Methodology and reporting of diagnostic accuracy studies of automated perimetry in glaucoma: evaluation using a standardised approach

Running head: Quality of reporting of diagnostic accuracy studies

Authors' names and institutional affiliations: Bruno M R Fidalgo, David P Crabb, John G Lawrenson

Division of Optometry and Visual Science, City University London, London EC1V 0HB, UK

Corresponding Author:

John G. Lawrenson

E-mail address: J.G.Lawrenson@city.ac.uk

Keywords: perimetry, glaucoma, STARD, QUADAS, methodological quality

Abstract

Purpose: To evaluate methodological and reporting quality of diagnostic accuracy studies of perimetry in glaucoma and to determine whether there had been any improvement since the publication of the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines.

Methods: A systematic review of English language articles published between 1993 and 2013 reporting the diagnostic accuracy of perimetry in glaucoma. Articles were appraised for methodological quality using the 14-item Quality assessment tool for diagnostic accuracy studies (QUADAS) and evaluated for quality of reporting by applying the STARD checklist.

Results: Fifty-eight articles were appraised. Overall methodological quality of these studies was moderate with a median number of QUADAS items rated as 'yes' equal to 9 (out of a maximum of 14) (IQR 7-10). The studies were often poorly reported; median score of STARD items fully reported was 11 out of 25 (IQR 10-14). A comparison of the studies published in 10-year periods before and after the publication of the STARD checklist in 2003 found quality of reporting had not substantially improved.

Conclusions: Methodological and reporting quality of diagnostic accuracy studies of perimetry is sub-optimal and appears not to have improved substantially following the development of the STARD reporting guidance. This observation is consistent with previous studies in ophthalmology and in other medical specialities.

Introduction

It is estimated that at least half of all glaucoma cases are undetected in developed countries.[1] Imperfect diagnostic accuracy of current tests used to detect cases is one reason among others[2] for this.

There are a number of tests used for the detection and management of glaucoma, which involve either the assessment of structural changes to the optic nerve head, detection of functional visual loss by testing the visual field (VF), or the measurement of the intraocular pressure (IOP).

Standard automated perimetry (SAP) is the most frequently performed vision function test in patients with glaucoma or suspect glaucoma, although a variety of newer technologies have been introduced.[3] The aim of testing the VF is to detect functional changes in patients with suspected disease and to monitor the progression of disease in those with established glaucoma.

Diagnostic accuracy studies are used to certify new tests before they are brought into clinical practice. Most diagnostic accuracy studies report sensitivity and specificity, positive and negative predictive values, or receiver-operator characteristics (ROC) curves as measures of diagnostic performance.[4] This information allows a clinician to make an informed decision regarding the potential value of the new test. However, if those studies are not conducted properly or their reporting is inconsistent there is potential for bias (internal validity) or difficulty in estimating the generalisability of the study findings (external validity). Furthermore, overstated or understated results could lead to the premature adoption of a poorly performing diagnostic test or delay the adoption of a high quality test respectively. Diagnostic accuracy studies in glaucoma present a particular problem since there is no universally agreed reference standard for the disease. Ophthalmologist diagnosed glaucoma following a detailed ophthalmic examination to detect structural and/or functional damage is the accepted reference standard in clinical practice.[5] However, this may lead to variability in test performance as some patients may show evidence of structural damage without a detectable visual field abnormality, whilst others may have glaucomatous visual field loss without a detectable structural abnormality. Furthermore, studies evaluating diagnostic accuracy typically use a case-control design where glaucoma patients (cases) are compared with normal subjects (controls). The adoption of stringent inclusion and exclusion criteria in these studies may mean that the sample is unrepresentative of the population on which the test would be used in practice.

Methodological rigour and accuracy of reporting are essential pre-requisites for the evaluation of diagnostic test performance. In 2003, a quality assessment tool was introduced for evaluating the methodological quality of diagnostic accuracy studies (QUADAS: Quality Assessment of Studies of Diagnostic Accuracy).[6] In the same year, a checklist was published for improving the quality of reporting of these studies (STARD: Standards for the Reporting of Diagnostic accuracy studies).[7]

The QUADAS tool was designed for use by those undertaking systematic reviews of diagnostic accuracy studies. It consists of 14 items each of which is rated as "yes", "no", or "unclear". The items cover the major sources of bias and to a lesser extent, the quality of reporting.

The purpose of the STARD initiative was to develop a checklist to improve the quality of reporting of studies of diagnostic accuracy. The tool uses a list of 25 items covering the main sections of the paper and promoted the use of a flow diagram to present the study design and report the exact number of patients at each stage of the study.

The aim of this study is to assess the methodological quality of diagnostic accuracy studies using a variety of perimetric tests in glaucoma, using the QUADAS tool and evaluate the accuracy and completeness of reporting of these studies based on the STARD checklist. We also sought to determine whether there had been any improvement in study quality and reporting since the publication of these standards.

Methods

Search strategy

We used the OVID platform to search relevant electronic databases (MEDLINE, EMBASE, and Global Health) to identify diagnostic accuracy studies of perimetry published over a 20-year period between January 1993 and August 2013.[8] The keywords and search terms used are described in Table 1. The search was limited to publications in the English language and studies on human subjects.

