

## SCIENTIFIC REPORT

# Sample size in studies on diagnostic accuracy in ophthalmology: a literature survey

Frank Bochmann, Zoe Johnson, Augusto Azuara-Blanco

*Br J Ophthalmol* 2007;**91**:898–900. doi: 10.1136/bjo.2006.113290

**Aim:** To assess the sample sizes used in studies on diagnostic accuracy in ophthalmology.

**Design and sources:** A survey literature published in 2005.

**Methods:** The frequency of reporting calculations of sample sizes and the samples' sizes were extracted from the published literature. A manual search of five leading clinical journals in ophthalmology with the highest impact (*Investigative Ophthalmology and Visual Science*, *Ophthalmology*, *Archives of Ophthalmology*, *American Journal of Ophthalmology* and *British Journal of Ophthalmology*) was conducted by two independent investigators.

**Results:** A total of 1698 articles were identified, of which 40 studies were on diagnostic accuracy. One study reported that sample size was calculated before initiating the study. Another study reported consideration of sample size without calculation. The mean (SD) sample size of all diagnostic studies was 172.6 (218.9). The median prevalence of the target condition was 50.5%.

**Conclusion:** Only a few studies consider sample size in their methods. Inadequate sample sizes in diagnostic accuracy studies may result in misleading estimates of test accuracy. An improvement over the current standards on the design and reporting of diagnostic studies is warranted.

Diagnostic tests help the clinicians to make a diagnosis and to evaluate the severity of a disease. When using the information gained from diagnostic tests in clinical practice, their performance must be known. Therefore, the design and the reporting of studies on diagnostic accuracy should comply with methodological standards. Calculation of sample size plays an important part in the design of a diagnostic accuracy study, as it determines how precise the estimates for test accuracy should be for a particular diagnostic situation. This becomes even more important when a subgroup analysis is planned as the sample size in different subgroups has to be considered as well. If too few patients with and without the target condition have been evaluated, the indexes of accuracy (sensitivity and specificity) may be unstable. The quantitative instability can be appraised—for example, from CIs, which progressively narrow as sample size increases. In addition, investigators performing meta-analysis would greatly benefit from the reporting of relevant data such as sample size.

A recent publication assessing the studies on test accuracy published in 2002 in leading medical journals showed that only 4.7% of the studies reported a calculation of sample size.<sup>1</sup> It is possible that the literature on diagnostic tests in ophthalmology has similar limitations, and that the calculation of the sample size is not routinely reported. In this study, we investigated how often calculation of sample sizes was reported in leading ophthalmology journals.

## METHODS

Two reviewers independently and manually screened all issues of the five leading clinical journals in ophthalmology (*Investigative Ophthalmology and Visual Science*, *Ophthalmology*, *Archives of Ophthalmology*, *American Journal of Ophthalmology* and *British Journal of Ophthalmology*) published in 2005. Leading journals were defined according to their current impact index, excluding subspecialty journals. Diagnostic accuracy studies were identified (fig 1). Disagreement between reviewers was settled by consensus. From each report, data were extracted about the condition, type of test, number of participants, the prevalence and whether a prior calculation of sample size was described.

## RESULTS

Of the 1698 articles published in 2005, 43 were studies on diagnostic accuracy. Three articles focused on a screening test, and were excluded. Figure 1 shows a flow chart highlighting the process and the results of this literature survey. Table 1 shows the data extracted from each report.

The median sample size reported was 122.5, whereas the mean (SD) was 172.6 (218.9). The median prevalence of the target condition was 50.5%.

One study (2.5%) reported a prior calculation of sample size for a planned sensitivity and specificity of 80% with a 95% CI. In another study, consideration of sample size was stated on the basis of the estimates of the prevalence of the visual impairment. However, the sample size itself was not calculated.

