

RESEARCH ARTICLE

Open Access

# The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review

Brian H Willis<sup>1\*</sup> and Muireann Quigley<sup>2</sup>

## Abstract

**Background:** Over the last decade there have been a number of guidelines published, aimed at improving the quality of reporting in published studies and reviews. In systematic reviews this may be measured by their compliance with the PRISMA statement. This review aims to evaluate the quality of reporting in published meta-analyses of diagnostic tests, using the PRISMA statement and establish whether there has been a measurable improvement over time.

**Methods:** Eight databases were searched for reviews published prior to 31<sup>st</sup> December 2008. Studies were selected if they evaluated a diagnostic test, measured performance, searched two or more databases, stated the search terms and inclusion criteria, and used a statistical method to summarise a test's performance. Data were extracted on the review characteristics and items of the PRISMA statement. To measure the change in the quality of reporting over time, PRISMA items for two periods of equal duration were compared.

**Results:** Compliance with the PRISMA statement was generally poor: none of the reviews completely adhered to all 27 checklist items. Of the 236 meta-analyses included following selection: only 2(1%) reported the study protocol; 59(25%) reported the searches used; 76(32%) reported the results of a risk of bias assessment; and 82 (35%) reported the abstract as a structured summary. Only 11 studies were published before 2000. Thus, the impact of QUOROM on the quality of reporting was not evaluated. However, the periods 2001-2004 and 2005-2008 (covering 93% of studies) were compared using relative risks (RR). There was an increase in the proportion of reviews reporting on five PRISMA items: eligibility criteria (RR 1.13, 95% CI 1.00 - 1.27); risk of bias across studies (methods) (RR 1.81, 95% CI 1.34 - 2.44); study selection results (RR 1.48, 95% CI 1.05 - 2.09); results of individual studies (RR 1.37, 95% CI 1.09 - 1.72); risk of bias across studies (results) (RR 1.65, 95% CI 1.20 - 2.25).

**Conclusion:** Although there has been an improvement in the quality of meta-analyses in diagnostic research, there are still many deficiencies in the reporting which future reviewers need to address if readers are to trust the validity of the reported findings.

## Background

Systematic reviews have become increasingly important in diagnostic research [1,2]. With the development of new statistical methods used to aggregate primary studies [3,4], and increasing numbers of diagnostic reviews appearing in the literature [1,2], the need for high quality meta-analyses of diagnostic tests has, perhaps, never been greater.

Unfortunately, like all other types of systematic reviews, they are prone to a number of shortcomings. These may arise at a review level, due to inaccessibility to all pertinent studies [5], failings in the selection process [6], or heterogeneity [7,8], which often blights reviews of diagnostic tests [2]. They may also arise at a primary study level, due to flaws in the design, execution and reporting of the component studies [9,10].

To help identify and mitigate potential weaknesses, quality assessment of the primary studies has become an intrinsic element of the review process [11]. Following the publication of the Standards for the Reporting of

\* Correspondence: brian.willis@manchester.ac.uk

<sup>1</sup>Biostatistics, University of Manchester, Jean McFarlane Building, Oxford Road, Manchester, M13 9PL, United Kingdom

Full list of author information is available at the end of the article

Diagnostic accuracy studies (STARD) statement [10], which set out to improve the reporting of primary research on diagnostic tests, the assessment of quality has been recently formalized. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool [12] is a generic tool that covers the major domains affecting diagnostic study validity, thus placing quality assessment on a firmer ground and allowing inter-study comparison.

There have been parallel developments in meta-analysis. As major undertakings of work, their results may be influential to health care providers, researchers, and decision makers. Thus, the need for a consistent framework of reporting was recognised. This led to the compilation of the Quality of Reporting of Meta-analyses (QUOROM) statement [13], which was aimed at improving the quality of published meta-analyses of randomised controlled trials. Recently, the QUOROM statement [13] has been superseded by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [14]. This was in response to developments in systematic review methodology and to widen the scope beyond randomised controlled trials. Currently this is the standard for investigators when reporting their findings and also provides a benchmark by which meta-analyses may be appraised.

Meta-analyses of diagnostic tests will potentially have an increasing role in healthcare as decision makers look to the evidence before implementing new diagnostic technologies. It is important that such analyses provide reliable results and this, in part, is determined by the quality of reporting [2,14,15].