"Perimetry"* or "Perimeter"* or "Standard Automated Perimetry" or "SAP" or "visual field" or "Motion displacement test" or "frequency doubling technology" or "Flicker Defined Form" or "High pass resolution" or "HRP" or "OKP" or "Humphrey" or "Henson" or "Octopus" or "Heidelberg" or "Dicon" or "Medmont" or "rarebit" or "Ophthimus".
"Glaucoma"**.
"Diagnostic Accuracy" or "diagnostic performance" or "precision" or "ROC" or "Receiver operating characteristic" or "Sensitivity and specificity" or "diagnostic odds ratio" or "DOR" or "area under the receiver operating characteristic curve" or "likelihood ratio". "Early diagnosis" or "differentiate" or "identify" or "detect" or "diagnosis" or "screening" or "case finding".

Table 1. Key words used for the literature search

** - Since the word perimeter or perimetry appears in most of the tests and synonyms of the test we assumed that those terms would encapsulate most forms of the test.*
***- Synonyms for the disease all tend to contain the word 'glaucoma' we assumed that the term "glaucoma" would encapsulate all forms of the disease.*

Study selection

The titles and abstracts of all reports identified by the search strategy were screened by a single reviewer (BF). Full text copies of all studies deemed to be potentially relevant were obtained and assessed for inclusion. Reasons for exclusion at this stage were tabulated and any uncertainty in eligibility was resolved by discussion with the other members of the review team. A flow diagram of the selection process is shown in Figure 1.

Assessment of methodological quality (QUADAS) and quality of reporting (STARD)

A quality assessment was then performed on all included studies by a single reviewer (BF) using both quality assessment tools (QUADAS and STARD)[6,7] using custom checklists developed in Microsoft Excel to score each quality item. A written justification for each decision was also included in the spreadsheet. For each of the included studies the diagnostic reference standard for glaucoma was documented (Table 2).

The QUADAS tool (Table 3) is structured as a list of 14 items which are answered as 'yes', 'no' or 'unclear'. The items cover the main sources of bias, including: spectrum bias, verification bias, disease progression bias as well as items asking about the execution of index and reference tests. The STARD checklist (Table 4) is divided into 5 sections, covering the main sections of the paper: title/abstract/keywords, introduction, methods, results and conclusion. The majority of items in the checklist are scored as 'reported', 'not reported' or 'partially reported'.

Two reviewers (JL and DC) independently assessed the quality of a random 20% sample of included studies.

Journal guidance regarding the use of STARD

We checked the 'Instructions to Authors' of the journals in which the articles were published for any advice or reference to the use of the STARD guidelines for reporting diagnostic accuracy studies.

Data synthesis and statistical analysis

Descriptive statistics were used to summarise the number and proportion of items that met the QUADAS and STARD criteria. Studies were stratified according to two time periods, 1993-2003 and 2004-2013, corresponding to periods before and after publication of the quality assessment tools to determine the impact of their publication on quality.

An inter-rater reliability analysis using the weighted Kappa statistic was performed to determine consistency among reviewers.[9]

Results

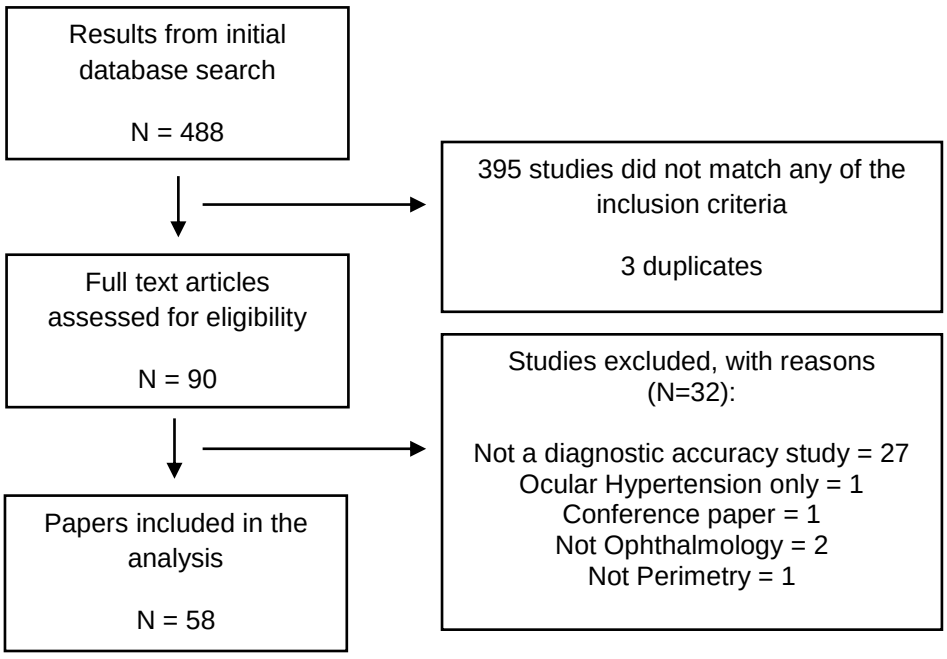

Figure 1. Flow diagram showing selection process.

The initial search identified a total of 488 articles. Subsequent title and abstract screening excluded 398 articles that failed to meet any of the inclusion criteria. Following the assessment of the full text of the remaining 90 articles, a further 32 were excluded. Therefore, 58 studies were eligible for inclusion in the review.[10-67] Inter-rater reliability, estimating consistency amongst the reviewers, as determined by weighted kappa for QUADAS and STARD was 0.70 and 0.81 respectively. The majority of studies used ophthalmologist diagnosis based on combined optic disc assessment, IOP and SAP as the reference standard for glaucoma diagnosis.

Reference Standard	Number of studies (%) Total = 58
Optic disc examination (ODE)	3 (5.2)
ODE+ IOP + SAP + Gonioscopy	11 (19)
ODE + IOP + SAP	11 (19)
ODE + IOP + SAP + (OCT or HRT or SLO or GDx)	5 (8.6)
ODE + IOP	3 (5.2)
ODE + SAP	8 (13.8)
SAP	3 (5.2)
ODE + Gonioscopy + IOP	4 (6.9)
ODE + Gonioscopy	1 (1.7)
IOP + SAP	1 (1.7)
Fundus Stereo photographs + SAP + IOP	1 (1.7)
Fundus Stereo photographs + SAP + IOP + SLO	1 (1.7)
Fundus Stereo photographs + SAP	1 (1.7)
Fundus Stereo photographs + IOP	1 (1.7)
Fundus Stereo photographs	2 (3.4)
Optic disc photography	1 (1.7)
Manual Suprathreshold Static Perimetry	1 (1.7)

Table 2. Reference standard for glaucoma diagnosis in the included studies.

Tables 3 and 4 summarise the results for each item using the QUADAS and STARD assessment tools respectively.

Of the 58 articles, the number of QUADAS items with a ‘yes’ response ranged from 2 to 14 with a median score of 9. For the papers published before 2003 (n=22) and after 2003 (n=36) this median score was 8 (interquartile range [IQR] 5-9) and 10 (IQR 8-10) respectively. Figure 2 shows the QUADAS assessment of each article arranged chronologically by year of publication.

	Item	1993 to 2003 (%) N = 22	2004 to 2013 (%) N = 36	Total (%) N = 58
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	9 (41)	16 (44)	25 (43)
2	Were selection criteria clearly described?	18 (82)	34 (94)	52 (90)
3	Is the reference standard likely to correctly classify the target condition?	17 (77)	29 (81)	46 (79)
4	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	2 (9)	21 (58)	23 (40)
5	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	12 (55)	32 (89)	44 (76)
6	Did patients receive the same reference standard regardless of the index test result?	13 (59)	25 (69)	38 (66)
7	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	21 (95)	33 (92)	54 (93)
8	Was the execution of the index test described in sufficient detail to permit replication of the test?	15 (68)	26 (72)	41 (71)
9	Was the execution of the reference standard described in sufficient detail to permit its replication?	7 (32)	24 (67)	31 (53)
10	Were the index test results interpreted without knowledge of the results of the reference standard?	4 (18)	12 (33)	16 (28)
11	Were the reference standard results interpreted without knowledge of the results of the index test?	8 (36)	19 (53)	27 (47)
12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	17 (77)	29 (81)	46 (79)
13	Were uninterpretable/ intermediate test results reported?	6 (27)	22 (61)	28 (48)
14	Were withdrawals from the study explained?	7 (32)	11 (31)	18 (31)

Table 3. QUADAS item (scored as “Yes”)

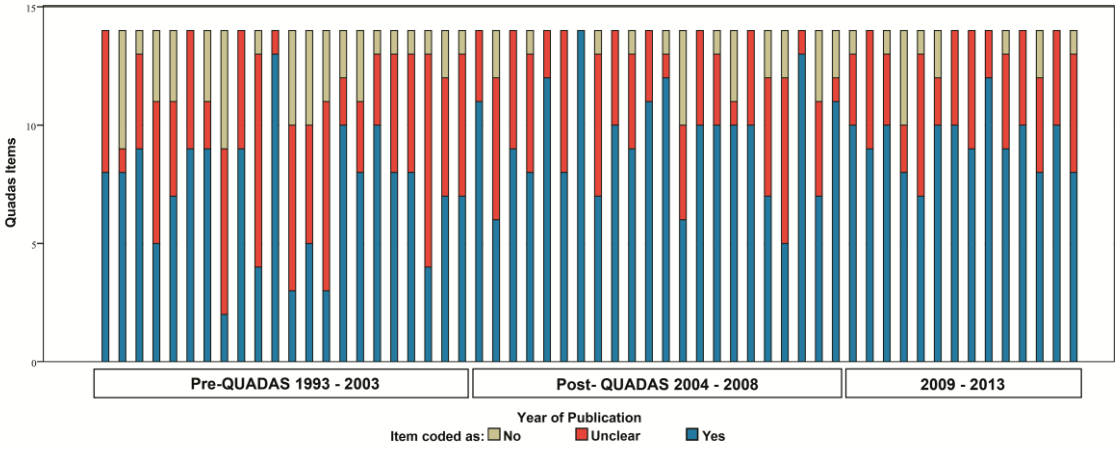

Figure 2. Stacked bar chart showing the QUADAS items reported and year of publication

In STARD, item 2, 4 and 25 were fully reported by all 58 articles. Item 20 was not scored due to the non-invasive nature of perimetry. No article fulfilled all items and only 41% reported in full more than half of the items.[9] The highest and lowest number of items reported was 18 and 6 respectively. Statistical methods were often partially described and only 13 (22%) of 58 publications reported measures of statistical uncertainty. The median score of items fully reported was 11. For the papers published before 2003 (n=22) and after 2003 (n=36) this median score was 11 (interquartile range [IQR] 8-13) and 12.5 (IQR 10-16). Only one study explicitly mentioned the use of STARD for preparing the manuscript. Figure 3 shows the STARD assessment of each article arranged chronologically by year of publication.

The included papers were published in 16 different journals, from those journals only the *Archives of Ophthalmology* and the *British Journal of Ophthalmology* encouraged the authors to use the STARD flow diagram and checklist for reports of diagnostic accuracy. We were unable to access instructions for authors for two of the included journals.

	Item	1993 to 2003 (%) N = 22	2004 to 2013 (%) N = 36	Total (%) N = 58
	TITLE/ ABSTRACT			
1*	Identify the article as a study of diagnostic accuracy	18 (82)	36 (100)	54 (93)
	INTRODUCTION			
2	State the research questions of study aims.	22 (100)	36 (100)	58 (100)
	METHODS			
3	The study population	4 (18)	12 (33)	16 (28)
4	Participant recruitment	22 (100)	36 (100)	58 (100)
5	Participant sampling	16 (73)	23 (64)	39 (67)
6	Data collection strategy	8 (36)	19 (53)	27 (47)
7	The reference standard	6 (27)	24 (67)	30 (52)
8	Technical specifications of material and methods	13 (59)	29 (81)	42 (72)
9	Definition of and rationale of the index tests and the reference standard	10 (45)	26 (72)	36 (62)
10	The number, training and expertise of the persons executing and reading the tests	7 (32)	7 (19)	14 (24)
11	Whether or not the tests were masked to the readers	2 (9)	9 (25)	11 (19)
12[$]	Methods for calculating and comparing measures of diagnostic accuracy and uncertainty	2 (9)	3 (8)	5 (9)
13	Methods to calculate reproducibility	1 (5)	4 (11)	5 (9)
	RESULTS			
14	When the study was done	3 (14)	12 (33)	15 (26)
15	Clinical and demographic characteristics	13 (59)	27 (75)	40 (69)
16	Participant flow	5 (23)	14 (39)	19 (33)
17	Time interval between index and reference standard	3 (14)	6 (17)	9 (16)
18	Distribution of severity of disease and other diagnoses	18 (82)	33 (92)	51 (88)
19	Cross tabulation of the results of the index tests	12 (55)	22 (61)	34 (59)
20	Adverse effects from performing index or reference tests	NA	NA	NA
21[$]	Estimates of diagnostic accuracy and measures of uncertainty	3 (14)	10 (28)	13 (22)
22	How indeterminate results, missing responses and outliers of the index tests were handled	6 (27)	14 (39)	20 (34)
23 **	Estimates of diagnostic accuracy between subgroups	11 (50)	19 (53)	30 (52)
24	Estimates of test reproducibility	2 (9)	4 (11)	6 (10)
	DISCUSSION			
25	Discuss the clinical applicability of the study findings	22 (100)	36 (100)	58 (100)

Table 4. STARD item (scored as “Yes”)

NA - not applicable, due to the non-invasive nature of the test.
** Considered to be positive if the words diagnostic accuracy appeared in the title or abstract, or if the article was identified using the MeSH term sensitivity and specificity.*
*** Considered to be NA if there were no subgroups. Not applicable for 23 articles.*
$ If only estimates of diagnostic accuracy without a measure of uncertainty were given, this was scored as “partial.”

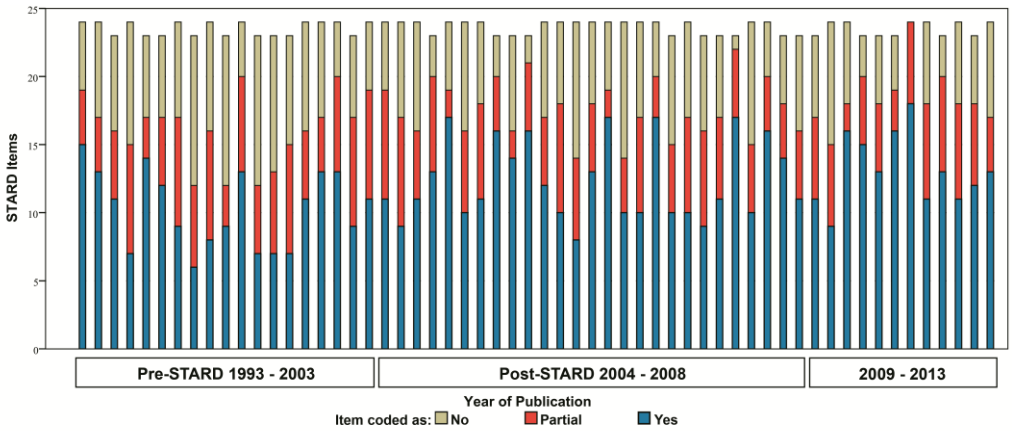

Figure 3. Stacked bar chart showing the STARD items reported and year of publication

Discussion

Automated perimetry is an effective tool for glaucoma diagnosis, as well as providing a means of monitoring disease deterioration to determine the effectiveness of treatment. However, a perfect diagnostic test for glaucoma does not exist, meaning that researchers commonly develop new tests that need to be scrutinised for diagnostic performance. High quality diagnostic accuracy studies are essential to evaluate test performance and reduce the risk of bias. The present study investigated the methodological quality and quality of reporting of diagnostic accuracy studies, using various types of perimeter to detect glaucomatous visual field loss, published between 1993 and 2003.

Our results showed that the overall methodological quality of these studies was low to moderate and they were often poorly reported. There were also differences in the reference standard used for glaucoma diagnosis, which could potentially lead to heterogeneity in test performance. A comparison of the studies published before and after the development of the QUADAS assessment tool and the publication of the STARD checklist in 2003 found that methodological quality and reporting had not substantially improved. These results are consistent with previous studies in ophthalmology [68-74] and in other medical specialities.[75, 76]

QUADAS items that scored particularly poorly (rated as ‘no’ or ‘unclear’) were: item 1 (inclusion of a representative spectrum of patients), item 4 (disease progression bias), item 10 (masking of reference test result) and item 14 (explanation of withdrawals). These sources of bias may have had a significant impact on the internal and external validity of the study findings reported in these papers.

The assessment of the methodological quality of a diagnostic accuracy study is closely related to how well the study is reported; a well-conducted study will score poorly on a quality assessment tool if the methods and results are not reported in sufficient detail. The guiding principle of the STARD initiative is *'to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study and to evaluate its generalisability'*.[7] The checklist consists of 25 items that roughly follow the structure of a journal article (Title/abstract, Introduction, Methods, Results and Discussion). The average number of items fully reported prior to the introduction of STARD was 9.2 (36.8%) compared to 12.8 (51.2%) for the period after publication. Only 3 items were fully reported by all studies including: study aims, method of participant recruitment and discussion of the clinical applicability of the study findings. Significant omissions included a full description of the study population in terms of inclusion and exclusion criteria (STARD item 3), which was reported by only 28% of studies. This is important for the assessment of the potential for spectrum bias which may impact on test performance and the applicability of the study findings. For example, tests with diagnostic promise normally become more sensitive as a disease becomes more severe; a study including participants with advanced disease only would by default report better sensitivity. Very few articles (19%) reported if the investigators are masked to the results of the other tests (STARD item 11). There is the possibility of inflation of the diagnostic accuracy estimates if readers of one test are aware of the results of either the reference standard or other clinical information when interpreting the index test. Whilst the majority of articles reported estimates of diagnostic accuracy, typically as sensitivities and specificities, only 22% of studies fully provided measures of statistical uncertainty e.g. confidence intervals. Only a third of the articles reported how withdrawals (STARD item 16) and indeterminate or missing results were handled (STARD item 22).

Over the past 20 years, a number of reporting guidelines have been developed to assist health researchers in writing up their work for publication, by specifying a minimum set of items that are required to provide a transparent account of what was done together with a clear description of the results of the study. The most widely known guidelines include the CONSORT statement for reporting parallel group randomised controlled trials, the STROBE statement for observational studies and the PRISMA statement for authors of systematic reviews (see the EQUATOR network for a full description of these guidelines).[77] Although an increase in reporting quality has been observed following the introduction of these guidelines, their adoption has been slow.[78] However, journal endorsement of the reporting guideline may improve the

completeness of reporting.[78] Only one of the articles included in the current review mentioned the use of STARD in the preparation of the manuscript and only two of the 16 journals in which these articles were published made reference to STARD in their 'Instructions to Authors'.

Conclusions

The major strength of this study was the comprehensive search strategy used, along with rigorous methods for study selection and quality appraisal. Both QUADAS and STARD checklists were examined, which provide a complementary assessment of both methodological quality and quality of reporting. The overall conclusion is that the quality of reporting of the diagnostic accuracy of perimetric tests in glaucoma is sub-optimal and appears not to have improved substantially following the development of the STARD initiative. Adoption and enforcement of STARD by ophthalmic journals is likely to lead to quality improvements in reporting of these studies.

References

1. Quigley, HA. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006; 90: 262–7.
2. Prior M, Francis JJ, Azuara-Blanco A, Anand N, Burr JM; Glaucoma screening Platform Study group. Why do people present late with advanced glaucoma? A qualitative interview study.*Br J Ophthalmol* 2013;97:1574-8.
3. Jampel, HD, Singh, K, Lin, SC, Chen, TC, Francis, BA, Hodapp, E, Samples, JR, Smith, SD.Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology* 2011; 118:986-1002.
4. Deeks, JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323:157–62.
5. Mowatt G, Burr JM, Cook JA, Siddiqui MAR, Ramsay C, Fraser C, et al. Screening Tests for Detecting Open-Angle Glaucoma: Systematic Review and Meta-analysis. Investigative Ophthalmology & Visual Science. 2008 Dec 1;49(12):5373–85.
6. 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. *BMC Med Res Methodol* 2003; 3: 25.
7. Bossuyt, PM, Reitsma, JB, Bruns, DE, Gatsonis, CA, Glasziou, PP, Irwig, LM, Lijmer, JG, Moher, D, Rennie, D, de Vet, HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003; 49:1-6.

8. Deville, WL, Bezemer, PD, Bouter, LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *J Clin Epidemiol* 2000; 53:65–9.
9. Landis, JR, Koch, GG. The measurement of observer agreement for categorical data. *Biometrics 1977;* 33:159-174.
10. Wishart, PK. Oculokinetic perimetry compared with Humphrey visual field analysis in the detection of glaucomatous visual field loss. *Eye* (Lond) 1993; 7:113–21.
11. Mutlukan, E, Damato, BE, Jay, JL. Clinical evaluation of a multi-fixation campimeter for the detection of glaucomatous visual field loss. *Br J Ophthalmol* 1993;77:332–8.
12. Katz, J, Tielsch, JM, Quigley, HA, Javitt, J, Witt, K, Sommer, A. Automated suprathreshold screening for glaucoma: the Baltimore Eye Survey. I*nvest Ophthalmol Vis Sci* 1993; 34: 3271–7.
13. Adachi, M, Shirato, S. The usefulness of the Noise-Field Test as a screening method for visual field defects. *Jpn J Ophthalmol* 1994; 38: 392–9.
14. Christoffersen, T, Fors, T, Waage, S, Holtedahl, K. Glaucoma screening with oculokinetic perimetry in general practice: is its specificity acceptable? *Eye* (Lond). 1995; 9: 36–9.
15. Sponsel, WE, Ritch, R, Stamper, R, Higginbotham, EJ, Anderson, DR, Wilson, MR, Zimmerman, TJ. Prevent Blindness America visual field screening study. The Prevent Blindness America Glaucoma Advisory Committee. *Am J Ophthalmol* 1995;120: 699–708.
16. Graham, SL, Drance, SM, Chauhan, BC, Swindale, NV, Hnik, P, Mikelberg, FS, Douglas, GR. Comparison of psychophysical and electrophysiological testing in early glaucoma. *Invest Ophthalmol Vis Sci* 1996; 37: 2651–62.
17. Huang, AS, Smith, SD, Quigley, HA. The efficacy of the Dicon screening field to detect eyes with glaucomatous field loss by Humphrey threshold testing. *J Glaucoma* 1998;7:158–64.
18. Bosworth, CF, Sample, PA, Gupta. N, Bathija, R, Weinreb, RN. Motion automated perimetry identifies early glaucomatous field defects. *Arch Ophthalmol* 1998;116:1153–8.
19. Sim, DH, Goh, LG. Screening for glaucoma in the Chinese elderly population in Singapore. *Singapore Med J* 1999; 40: 644–7.
20. Yamada, N, Chen, PP, Mills, RP, Leen, MM, Lieberman, MF, Stamper, RL, Stanford, DC. Screening for glaucoma with frequency-doubling technology and Damato campimetry. *Arch Ophthalmol* 1999 ;117:1479–84.
21. Fabré, K, Michiels, I, Zeyen, T. The sensitivity and specificity of TOP, FDP and GDX in screening for early glaucoma. *Bull Soc Belge Ophtalmol* 2000;275:17–23.

22. Burnstein, Y, Ellish, NJ, Magbalon, M, Higginbotham, EJ. Comparison of frequency doubling perimetry with Humphrey visual field analysis in a glaucoma practice. *Am J Ophthalmol* 2000;129: 328–33.
23. Cello, KE, Nelson-Quigg, JM, Johnson, CA. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. Am J Ophthalmol 2000;129: 314–22.
24. Trible, JR, Schultz, RO, Robinson, JC, Rothe, TL. Accuracy of glaucoma detection with frequency-doubling perimetry. *Am J Ophthalmol* 2000;129:740–5.
25. Vitale, S, Smith, TD, Quigley, T, Kerrigan-Baumrind, TA, Pease, TE, Varma, R, Friedman, TS, Katz, J, Tielsch, JM. Screening performance of functional and structural measurements of neural damage in open-angle glaucoma: a case-control study from the Baltimore Eye Survey. *J Glaucoma* 2000; 9:346–56.
26. Paczka, JA, Friedman, DS, Quigley, HA, Barron, Y, Vitale, S. Diagnostic capabilities of frequency-doubling technology, scanning laser polarimetry, and nerve fiber layer photographs to distinguish glaucomatous damage. *Am J Ophthalmol* 2001;131:188–97.
27. Wadood, AC, Azuara-Blanco, A, Aspinall, P, Taguri, A, King, AJ. Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, and Humphrey Swedish interactive threshold algorithm-fast perimetry in a glaucoma practice. *Am J Ophthalmol* 2002;133:327–32.
28. Bayer, AU, Maag, K-P, Erb, C. Detection of optic neuropathy in glaucomatous eyes with normal standard visual fields using a test battery of short-wavelength automated perimetry and pattern electroretinography. *Ophthalmology* 2002;109:1350–61.
29. Horn, FK, Wakili, N, Jünemann, AM, Korth, M. Testing for glaucoma with frequency-doubling perimetry in normals, ocular hypertensives, and glaucoma patients. *Graefes Arch Clin Exp Ophthalmol* 2002; 240:658–65.
30. Iwasaki, A, Sugita, M. Performance of glaucoma mass screening with only a visual field test using frequency-doubling technology perimetry. *Am J Ophthalmol* 2002;134: 529–37.
31. Horn, FK, Nguyen, NX, Mardin, CY, Jünemann, AG. Combined use of frequency doubling perimetry and polarimetric measurements of retinal nerve fiber layer in glaucoma detection. *Am J Ophthalmol* 2003;135:160–8.
32. Detry-Morel, M, Zeyen, T, Kestelyn, P, Collignon, J, Goethals, M. Screening for glaucoma in a general population with the non-mydriatic fundus camera and the frequency doubling perimeter. *Eur J Ophthalmol* 2004;14:387–93.
33. Spry, PG, Johnson, CA, Mansberger, SL, Cioffi, GA. Psychophysical investigation

of ganglion cell loss in early glaucoma: *J Glaucoma.* 2005;14:11–9.

34. Brusini, P. Probing glaucoma visual damage by rarebit perimetry. *Br J Ophthalmol* 2005; 89:180–4.
35. Heeg, GP, Blanksma, LJ, Hardus, PLLJ, Jansonius, NM. The Groningen Longitudinal Glaucoma Study. I. Baseline sensitivity and specificity of the frequency doubling perimeter and the GDx nerve fibre analyser: *Acta Ophthalmologica* 2005; 83:46–52.
36. Mansberger, SL, Johnson, CA, Cioffi, GA, Choi, D, Krishnadas, SR, Srinivasan, M, et al. Predictive value of Frequency Doubling Technology perimetry for detecting glaucoma in a developing country. *J Glaucoma* 2005; 14:128–34.
37. Babalola, OE. The motion sensitivity screening test in clinical practice in abuja, Nigeria: affordable automated perimetry for the third world? *Afr J Med Med Sci* 2005; 34:119–24.
38. Robin, TA, Müller, A, Rait, J, Keeffe, JE, Taylor, HR, Mukesh, BN. Performance of community-based glaucoma screening using Frequency Doubling Technology and Heidelberg Retinal Tomography. *Ophthalmic Epidemiol* 2005;12:167–78.
39. Fogagnolo, P, Mazzolani, F, Rossetti, L, Orzalesi, N. Detecting glaucoma with Frequency Doubling Technology perimetry: A comparison between N-30 and C-20 screening programs. *J Glaucoma* 2005;14:485–91.
40. Matsumoto, C, Takada, S, Okuyama, S, Arimura, E, Hashimoto, S, Shimomura, Y. Automated flicker perimetry in glaucoma using Octopus 311: a comparative study with the Humphrey Matrix. *Acta Ophthalmologica* 2006;84:210–5.
41. Brusini, P, Salvetat, ML, Zeppieri, M, Parisi, L. Frequency Doubling Technology perimetry with the Humphrey Matrix 30-2 test, *J Glaucoma* 2006 Apr;15(2):77–83.
42. Pierre-Filho, P, Schimiti, RB, De Vasconcellos, JPC, Costa, VP. Sensitivity and specificity of Frequency Doubling Technology, tendency-oriented perimetry, SITA Standard and SITA Fast perimetry in perimetrically inexperienced individuals. *Acta Ophthalmologica* 2006;84:345–50.
43. North, RV, Jones, AL, Hunter, E, Morgan, JE, Wild, JM. Evaluation of the high specificity Screening Program (C-20-1) of the Frequency Doubling Technology (FDT) perimeter in clinical practice. *Eye* 2006; 20:681–7.
44. Gardiner, SK, Anderson, DR, Fingeret, M, McSoley, JJ, Johnson, CA. Evaluation of decision rules for Frequency Doubling Technology screening tests. *Optom Vis Sci* 2006; 83: E432–E439.
45. Sample, PA, Medeiros, FA, Racette, L, Pascual, JP, Boden, C, Zangwill, LM, Bowd, C, Weinreb, RN. Identifying glaucomatous vision loss with visual-function-specific perimetry in the diagnostic innovations in glaucoma study. *Invest Ophthalmol Vis*

Sci 2006; 47:3381–9.

46. Iwase, A, Tomidokoro, A, Araie, M, Shirato, S, Shimizu, H, Kitazawa, Y. Performance of Frequency Doubling Technology Perimetry in a population-based prevalence survey of glaucoma. *Ophthalmology* 2007;114: 27–32.
47. 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;143:263–271.
48. Fortune, B, Demirel, S, Zhang, X, Hood, DC, Patterson, E, Jamil, A, Mansberger, SL, Cioffi, GA, Johnson, CA. Comparing multifocal VEP and standard automated perimetry in high-risk ocular hypertension and early glaucoma. *Invest Ophthalmol Vis Sci* 2007;48:1173–80.
49. Kumar, S, Giubilato, A, Morgan, W, Jitskaia, L, Barry, C, Bulsara, M, Constable, IJ, Yogesan, K. Glaucoma screening: analysis of conventional and telemedicine-friendly devices. *Clin Exp Ophthalmol* 2007; 35:237–43.
50. Leeprechanon, N, Giaconi, JA, Manassakorn, A, Hoffman, D, Caprioli, J. Frequency Doubling perimetry and short-wavelength automated perimetry to detect early glaucoma. *Ophthalmology* 2007;114:931–7.
51. Ferreras, A, Polo, V, Larrosa, JM, Pablo, LE, Pajarin, AB, Pueyo, V, Honrubia, FM. 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;16:372–83.
52. Horn FK, Brenning A, Junemann AG, Lausen B. Glaucoma Detection With Frequency Doubling Perimetry and Short-wavelength Perimetry: Journal of Glaucoma. 2007 Jun;16(4):363–71.
53. Tóth, M, Kothy, P, Vargha, P, Hollo, G. Accuracy of combined GDx-VCC and Matrix FDT in a glaucoma screening trial: *J Glaucoma* 2007;16:462–70.
54. Wang, YX, Xu, L, Zhang, RX, Jonas, JB. Frequency-doubling threshold perimetry in predicting glaucoma in a population-based study: The Beijing Eye Study. *Arch Ophthalmol* 2007;125:1402–6.
55. Spry, PGD, Hussin, HM, Sparrow, JM. Performance of the 24-2-5 Frequency Doubling Technology screening test: a prospective case study. Br J Ophthalmol 2007;91:1345–9.
56. Tafreshi, A, Sample, PA, Liebmann, JM, Girkin, CA, Zangwill, LM, Weinreb, RN, et al. Visual function-specific perimetry to identify glaucomatous visual loss using three different definitions of visual field abnormality. *Invest Ophthalmol Vis Sci* 2009; 50:1234–40.
57. Salim, S, Netland, PA, Fung, KH, Smith, ME, Aldridge, A. Assessment of the

Student Sight Savers Program methods for glaucoma screening. *Ophthalmic Epidemiol* 2009;16:238–42.

58. Cook, C, Cockburn, N, van der Merwe, J, Ehrlich, R. Cataract and glaucoma case detection for Vision 2020 programs in Africa: An evaluation of 6 possible screening tests. *J Glaucoma* 2009;18:557–62.
59. Racette, L, Chiou, CY, Hao, J, Bowd, C, Goldbaum, MH, Zangwill, LM, Lee, TW, Weinreb, RN, Sample, PA. Combining functional and structural tests improves the diagnostic accuracy of relevance vector machine classifiers. *J Glaucoma* 2010;19:167–75.
60. Salvetat, ML, Zeppieri, M, Tosoni, C, Parisi, L, Brusini, P. Non-conventional perimetric methods in the detection of early glaucomatous functional damage. *Eye* 2010;24:835–42.
61. Rowe, FJ, Sueke, H, Gawley, SD. Comparison of Damato campimetry and Humphrey automated perimetry results in a clinical population. *Br J Ophthalmol* 2010; 94:757–62.
62. 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;45:494–500.
63. Kamdeu Fansi, AA, Li, G, Harasymowycz, PJ. The validity of screening for open-angle glaucoma in high-risk populations with single-test screening mode Frequency Doubling Technology perimetry (FDT). *J Glaucoma* 2011; 20:167–71.
64. Francis, BA, Varma, R, Vigen, C, Lai, MY, Winarko, J, Nguyen, B, Azen, S, Los Angeles Latino Eye Study Group. Population and high-risk group screening for glaucoma: The Los Angeles Latino Eye Study. *Invest Ophthalmol Vis Sci* 2011; 52:6257–64.
65. Gonzalez de la Rosa, M. Pulsar perimetry in the diagnosis of early glaucoma. *Am J Ophthalmol* 2011;152: 500–1.
66. Horn, FK, Lämmer, R, Mardin, CY, Jünemann, AG, Michelson, G, Lausen, B, Adler, W. Combined evaluation of Frequency Doubling Technology perimetry and scanning laser ophthalmoscopy for glaucoma detection using automated classification. *J Glaucoma* 2012;21:27–34.
67. Anton, A, Capilla, P, Morilla-Grasa, A, Luque, MJ, Artigas, JM, Felipe, A. Multichannel functional testing in normal subjects, glaucoma suspects, and glaucoma patients. *Invest Ophthalmol Vis Sci* 2012; 53:8386–95.
68. Siddiqui, MA, Azuara-Blanco, A, Burr, J. The quality of reporting of diagnostic accuracy studies published in ophthalmic journals. *Br J Ophthalmol* 2005; 89:261-5.
69. Shunmugam, M, Azuara-Blanco, A. The quality of reporting of diagnostic accuracy

studies in glaucoma using the Heidelberg retina tomograph. *Invest Ophthalmol Vis Sci* 2006; 47: 2317–23

70. 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.
71. Paranjothy, B, Shunmugam, M, Azuara-Blanco, A. The quality of reporting of diagnostic accuracy studies in glaucoma using scanning laser polarimetry. *J Glaucoma* 2007;16:670-5.
72. Zafar, A, Khan, GI, Siddiqui, MA. The quality of reporting of diagnostic accuracy studies in diabetic retinopathy screening: a systematic review. *Clin Experiment Ophthalmol* 2008;36:537-42.
73. Castillo, MM, Mowatt, G, Lois, N, Elders, A, Fraser, C, Amoaku, W, Burr, JM, Lotery, AJ, Ramsay, CR, Azuara-Blanco, A. Optical coherence tomography for the diagnosis of neovascular age-related macular degeneration: a systematic review. *Eye* (Lond). 2014;28:1399-1406.
74. Harper R, Reeves B. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. Invest Ophthalmol Vis Sci. 1999 Jul; 40(8):1650–7.
75. Coppus, SF, van der Veen, F, Bossuyt, PM, Mol, BW. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertil Steril* 2006;86:1321-9.
76. Wilczynski, NL. Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication--before-and-after study. *Radiology* 2008; 248:817-23.
77. The EQUATOR Network. www.equator-network.org/ (accessed 3.1.15)
78. Turner, L, Shamseer, L, Altman, DG, Weeks, L, Peters J, Kober, T, Dias,S, Schulz, KF, Plint, AC, Moher, D. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11: MR000030.

Acknowledgements

The authors thank Optovue Inc. for part-funding a PhD studentship for Bruno Fidalgo

Disclosure

The authors report no conflicts of interest

Figure captions

Figure 1. Flow diagram showing selection process.

Figure 2. Stacked bar chart showing the QUADAS items reported and year of publication

Figure 3. Stacked bar chart showing the STARD items reported and year of publication