Of the 40 articles appraised, 29 (72.5%) evaluated the diagnostic performance of imaging technology. Four articles reported results of clinical examination, seven about functional tests, two about results of laboratory tests and one study about the diagnostic performance of patient's history. About half of the studies (52.5%) on diagnostic accuracy evaluated imaging technologies in patients with glaucoma.

## DISCUSSION

In studies on diagnostic accuracy, the performance of a test to identify a target condition is determined. A misleading estimation of the test performance may result in unwanted consequences in clinical practice as it is difficult to assess how accurate a test might be. In diagnostic studies, the sample size plays a central role as it directly influences the width of the CIs. In studies with small sample sizes, the estimation of sensitivity and specificity may be imprecise as the CIs can be wide. If, for example, a new test correctly detects disease in 1140 of 1770 patients, the sensitivity would be 64.5% with a narrow CI of 0.632 to 0.667. If the same test is used to assess 177 patients, with the same sensitivity of 64.5% 114 patients would be diagnosed correctly, but the CI would be much wider (0.565 to 0.719). When subgroups are analysed separately, this effect may become even more important.<sup>2</sup> In the planning stage of a

**Abbreviation:** CONSORT, Consolidated Standards of Reporting Trials

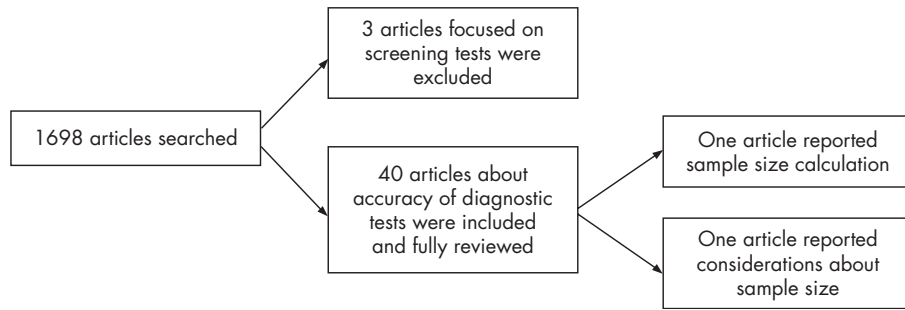


Figure 1 Flow chart of the literature survey.

study, investigators can influence this issue by calculating the sample size needed to obtain narrow CIs as it is a common practise in randomised trials.<sup>3,4</sup> In addition, reserves for patients who may drop out of the study can be considered at this stage. Tables have been recently published to ease the calculation of sample sizes in diagnostic test studies.<sup>5</sup> Using these tables to determine the number of cases that are necessary to assess a new test, the examiner must only specify the expected sensitivity of the test and the maximal distance of the lower confidence limit from this sensitivity. With this

information, the number of necessary cases can be easily extracted from the tables.

Sample size calculation is only one of the several aspects that are relevant in planning a study on a diagnostic test. Others include an independent, masked comparison with a reference standard, appropriate spectrum of included patients to whom the test will be applied, and absence of influence of the test results on the decision to perform the reference standard. Recently, tools have been designed to improve the standards and reporting of studies on diagnostic accuracy. The Standards