As meta-analysis in diagnostic research has been developing for nearly two decades, it is an appropriate time to assess the overall quality of reporting of meta-analyses of diagnostic test studies. Furthermore it is of interest to know whether there has been a measureable improvement in the quality of reporting. Thus, the objective of this systematic review was to examine the quality of reporting of published meta-analyses of diagnostic tests studies, by their compliance with the PRISMA statement and to assess whether there is evidence of an improvement in the overall quality of reporting. This review was part of a wider investigation into meta-analyses of diagnostic test accuracy studies that has been published elsewhere [16].

## Methods

### Data sources and searches

The electronic databases, Medline, CINAHL, Cochrane library (including the Cochrane Database of Systematic Reviews, DARE, Health Technology Assessment Database and NHS Economic Evaluation Database) EMBASE, PsychInfo, Global health, HMIC, and AMED were all searched for relevant reviews (example search algorithms are listed in Additional file 1). The searches

were conducted initially in September 2008 and updated in September 2009. The cut off for inclusion of the meta-analyses was December 31<sup>st</sup> 2008.

### Selection criteria

For the purpose of this review, the term 'meta-analysis' is taken to mean a special type of systematic review, in which standard systematic review methodology has been followed and a quantitative summary of the results has been derived.

All citations retrieved from the electronic searches were subject to a six-step algorithm for inclusion in the review. The title and abstracts of the citations were initially screened using step 1, before retrieving the full text. Steps 2 to 6 were then applied to the full text of the articles, where non-compliance with any of the steps resulted in the article's exclusion.

The steps in the inclusion criteria were as follows:

1. Is the citation an original study of a diagnostic/screening test?
2. Was one of the objectives to measure the test(s) performance?
3. Were two or more of the major databases searched to identify the relevant articles? Examples included MEDLINE, EMBASE, Cochrane, and CINAHL.
4. Were the search terms explicitly stated?
5. Were the inclusion criteria explicitly stated?
6. Was at least one statistical method used to summarize the overall test performance across the primary studies?

The first criterion (step 1) requires clarification as it encompasses a number of terms. The word original was defined here as a primary or secondary (systematic review) evaluation of the technology. Narrative reviews, editorials and commentaries were excluded, although primary studies were still included at this stage. A diagnostic/screening test was defined as a technology aimed at identifying a target disorder, which was present at the time of testing. Target disorders were considered to be pathological processes and not related to a success or failure of an intervention, such as successful placing of stents. Furthermore, technologies, which predicted the future occurrence of a target disorder, which was not present at the time of testing, were not considered diagnostic technologies in this review.

The question of what constitutes a systematic review is, to a degree, open to debate. But the view taken here is that, an important part of the systematic review process is that investigators should make every effort to identify *all* the relevant studies. Since research has demonstrated that to search a single database runs a high risk of missing relevant studies [17,18], a minimum requirement that investigators should have searched two or more databases was imposed in the inclusion criteria.

The first author (BHW) screened and applied the inclusion criteria to all the citations and reviews. The second author (MQ) independently screened and applied the inclusion criteria to a random sample of 10% of all the citations, and discrepancies were decided by consensus agreement.

#### Data extraction and quality assessment

Data were abstracted on the following items: publication year; objective; diagnostic test; target disorder; search terms (or whether an algorithm was given); databases searched; inclusion criteria (and whether they were made explicit); process of data extraction; method used to assess quality including QUADAS [12]; presence of heterogeneity; and responses to the PRISMA statement [14].

The PRISMA statement was used to evaluate the overall quality of reporting of the meta-analyses and consists of a twenty seven-point checklist. To indicate the degree of compliance, each checklist item was assigned one of three responses: 'yes' for total compliance; 'partial' for partial compliance; and 'no' for non-compliance. As a large number of checklist items may not be satisfactorily answered with a binary response (yes/no), the intermediate category (partial) was included to represent the situation where a review had satisfied some, but not all of the criteria for an individual item.

Extraction of data was performed by BHW and independently on a random sample of 10%, by MQ. Discrepancies were decided by consensus agreement.

#### Data synthesis and analysis

For making comparisons over time, two cohorts of equal duration were compared.

Ideally, these would have been periods either side of the introduction of the QUOROM statement (published in November 1999) so that the impact it had on the quality of reporting could be evaluated. Unfortunately, the sample size of studies prior to its introduction was too small to address this question adequately. Hence, both periods were chosen to be later than the publication of QUOROM to avoid it having a heterogeneous effect on one of the cohorts (see Results).

When analysing responses to the PRISMA statement [14], the relative risk or risk ratio was used as the summary statistic for sub-group comparisons [19,20]. If the relative risk was undefined, then Fisher's exact test was used [21]. Statistical significance was set at  $p < 0.05$ . Where appropriate, the kappa statistic was used to assess the level of agreement between the reviewers [22,23], and interpretation was made using accepted criteria [23,24]. For all statistical analyses, EXCEL and the programming software R (version 2.10.1) were used.

## Results

Over 4000 unduplicated citations were retrieved from the electronic searches and after applying the selection criteria 236 articles were included for appraisal (Figure 1). For a list of the included reviews see Additional file 2. As measured by the kappa score, agreement between the reviewers on review selection was excellent, with a kappa score of 0.86. Disagreements were over the stage in which certain reviews were excluded, not the decision whether to exclude. Thus, there was 100% agreement on which reviews to include. For the PRISMA items, the median kappa score between the two reviewers was 0.88 (range: 0.66 - 1).

#### Characteristics of the included reviews

The majority of the included reviews were reported in specialist journals (78%), with 39 (17%) being published in radiology journals. Thirty meta-analyses were reported in general medical journals, such as the BMJ or Annals of Internal Medicine and 15 were commissioned health technology assessments (HTA).

Figure 2 illustrates the number of reviews per publication year in the included set. Nearly 93% (219 meta-analyses) were published after the year 2001.

A wide range of diagnostic tests featured in the reviews, with imaging technologies (47%) being the most common category of test evaluated. These included 45 reviews on ultrasound tests, 42 on computer tomography and 34 on magnetic resonance imaging. Nearly a fifth of reviews investigated diagnostic tests found in the clinical examination and 35 (15%) evaluated biochemical tests.

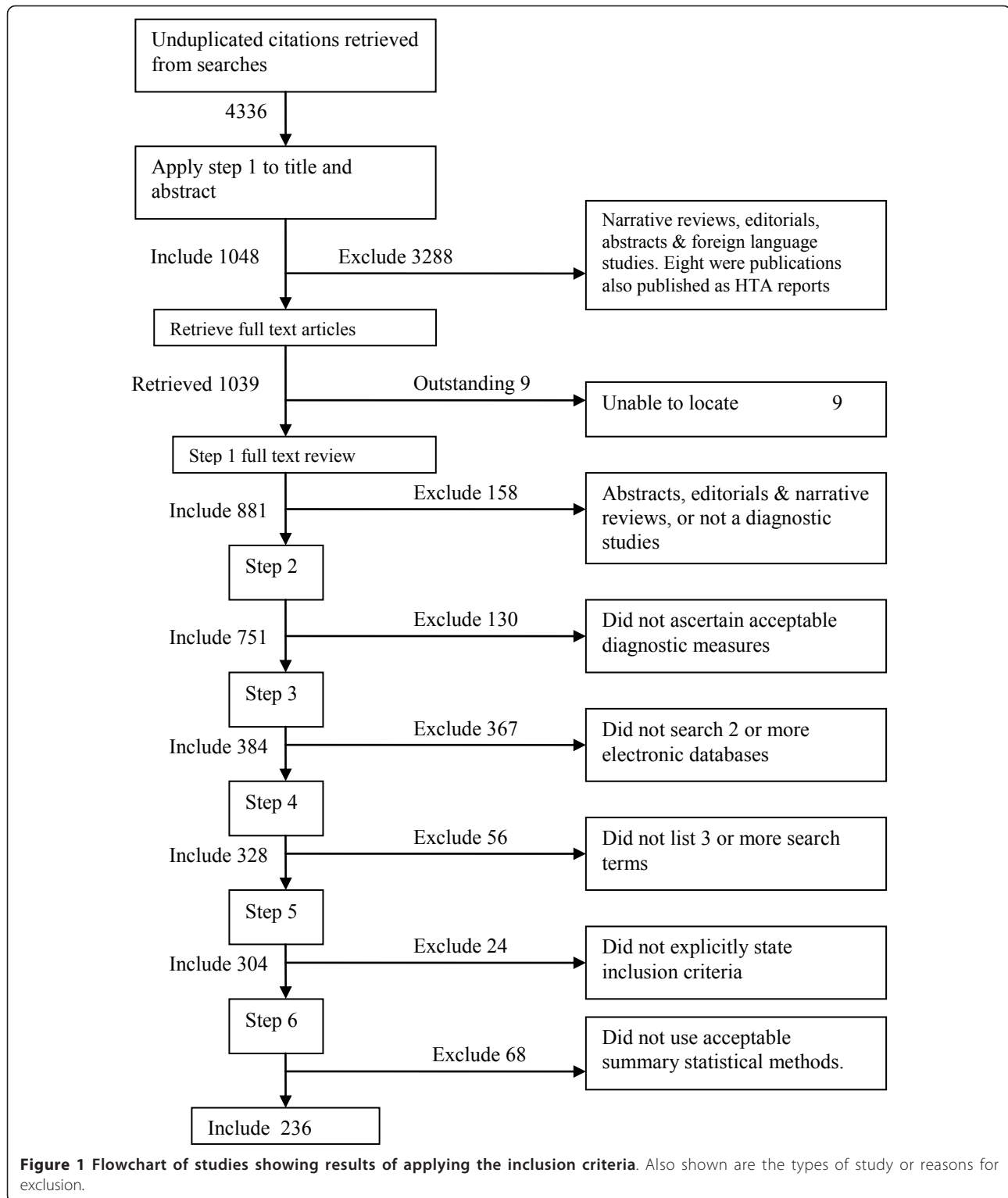
An equally diverse spectrum of target disorders was covered by the reviews. Cancer (25%) and infection (20%) were the two most frequent categories of target disorder in the reviews; 20 reviews evaluated diagnostic tests used to detect tuberculosis alone (the characteristics of the individual reviews are detailed elsewhere [16]).

#### Quality of reporting and compliance with PRISMA (table 1)

Generally compliance with PRISMA was poor: none of the 27 checklist criteria was complied with by all reviews, nor did any one review fulfil all 27 items of PRISMA.

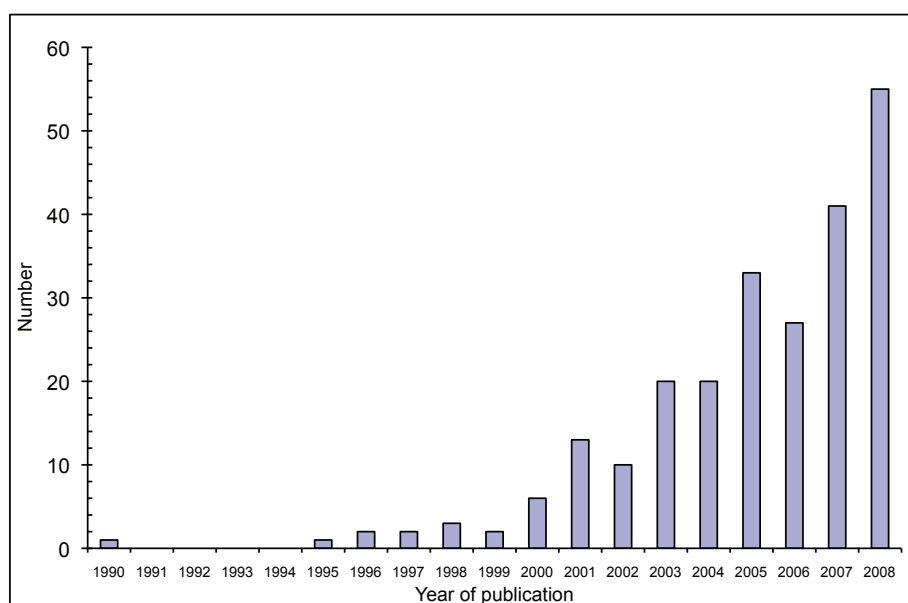
All of the reviews did at least achieve partial compliance in nearly a third (8/27) of the criteria.

However, some of these items relate to the definition of a systematic review and coincide with the eligibility criteria for this study. Thus a review's partial compliance with PRISMA items 6,7,8 and 14 follows directly from its eligibility for inclusion in this study. Full details of applying the PRISMA statement to each of the included meta-analyses are given in Additional file 3.



There were individual PRISMA items which were adequately reported in over 95% of the reviews and these included: the rationale for the review; the description of the data sources; the measures used to summarize the

primary studies; the methods used to aggregate the data; the results of the meta-analysis; a summary of the main findings and the concluding remarks in the discussion.



**Figure 2** Number of included meta-analyses per year of publication.

**Table 1** Summary of the results of applying the PRISMA statement to the individual meta-analyses

PRISMA statement results			Yes	Partial	No
<b>Title</b>	1	Title	221 (93%)	13 (5%)	2 (0%)
<b>Abstract</b>	2	Structured summary	82 (35%)	147 (62%)	7 (3%)
<b>Introduction</b>	3	Rationale	233 (98%)	3 (1%)	0 (0%)
<b>Methods</b>	4	Objectives	143 (60%)	90 (38%)	3 (1%)
	5	Protocol and registration	2 (1%)	35 (15%)	199 (84%)
	6	Eligibility criteria	209 (88%)	27 (11%)	0 (0%)
	7	Information sources	235 (99%)	1 (0%)	0 (0%)
	8	Search	59 (25%)	177 (75%)	0 (0%)
	9	Study selection	134 (57%)	30 (13%)	72 (30%)
	10	Data collection process	149 (63%)	16 (7%)	71 (30%)
	11	Data items	127 (54%)	64 (27%)	45 (19%)
	12	Risk of bias in individual studies	103 (43%)	79 (33%)	54 (23%)
	13	Summary measures	235 (99%)	0 (0%)	1 (0%)
	14	Synthesis of results	234 (99%)	2 (1%)	0 (0%)
	15	Risk of bias across studies	152 (64%)	30 (13%)	54 (23%)
	16	Additional analyses	101 (43%)	12 (5%)	123 (52%)
<b>Results</b>	17	Study selection	112 (47%)	91 (38%)	33 (14%)
	18	Study characteristics	132 (56%)	94 (40%)	10 (4%)
	19	Risk of bias within studies	76 (32%)	84 (36%)	76 (32%)
	20	Results of individual studies	162 (68%)	36 (15%)	39 (16%)
	21	Synthesis of results	235 (99%)	1 (0%)	0 (0%)
	22	Risk of bias across studies	136 (57%)	35 (15%)	65 (27%)
	23	Additional analyses	90 (38%)	24 (10%)	122 (51%)
<b>Discussion</b>	24	Summary of evidence	232 (98%)	4 (2%)	0 (0%)
	25	Limitations	139 (59%)	38 (16%)	59 (25%)
	26	Conclusions	232 (98%)	4 (2%)	0 (0%)
<b>Funding</b>	27	Funding	114 (48%)	0 (0%)	122 (51%)

In over 84% (199/236) of publications the existence of a review protocol, or whether the review had been registered, was not reported. The test being investigated was usually stated in the objective, but, in general, the reporting in the included reviews fell short of stating clear focused objectives. In six reviews the objective was not clear at all (see Additional file 4)

Search algorithms used to locate the primary studies were reported in only 23% (55/236) of the meta-analyses. In a number of the included reviews, the reference lists of the primary studies were searched, for further citations. Nevertheless, in over a third, the searches were confined to only two electronic databases, principally Medline and EMBASE, thereby diminishing the likelihood of achieving completeness (see Additional file 4).

In all of the reviews, the eligibility criteria were explicitly stated in the methods; however, there was significant variability in the clarity of terminology and in only 67 (28%) reviews could they be feasibly described as being algorithmic. The process of selection was also variable: in only 96 (41%) reviews was it clearly described that at least two investigators had independently screened, selected and abstracted data from the primary studies (see Additional file 4).

From the PRISMA analysis, screening and eligibility results were displayed in the form of a flow chart in 113 (48%) reviews. However, on closer inspection, only 44 (19%) of these meta-analyses gave sufficient information on the decisions behind including or excluding the primary studies. Either the reason was not given, or vague terms, such as 'not relevant', were used without elaboration (see Additional file 4).

The descriptions of data abstraction also varied widely. In 103 (43%) reviews, the authors were unclear on either the abstraction process or the data items abstracted (Additional file 4). Many reviews listed no more than abstracting data to complete 2 × 2 tables, when clearly other information had been sought from the primary studies. In eighteen reviews, data abstraction was not reported at all.

For more than half of the reviews (123/236), additional analyses in the methods, such as meta-regression or sub-group analyses, were not described. In a similar proportion there was neither reporting of the results of any additional analyses (52%) nor whether the review had been funded (52%).

#### **Comparison of meta-analyses published between 2001-2004 and 2005-2008**

To determine whether there had been an improvement in the quality of reporting over time, cohorts of meta-analyses published over periods of similar duration were compared. The question of what impact the QUOROM statement had on the quality of reporting was not

addressed, owing to the small sample size of the studies published before QUOROM (11 included studies).

Therefore, the period 2001-2004 (63 reviews) was compared with 2005-2008 (156 reviews) for full compliance in each of the PRISMA criteria. Although not pre-specified at the inception stage of the review, the periods 2001-2004 and 2005-2008 were chosen as they not only capture the vast majority of reviews, both are later than the publication of the QUOROM statement (November 1999). The advantage of this latter point is that the earlier cohort does not contain studies published both pre and post publication of QUOROM and so is less likely to be heterogeneous.

Over the two periods, there was a significant improvement demonstrated in five of the criteria and these are illustrated in Figure 3. In only one of the items (abstract) was the reporting poorer in the later period (Figure 3). In the previous decade there also seems to have been an improvement in the number of reviews reporting on the assessment of the quality of the primary studies. Over 50% of the reviews published in 2000 reported no formal quality assessment of the primary studies, compared with 20% in 2008. This has coincided with a greater number of investigators using the QUADAS tool for quality assessment, although other methods of quality assessment continue to be used (see Figure 4).

As already noted, some items of the PRISMA statement feature in the eligibility criteria for this review. The effect is to potentially increase compliance with these items across both periods and bias the relative risk towards 1. This explains the relative risks and narrow confidence intervals observed in items 7 and 14 of Figure 3.

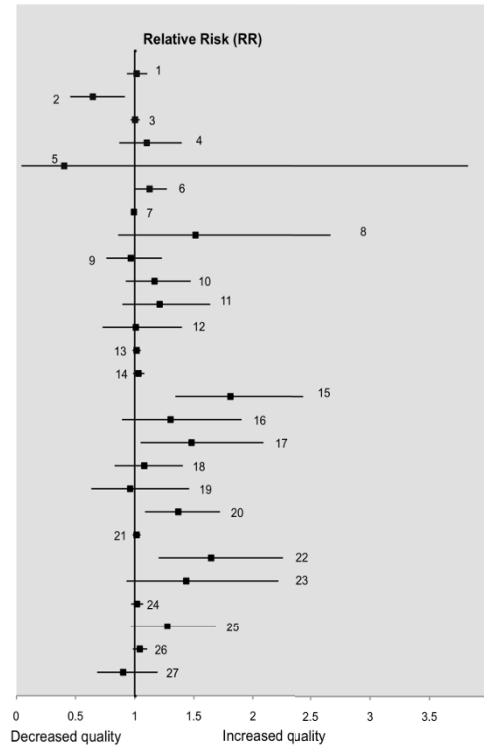
#### **Comparison of HTA reports with other meta-analyses**

In the included reviews there were 15 Health technology assessment (HTA) reports. These are often commissioned reports conducted by experienced reviewers with a remit of providing 'high quality research information for decision makers' [25] and are not constrained by the word count restrictions imposed by many journals. As might be expected, the reporting in these was of a higher quality. In nine of the PRISMA criteria, there was a significantly higher proportion of HTA reviews adequately reporting on these, compared with the other meta-analyses (see Figure 5). For the individual results of applying PRISMA, refer to Additional file 3.

#### **Sensitivity analysis**

The HTA reports could potentially confound the differences in the quality observed between years 2001-2004 and 2005-2008, particularly if a disproportionate number were in the second cohort (2005-2008). There were 10

#	2001-2004	2005-2008	RR (95% CI)
1	58/63	146/156	1.02 (0.93-1.11)
2	30/63	48/156	0.65 (0.46-0.92)
3	62/63	154/156	1.00 (0.97-1.04)
4	37/63	101/156	1.10 (0.87-1.40)
5	1/63	1/156	0.40 (0.04-3.82)
6	52/63	145/156	1.13 (1.00-1.27)
7	63/63	155/156	0.99 (0.98-1.01)
8	12/63	45/156	1.51 (0.86-2.67)
9	38/63	91/156	0.97 (0.76-1.23)
10	37/63	107/156	1.17 (0.93-1.47)
11	29/63	87/156	1.21 (0.90-1.64)
12	28/63	70/156	1.01 (0.73-1.40)
13	62/63	156/156	1.02 (0.98-1.05)
14	61/63	156/156	1.03 (0.99-1.08)
15	27/63	121/156	1.81 (1.34-2.44)
16	22/63	71/156	1.30 (0.89-1.90)
17	24/63	88/156	1.48 (1.05-2.09)
18	34/63	91/156	1.08 (0.83-1.41)
19	21/63	50/156	0.96 (0.63-1.46)
20	36/63	122/156	1.37 (1.09-1.72)
21	62/63	156/156	1.02 (0.98-1.05)
22	26/63	106/156	1.65 (1.20-2.25)
23	18/63	64/156	1.44 (0.93-2.22)
24	61/63	154/156	1.02 (0.97-1.07)
25	31/63	98/156	1.28 (0.97-1.69)
26	60/63	155/156	1.04 (0.98-1.11)
27	34/63	76/156	0.90 (0.80-1.19)

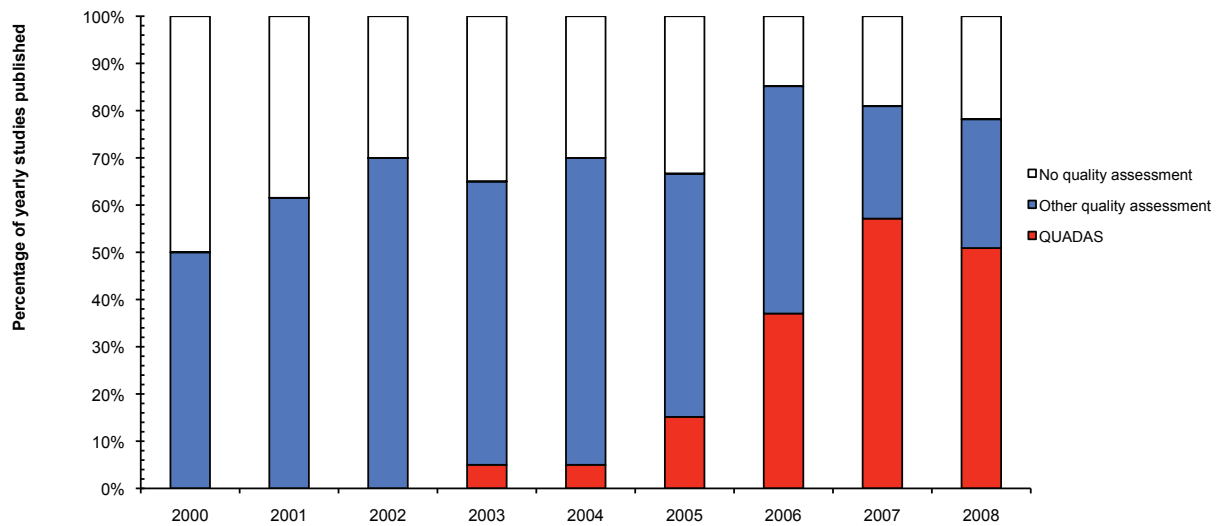


**Figure 3 Comparison of periods 2001-04 and 2005-08 by compliance with the PRISMA statement.** The numbered items (#) correspond to the PRISMA item numbers (see table 1). RR (95% CI) denotes the relative risk with the associated 95% confidence interval.

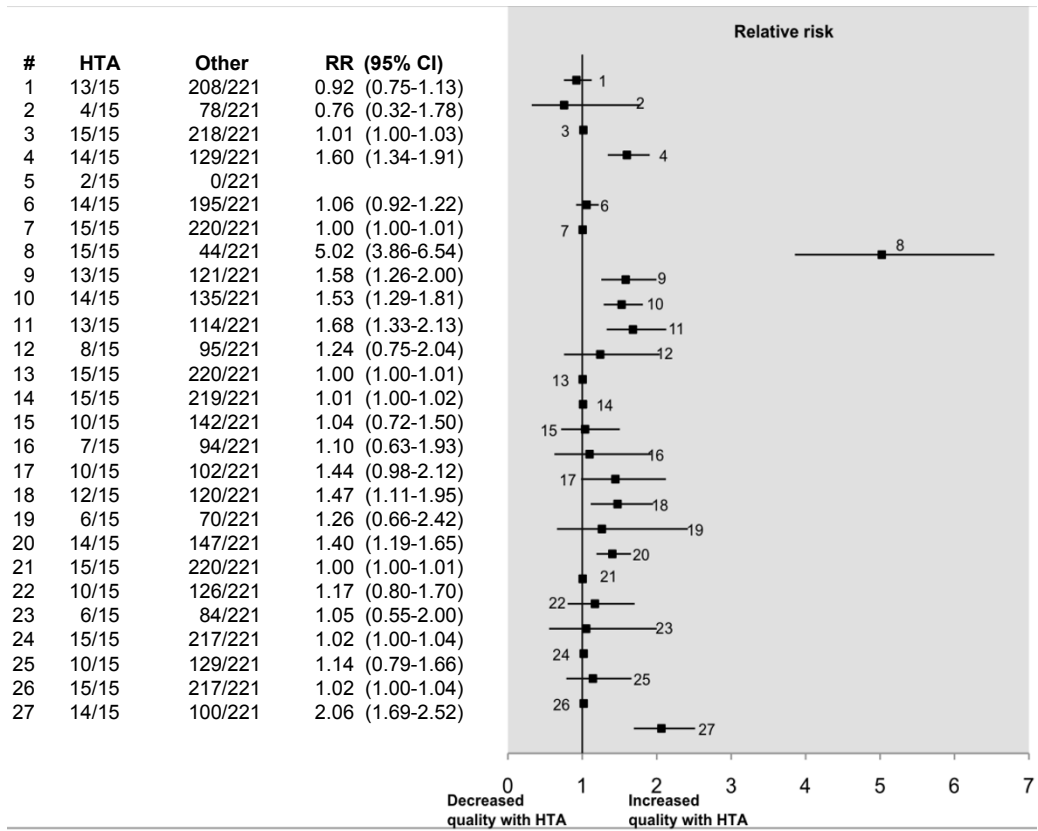
HTA reports published between years 2005-2008, and 4 published between 2001-2004 (ratio of 2.5:1). Thus, compared with the overall distribution of meta-analyses, (156 reviews published between 2005-2008 and 63 published

between 2001-2004, a ratio of 2.47:1) there were a similar proportion of HTA reports in the later cohort.

To test whether this affected the results, a sensitivity analysis was conducted, where the HTA reports were



**Figure 4 Changing pattern of quality assessment in meta-analyses of diagnostic tests.** Comparison of the percentage of reviews published per year using the QUADAS tool, other forms of quality assessment and no quality assessment. Earlier years not included due small sample sizes (around 2 studies per year).



**Figure 5 Comparison of HTA reviews with other reviews using PRISMA.** In nine PRISMA items the HTA reviews were significantly better reported than in other types of reviews. The numbered items (#) correspond to the PRISMA item numbers (see table 1). In item #5, the relative risk was undefined, but, Fisher's exact test demonstrated a significant difference, ( $p = 0.0038$ ), in favour of the HTA reports. RR (95% CI) denotes the relative risk with the associated 95% confidence interval.

excluded from the data and the analysis repeated. Despite the exclusion of the HTA reports, the results remained robust and the only significant PRISMA items were those in the earlier analysis, that is, items 2, 6, 15, 17, 20 and 22 (Figure 6).

## