bias of the review
Ahmed *et al*[20]	High	Unclear	Unclear	High	High
Fallon *et al*[21]	Low	Low	Low	Low	Low
Kansal *et al*[22]	High	Low	Unclear	Low	High
Michelessi *et al*[28] 2015	Low	Low	Low	Low	Low

Table 3

Diagnostic accuracy parameters as estimated in the included systematic reviews

	RNFL parameters	GCC/GCIPL parameters	other parameters
Ahmed *et al*[20]			▲ *Unqualified parameter* (n. 17) Se: 83.3 (77.2 to 88) Sp: 91.6 (87.8 to 94.2)
Fallon *et al*[21]	▲ *RNFL average* (n. 50) Se: 0.78 (0.74 to 0.83) Sp: 0.91 (0.88 to 0.93)	▲ *GCC average* (n. 14) Se: 0.79 (0.71 to 0.88) Sp: 0.88 (0.82 to 0.94)	▲ *C/D vertical ratio* (n. 14) Se: 0.78 (0.74 to 0.83) Sp: 0.91 (0.88 to 0.93)
Kansal *et al*[22]	▲ *RNFL average* (n. 119) AUC: 0.901 (0.877 to 0.917) ▲ *RNFL inferior* (n. 95) AUC: 0.905 (0.869 to 0.925)	▲ *GCIPL average* (n. 28) AUC: 0.858 (0.835 to 0.880) ▲ *GCC average* (n. 39) AUC: 0.885 (0.869 to 0.901)	
Michelessi *et al*[28]	▲ *RNFL average* (n. 57) Se: 0.69 (0.63 to 0.73) Sp: 0.94 (0.93 to 0.95) ▲ *RNFL inferior sector* (n. 45) Se: 0.72 (0.65 to 0.77) Sp: 0.93 (0.92 to 0.95)	▲ *GCC/GCIPL average* (n. 32) Se: 0.63 (0.57 to 0.70) Sp: 0.93 (0.91 to 0.94)	▲ *C/D vertical ratio* (n.15) Se: 0.72 (0.60 to 0.81) Sp: 0.94 (0.90 to 0.94)

AUC, area under the receiver operating characteristic curve; C/D, cup/disc; GCC, ganglion cell complex; GCIPL, ganglion cell layer +Inner plexiform layer; RNFL, retinal nerve fibre layer; Se, sensitivity; Sp, specificity.