

SCIENTIFIC REPORT

The quality of reporting of diagnostic accuracy studies published in ophthalmic journals

M A R Siddiqui, A Azuara-Blanco, J Burr

Br J Ophthalmol 2005;89:261–265. doi: 10.1136/bjo.2004.051862

Aim: To evaluate the quality of reporting of all diagnostic studies published in five major ophthalmic journals in the year 2002 using the Standards for Reporting of Diagnostic Accuracy (STARD) initiative parameters.

Methods: Manual searching was used to identify diagnostic studies published in 2002 in five leading ophthalmic journals, the *American Journal of Ophthalmology (AJO)*, *Archives of Ophthalmology (Archives)*, *British Journal of Ophthalmology (BJO)*, *Investigative Ophthalmology and Visual Science (IOVS)*, and *Ophthalmology*. The STARD checklist of 25 items and flow chart was used to evaluate the quality of each publication.

Results: A total of 16 publications were included (*AJO*=5, *Archives*=1, *BJO*=2, *IOVS*=2, and *Ophthalmology*=6). More than half of the studies ($n=9$) were related to glaucoma diagnosis. Other specialties included retina ($n=4$) cornea ($n=2$), and neuro-ophthalmology ($n=1$). The most common description of diagnostic accuracy was sensitivity and specificity values, published in 13 articles. The number of fully reported items in evaluated studies ranged from eight to 19. Seven studies reported more than 50% of the STARD items.

Conclusions: The current standards of reporting of diagnostic accuracy tests are highly variable. The STARD initiative may be a useful tool for appraising the strengths and weaknesses of diagnostic accuracy studies.

Current ophthalmological practice relies on diagnostic tests using sophisticated technologies that are constantly evolving. Diagnostic accuracy studies determine the performance of the test in diagnosing the target condition. Improperly conducted and incompletely reported studies are prone to bias that, in turn, may lead to overly optimistic appraisal of evaluated tests.¹ The performance of a diagnostic test can be estimated in several ways, including sensitivity, specificity, receiver operating characteristic curves (ROC), positive and negative predictive values, likelihood ratios, and diagnostic odd ratios.^{2,3}

To improve the quality of reporting of diagnostic accuracy studies the Standards for Reporting of Diagnostic Accuracy (STARD) initiative was published.⁴ During a consensus conference in the year 2000, the STARD project group developed a checklist of 25 items and a prototypical flow chart.^{2,5}

The aim of this study was to examine the current standard of reporting of diagnostic accuracy studies using the STARD parameters. Current standards may provide a useful baseline to measure the impact of the introduction of the STARD statement in the future.

METHODS

The five leading ophthalmic journals (that is, according to the impact factor) with clinical research sections or articles were

selected. Basic science research and subspecialty journals were excluded. The journals evaluated were *AJO*, *Archives*, *BJO*, *IOVS*, and *Ophthalmology*. Since search strategies for diagnostic accuracy tests are suboptimal⁶ a hand search of all issues of 2002 was done. In these journals all manuscripts related to a diagnostic procedure were identified. Manuscripts were selected for inclusion if the diagnostic test was used in human subjects, the test was intended for clinical use, and measures of diagnostic accuracy were provided. Review articles, case reports, and longitudinal studies were excluded. The full paper was assessed for inclusion by one author, if uncertain; the study was selected as potentially suitable for inclusion. The selected papers were then independently assessed for inclusion by two investigators; if there was a disagreement, a consensus was reached.

The STARD checklist (table 1) was used to score the studies. Each item could be considered to be fully, partially, or not reported. If the item was “not applicable” it was marked as such. For example, item 21 required reporting of estimates of diagnostic accuracy and measure of statistical uncertainty. If a study reported estimates of accuracy but no measure of precision it was considered partially fulfilled. Similarly, item 20 (reporting of adverse events associated with the test) was scored non-applicable for non-invasive studies (for example, visual field tests, fundus photography).

One investigator assessed all the included studies. To evaluate the interobserver variability in the rating of the STARD criteria, a second investigator examined four randomly selected publications, masked to the results of the first investigator.

RESULTS

Twenty manuscripts were identified as potentially suitable for inclusion. After review of the full paper, four reports were excluded as they did not meet the inclusion criteria. One longitudinal study evaluated the value of short wavelength automated perimetry to predict the development of glaucoma.⁷ Another study discussed the use of magnetic resonance imaging (MRI) to differentiate between optic neuritis and non-arteritic anterior ischaemic neuropathy.⁸ Another study evaluated longitudinally changes in the wavefront aberration of patients with keratoconus.⁹ The fourth excluded paper described videokeratography findings in children with vernal keratoconjunctivitis and compared them with those of healthy children, without attempting to use these differences as a diagnostic test.¹⁰ A total of 16

Abbreviations: *AJO*, *American Journal of Ophthalmology*; *Archives*, *Archives of Ophthalmology*; *BJO*, *British Journal of Ophthalmology*; *IOVS*, *Investigative Ophthalmology and Visual Science*; *MRI*, magnetic resonance imaging; *ROC*, receiver operating characteristic; *STARD*, Standards for Reporting of Diagnostic Accuracy

Table 1 STARD checklist^{4, 5}

Section and topic	Item	Description
Title, abstract, and keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading "sensitivity and specificity")
Introduction	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
Methods		
Participants	3	Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected
	4	Describe participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
	5	Describe participant sampling: was this a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected
	6	Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	Describe the reference standard and its rationale
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or reference standard, or both
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard
	10	Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard
	11	Were the readers of the index tests and the reference standard blind (masked) to the results of the other test? Describe any other clinical information available to the readers.
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (eg, 95% confidence intervals)
	13	Describe methods for calculating test reproducibility, if done
Results		
Participants	14	Report when study was done, including beginning and ending dates of recruitment
	15	Report clinical and demographic characteristics (eg, age, sex, spectrum of presenting symptoms, co-morbidity, current treatments, and recruitment centre)
	16	Report how many participants satisfying the criteria for inclusion did or did not undergo the index tests or the reference standard, or both; describe why participants failed to receive either test (a flow diagram is strongly recommended)
Test results	17	Report time interval from index tests to reference standard, and any treatment administered between
	18	Report distribution of severity of disease (define criteria) in those with the target condition and other diagnoses in participants without the target condition
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, report the distribution of the test results by the results of the reference standard
	20	Report any adverse events from performing the index test or the reference standard
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (eg, 95% confidence intervals)
	22	Report how indeterminate results, missing responses, and outliers of index tests were handled
	23	Report estimates of variability of diagnostic accuracy between readers, centres, or subgroups of participants, if done
	24	Report estimates of test reproducibility, if done
Discussion	25	Discuss the clinical applicability of the study findings

studies (table 2) were included in this review (*AJO* = 5, *Archives* = 1, *BJO* = 2, *IOVS* = 2, and *Ophthalmology* = 6).

Glaucoma was the specialty with the highest number of studies (n = 9). Other specialties included retina (n = 4), cornea (n = 2), and neuro-ophthalmology (n = 1) (table 3). Interobserver rating agreement was observed in 92% of items. Among the 16 articles the range of fully reported positive STARD items was from eight to 19. Less than half the studies (n = 7) explicitly reported more than 50% of STARD items. Reporting of an individual STARD item ranged from 1/16 (item 24) to 16/16 (100%) (item 2 and item 25) (table 4). The commonest description of diagnostic accuracy was sensitivity and specificity values (n = 13), followed by area under the ROC curve (n = 4). The reporting of each of the items is described in table 4.

DISCUSSION

In 1978 Ransohoff and Feinstein¹¹ first reported a detailed analysis of diagnostic accuracy studies and identified the major sources of bias. Since then there have been numerous articles identifying a variety of biases as a potential source of inaccuracies in the indices of diagnostic accuracy.¹²⁻¹⁷ Reid *et al*¹² evaluated diagnostic accuracy studies published in four prominent medical journals between 1978 and 1993. They evaluated the quality of 20 diagnostic test studies published during this period against seven methodological standards. Their study showed that quality of reporting was of moderate or low quality, and that the essential elements of data required to evaluate a study were missing in the majority of the reports. Although there had been some improvement over time, most of the diagnostic accuracy tests were inadequately reported.

Table 2 Included studies

No	Authors	Title	Journal	Volume	Pages
1	Wadood AC, Azuara-Blanco A, Aspinall P, <i>et al</i>	Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, and Humphrey Swedish interactive threshold algorithm-fast perimetry in a glaucoma practice	AJO	133	327–32
2	Kesen MR, Spaeth GL, Henderer JD, <i>et al</i>	The Heidelberg retina tomograph vs clinical impression in the diagnosis of glaucoma	AJO	133	613–6
3	Lin DY, Blumenkranz MS, Brothers RJ, <i>et al</i>	The sensitivity and specificity of single-field non-mydratic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydratic colour photography	AJO	134	204–13
4	Tatemichi M, Nakano T, Tanaka K, <i>et al</i>	Performance of glaucoma mass screening with only a visual field test using frequency-doubling technology perimetry	AJO	134	529–37
5	Williams ZY, Schuman JS, Gamell L, <i>et al</i>	Optical coherence tomography measurement of nerve fiber layer thickness and the likelihood of a visual field defect	AJO	134	538–46
6	Kowalski RP, Karenchak LM, Shah C, <i>et al</i>	ELVIS: a new 24-hour culture test for detecting herpes simplex virus from ocular samples	Archives	120	960–2
7	Funaki S, Shirakashi M, Yaeeda K, <i>et al</i>	Specificity and sensitivity of glaucoma detection in the Japanese population using scanning laser polarimetry	BJO	86	70–4
8	Stanford MR, Gras L, Wade A, <i>et al</i>	Reliability of expert interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis	BJO	86	636–9
9	Greaney MJ, Hoffman DC, Garway-Heath DF, <i>et al</i>	Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma	IOVS	43	140–5
10	Wall M, Neahring RK, Woodward KR	Sensitivity and specificity of frequency doubling perimetry in neuro-ophthalmic disorders: a comparison with conventional automated perimetry	IOVS	43	1277–83
11	Rudnisky CJ, Hinz BJ, Tennant MT, <i>et al</i>	High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema	Ophthalmology	109	267–74
12	Soliman MA, de Jong LA, Ismaeil AA, <i>et al</i>	Standard achromatic perimetry, short wavelength automated perimetry, and frequency doubling technology for detection of glaucoma damage	Ophthalmology	109	444–54
13	Fransen SR, Leonard-Martin TC, Feuer WJ, <i>et al</i>	Clinical evaluation of patients with diabetic retinopathy: accuracy of the Inoveon diabetic retinopathy-3DT system	Ophthalmology	109	595–601
14	Budenz DL, Rhee P, Feuer WJ, <i>et al</i>	Sensitivity and specificity of the Swedish interactive threshold algorithm for glaucomatous visual field defects	Ophthalmology	109	1052–8
15	Bayer AU, Maag KP, 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	109	1350–61
16	Rao SN, Raviv T, Majumdar PA, <i>et al</i>	Role of Orbscan II in screening keratoconus suspects before refractive corneal surgery	Ophthalmology	109	1642–6

Harper and Reeves evaluated the quality of reporting of ophthalmic diagnostic tests¹⁵ published in the early and mid-1990s. They showed a limited compliance with accepted methodological standards. The compliance in ophthalmic journals was no worse than other evaluations published in general medical journals, but only 25% of articles complied with more than 50% of methodological standards

In this current appraisal of recent ophthalmic publications using the STARD checklist, similar flaws were found. Less than 50% of articles (n = 7) reported more than half of STARD items. Information on key elements of design, conduct, analysis, and interpretation of diagnostic studies were frequently missed. To our knowledge, STARD has not been used to appraise the quality of reporting of diagnostic accuracy studies in other medical specialties.

The importance of describing the selection of the study population in appraising a diagnostic test cannot be over-emphasised (item 3). For example, Harper *et al* showed how indices of diagnostic accuracy of tonometry for glaucoma greatly varied depending on the characteristics of the study population.¹⁶ Most publications reported this issue properly (n = 13).

Review bias, including test review bias (inflation of diagnostic accuracy indices by knowing the results of the gold standard while reviewing the index test), diagnostic review bias (knowledge of the outcome of the index test while reviewing gold standard), and clinical review bias (additional clinical information available to the reader, which would not normally be available when interpreting the index test results) can lead to inflation of the measures of

diagnostic accuracy. Reader masking (item 11) was reported in less than half of the studies (n = 6).

Methods for calculating test reproducibility or citation of reproducibility studies (item 13) was among the least commonly reported items from the STARD checklist (n = 2). There may be a lack of understanding of effects of poor reproducibility on the final outcome of a diagnostic accuracy test.

Verification or workup bias (item 16) occurs when gold standard test is performed only on people who have already tested positive for the index test.³ It is important to describe how many patients satisfying inclusion criteria failed to undergo index or reference tests and the reason of failing to do so. A flow diagram is highly recommended to clearly explain this issue.^{2,4} This item was reported in four studies.

Since the technology for existing tests is rapidly improving, it is important to report the actual dates when the study was performed. This will allow the reader to consider any technological advancement since the study was done. This information was provided in less than half of articles (n = 6).

Spectrum bias results from differences in the severity of target condition and co-morbidity. Incomplete reporting of clinical spectrum (item 18) may result in inaccurate diagnostic accuracy estimates—for example, advanced disease status would lead to increased sensitivity of a diagnostic test. This item was fully reported in 10 studies.

Confidence intervals (CIs) were reported in only a quarter (n = 4) of studies. A recent review by Harper and Reeves¹⁷ revealed that CIs were reported in only 50% of diagnostic evaluation reports published in the *BMJ* during the 2 year

Table 3 Details of included studies

No	Article	Target condition	Index test	Reference standard	Total number of participants
1	Wadood <i>et al</i>	Glaucoma	FDT, TOP, and SAP	Stereoscopic disc analysis	98
2	Kesen <i>et al</i>	Glaucoma	HRT	Clinical impression	200
3	Lin <i>et al</i>	Diabetic retinopathy screening	Single field non-mydratric monochromatic digital fundus photography	Standardised mydratric colour photography or ophthalmoscopy	197
4	Iwasaki <i>et al</i>	Glaucoma screening	FDT	SAP, Octopus and clinical	14, 814
5	Williams <i>et al</i>	Glaucoma	OCT	SAP or FDT	276
6	Kowalski <i>et al</i>	HSV ocular disease	ELVIS	Cell culture and Herp check	483
7	Funaki <i>et al</i>	Glaucoma	SLP	SAP	184
8	Stanford <i>et al</i>	Toxoplasma retinochoroiditis	Retinal photography	None	96
9	Greany <i>et al</i>	Glaucoma	ODP, CLSO, SLP, OCT	SAP	89
10	Wall <i>et al</i>	Neuro-ophthalmic disorders	FDT	SAP and clinical diagnosis	139
11	Rudinsky <i>et al</i>	CSMO	High resolution digital fundus photography	Contact lens biomicroscopy	120
12	Soliman <i>et al</i>	Glaucoma	SWAP and FDT	SAP and clinical	123
13	Fransen <i>et al</i>	Diabetic Retinopathy	Inoveon's DR-3DT system	DRS7 photography	290
14	Budenz <i>et al</i>	Glaucoma	SAP (SITA)	SAP (FT)	172
15	Bayer <i>et al</i>	Glaucoma	SWAP, FDT, and PERG	SAP	72
16	Rao <i>et al</i>	Keratoconus	ORBSKAN II indices	VKG indices	110

FDT, frequency doubling technology, TOP, tendency oriented perimetry, SAP, standard automated perimetry (SAP-SITA, Swedish Interactive Threshold Algorithm; SAP-FT, full threshold perimetry) HRT, Heidelberg retinal tomograph, OCT, optical coherence tomography, ODP, optic disc pictures, CLSO, confocal laser scanning ophthalmoscope, SLP, scanning laser polarimetry, CSMO, clinically significant macular oedema, SWAP, short wavelength automated perimetry, PERG, pattern electroretinogram

Table 4 Summary score of STARD items

	Item	Yes	Partial*	No	NA†
Title, abstract, and keywords	1	8	0	8	0
Introduction	2	16	0	0	0
Methods					
Participants	3	13	3	0	0
	4	13	0	3	0
	5	8	0	8	0
	6	13	2*	1	0
Test methods	7	5	6	5	0
	8	11	3*	2	0
	9	8	4*	4	0
	10	6	4	6	0
	11	6	1*	9	0
Statistical methods	12	14	0	2	0
	13	2	0	14	0
Results					
Participants	14	6	0	10	0
	15	12	0	4	0
	16	4	0	11	1†
Test results	17	9	0	6	1†
	18	10	1	5	0
	19	7	3	6	0
	20	0	0	0	16†
Estimates	21	4	12	0	0
	22	5	0	11	0
	23	4	0	12	0
	24	1	0	15	0
Discussion	25	16	0	0	0
Flow diagram		1			

*Item 6. Two publications did not mention whether the design was prospective or retrospective but in the text mentioned that the patients were consented for the study.

Item 8. Three studies only cited a reference for technical details of either the index or the reference test.

Item 9. Four articles described units and/or cut-off points but no rationale and were considered to partially fulfil this item. For example one study (table 2 no 7) classified glaucoma severity on the basis of mean deviations (MD) without providing any rationale or reference.

Item 11. Described that some information was provided to the investigators, although it was not known whether the investigators were masked or not

†Item 16. This item was classified as "not applicable" in a retrospective study that only included patients who had both the index and reference tests done (table 2 no 5).

Item 17. This item was classified as "not applicable" in a retrospective study in which frozen histopathological samples were re-analysed with a new diagnostic test (table 2 no 6)

Item 20. Reporting of adverse events was marked as "not applicable" because of the non-invasive nature and safety profile of the tests.

period of 1996 and 1997. Since the absolute values of diagnostic accuracy are only estimates, when evaluations of diagnostic accuracy are reported the precision of the sensitivity and specificity or likelihood ratios should be reported. Reporting of confidence interval is essential to allow a physician to know the range within which the true values of the indices are likely to lie.¹⁷

Intermediate, indeterminate, and uninterpretable results may not always be included in final assessment of the diagnostic accuracy of a test.¹⁸ The frequency of these results, by itself, is an important pointer of the overall usefulness of the test.² Approximately one third of studies (n = 5) reported this item (item 22). Diagnostic accuracy in subgroups was reported in only a quarter of studies (n = 4) (item 23).

The STARD group strongly recommends use of a flow diagram to clearly communicate the design of the study and provide the exact number of participants at each stage of the study.² A flow diagram has been a valuable addition to the report of randomised clinical trials. It has been reported that flow diagrams are associated with improved quality of reporting of randomised controlled trials.¹⁹ A flow diagram was present in only one of the evaluated studies.

In a similar and previous effort to improve the quality of reporting of the literature, to prevent shortcomings and biases in randomised control trials, the CONSORT statement was introduced in 1995.²⁰ Use of CONSORT has shown to improve the quality of reporting of randomised controlled trials (RCTs).²¹ Sanchez-Thorin *et al*²² compared RCTs published in *Ophthalmology* during 1999 with the ones published in 1991–4 before the adoption of the CONSORT statement, and found an improvement in the quality of reporting. Future research will be able to evaluate the impact of the STARD initiative on the accuracy and completeness of reporting of studies on diagnostic accuracy.

ACKNOWLEDGEMENTS

The Health Services Research Unit is funded by Chief Scientist Office of the Scottish Executive Health Department; the views expressed here are those of the authors.

Authors' affiliations

M A R Siddiqui, A Azuara-Blanco, Department of Ophthalmology, Grampian University Hospitals NHS Trust, UK
J Burr, Health Services Research Unit, University of Aberdeen, UK

Competing interests: none declared.

Correspondence to: Augusto Azuara-Blanco, PhD, FRCS(Ed), The Eye Clinic, Aberdeen Royal Infirmary, Aberdeen AB25 2ZN, UK; aazblanco@aol.com

Accepted for publication 8 August 2004

REFERENCES

- Lijmer JG, Mol BW, Heisterkamp S, *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
- Bossuyt PM, Reitsma JB, Bruns DE, *et al*. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;**49**:7–18.
- Deeks J. Systematic reviews in health care: systematic reviews of diagnostic tests. *BMJ* 2001;**323**:157–62.
- Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Standards for Reporting of Diagnostic Accuracy steering group Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *BMJ* 2003;**326**:41–4.
- STARD statement. Available at www.consort-statement.org/stardstatement.htm (accessed 19 February 2004).
- 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.
- Polo V, Larrosa JM, Pinilla I, *et al*. Predictive value of short-wavelength automated perimetry: a 3-year follow-up study. *Ophthalmology* 2002;**109**:761–5.
- Rizzo JF III, Andreoli CM, Rabinov JD. Use of magnetic resonance imaging to differentiate optic neuritis and nonarteritic anterior ischemic optic neuropathy. *Ophthalmology* 2002;**109**:1679–84.
- Maeda N, Fujikado T, Kuroda T, *et al*. Wavefront aberrations measured with Hartmann-Shack sensor in patients with keratoconus. *Ophthalmology* 2002;**109**:1996–2003.
- Lapid-Gortzak R, Rosen S, Weitzman S, *et al*. Videokeratography findings in children with vernal keratoconjunctivitis versus those of healthy children. *Ophthalmology* 2002;**109**:2018–23.
- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;**299**:926–30.
- Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;**274**:645–51.
- Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988;**259**:1699–702.
- Power EJ, Tunis SR, Wagner JL. Technology assessment and public health. *Annu Rev Public Health* 1994;**15**:561–79.
- Harper R, Reeves B. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. *Invest Ophthalmol Vis Sci* 1999;**40**:1650–7.
- Harper R, Henson D, Reeves BC. Appraising evaluations of screening/diagnostic tests: the importance of the study populations. *Br J Ophthalmol* 2000;**84**:1198–202.
- Harper R, Reeves B. Reporting of precision of estimates for diagnostic accuracy: a review. *BMJ* 1999;**318**:1322–3.
- Simel DL, Feussner JR, DeLong ER, *et al*. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making* 1987;**7**:107–14.
- Egger M, Juni P, Bartlett C, CONSORT Group. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;**285**:1996–9.
- CONSORT statement. Available at www.consort-statement.org (accessed 19 February 2004).
- Moher D, Jones A, Lepage L, CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;**285**:1992–5.
- Sanchez-Thorin JC, Cortes MC, Montenegro M, *et al*. The quality of reporting of randomized clinical trials published in *Ophthalmology*. *Ophthalmology* 2001;**108**:410–5.