Table 1 Key features of 43 studies on accuracy of diagnostic tests

First author	Type of test	Prevalence, %	Sample size (n)	Screening	Calculation of sample size
Budenz DL	Imaging	37	172	No	No
Burkat CN	Clinical examination	32	202	No	No
Leung CK	Imaging	65	133	No	No
Morgan JE	Imaging	49	106	No	No
PHPRG	Test of visual function	53	122	No	Yes
Kook MS	Imaging	49	136	No	No
Ben Simon GJ	Imaging	27/23	131	No	No
Harasymowycz PJ	Imaging	7,2	303	Yes	/
Jeoung JW	Imaging	100	55	No	No
Brusini P	Functional test	32	123	No	No
Sandhu SS	Imaging	64	131	No	No
Galvao Filho RP	Imaging	63	108	No	No
Spry PGD	Functional test	31	48	No	No
Koc F	Clinical examination	36	85	No	No
Grus FH	Laboratory test	55	159	No	No
Leung CK	Imaging	27/36	111	No	No
Bourne RRA	Imaging	27/38	104	No	No
Bowd C	Imaging	56	164	No	No
Patterson AJ	Imaging	60	50	No	No
Lin PY	Clinical Examination	34	1361	No	No
Ishikawa H	Imaging	51	47	No	No
Essock EA	Imaging	50	134	No	No
VPSG	Functional test	32	1452	Yes	/
Leung CK	Imaging	70	89	No	No
Bengtsson B	Functional test	42	463	No	No
Burgansky Z	Imaging	53	89	No	No
Huang M	Imaging	47	189	No	No
Moreno J	Imaging	77	86	No	No
Gelman R	Imaging	38	32	No	No
Martinez JM	Clinical examination	100	147	No	No
Smith TS	Imaging	100	20	No	No
van Overdam KA	History	2,7	270	No	No
Graham SL	Functional test	51	436	No	No
Radhakrishnan S	Imaging	52	31	No	No
Pirbhai A	Imaging	52	223	No	No
Ferraro JG	Imaging	69	72	No	No
Wollstein G	Imaging	50	74	No	No
Medeiros FA	Imaging	53	166	No	No
Migliori S	Imaging	20	518	No	No
Cavallerano JD	Imaging	36	102	No	No
Davis JL	Laboratory test	14/34	78	No	No
Medeiros FA	Imaging	52	136	No	No
Huynh SC	Functional test	4,5	1765	Yes	/

PHPRG, Preferential Hyperacuity Perimetry Research Group; VPSG, The Vision in Preschoolers Study Group.

for Reporting of Diagnostic Accuracy checklist gives a framework to improve the accuracy of reporting of studies on diagnostic accuracy.<sup>6</sup> QUADAS (Quality Assessment of Diagnostic Accuracy Studies) has been recently designed to adequately assess the methodological quality of studies included in systematic reviews of diagnostic accuracy studies and consists of several items, one of which evaluates possible source of bias when patients withdraw from a study.<sup>7</sup> If a sample size has not been calculated or reported, it is unlikely that considerations about the dropout rate and its influence on the power of a study have been done. In this literature review of all diagnostic performance studies in ophthalmology published in 2005 in five leading journals, sample size calculation of a variety of tests was only available in one publication.

In an attempt to improve the quality of reporting of randomised controlled trials, the Consolidated Standards of Reporting Trials (CONSORT) has been introduced.<sup>8</sup> Sample size calculation is one of these key methodological items. An evaluation of all new randomised controlled trials published during 1999 in the journal *Ophthalmology* before and after the adoption of the CONSORT statement by the editors found an overall improvement in the quality of publications compared with the published of early 1990s, from 20% of studies reporting sample size calculation before the publication and use of CONSORT to 35% afterwards.<sup>9</sup> An evaluation of the quality of controlled clinical trials in glaucoma found a pre-estimation of sample size in only 15% of them (34 of 226).<sup>10</sup>

The beneficial effect of CONSORT was also observed in general medical journals,<sup>11, 12</sup> but further improvement would still be desirable. In a recent study assessing the quality of reports of randomised trials, only one quarter of 162 trials (35, 21.6%) did not describe a sample size calculation.<sup>13</sup> In an investigation comparing the quality of clinical trials among journals that endorse CONSORT, 85% of medical journals reported sample size calculation compared with 55% of specialist journals. Other key aspects of clinical trials such as methods of randomisation, allocation concealment and implementation, masking status, and use of intention-to-treat analysis were better reported in general medical journals than in specialist ones.<sup>14</sup>

It would seem that diagnostic test studies have poorer quality than trials evaluating interventions. In this survey of studies on diagnostic accuracy in ophthalmology, the number of studies reporting calculation of sample size was minimal. Only in two (5%) studies sample size was taken into consideration when the study was planned, and only 1 (2.5%) study reported an exact calculation of sample size considering the targeted sensitivity and specificity as well as a possible dropout rate. The consideration for sample size calculation seems to be worse in our survey than in current medical journals.<sup>1</sup> Other important methodological aspects of diagnostic performance studies are often missing.<sup>15-19</sup> Harper and Reeves<sup>19</sup> highlighted that only two of 16 articles reported complete precision for the estimates of diagnostic accuracy and CIs.

An improvement in the reporting of methodological aspects of studies would facilitate systematic reviews and meta-analysis of the increasing number of publications on diagnostic tests. The assessment of data from studies on diagnostic tests to quantify bias and other sample size-related effects relies on the information about the methods. Only if the study design is entirely transparent and includes power calculation for testing hypotheses, it is possible to investigate whether there is any

sample size-related effect, which is especially important to know for reviews on test accuracy.<sup>20</sup> A possible contributing factor for under-reporting methodological characteristics can be limitation of space in journals or undervaluing the importance of the Methods section by reviewers and editors.

In conclusion, sample size calculations should be a part of the methods and published report of diagnostic performance studies. Currently, they are not being reported in the ophthalmology literature.

#### Authors' affiliations

Frank Bochmann, Department of Ophthalmology, Lucerne, Switzerland  
Zoe Johnson, Augusto Azuara-Blanco, Department of Ophthalmology, Aberdeen, UK

Competing interests: None.

Correspondence to: Dr F Bochmann, Department of Ophthalmology, Cantonal Hospital of Lucerne, CH-6000, Lucerne 16, Switzerland; frank.bochmann@ksh.ch

Accepted 2 February 2007

Published Online First 14 February 2007

#### REFERENCES

- 1 Bachman L, Puhan M, Riet G, *et al*. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006;**332**:1127-9.
- 2 Irwig L, Bossuyt PM, Glasziou P, *et al*. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002;**324**:669-71.
- 3 Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;**365**:1348-53.
- 4 Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;**21**:1525-37.
- 5 Flauhaut A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;**58**:859-62.
- 6 Bossuyt PM, Reitsma JB, Bruns DEE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2004;**326**:41-4.
- 7 Whiting P, Rutjes AW, 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.
- 8 Begg C, Cho M, Eastwood S. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;**276**:637-9.
- 9 Sanchez-Torin JC, Cortes MC, Montenegro M, *et al*. The quality of reporting of randomized clinical trials published in ophthalmology. *Ophthalmology* 2001;**108**:410-15.
- 10 Javier L, Martinez-Sanz F, Prieto-Salceda D, *et al*. Quality of controlled clinical trials on glaucoma and intraocular high pressure. *J Glaucoma* 2005;**14**:190-5.
- 11 Mills EJ, Wu P, Gagnier J, *et al*. The quality or randomised trial reporting in leading medical journals since the revised CONSORT statement. *Contemp Clin Trials* 2005;**26**:480-7.
- 12 Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;**285**:1992-5.
- 13 Le Henanff A, Giraudeau B, Baron G, *et al*. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006;**295**:1147-51.
- 14 Mills E, Wu P, Gagnier J, *et al*. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *J Clin Epidemiol* 2005;**58**:662-7.
- 15 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.
- 16 Harper R, Reeves B. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. *Invest Ophthalmol Vis Sci* 1999;**40**:1650-7.
- 17 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.
- 18 Siddiqui R, Azuara-Blanco A, Burr J. The quality of reporting of diagnostic accuracy studies in ophthalmology. *Br J Ophthalmol* 2005;**89**:261-5.
- 19 Harper R, Reeves B. Reporting of precision of estimates of diagnostic accuracy: a review. *BMJ* 1999;**318**:1322-3.
- 20 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882-93.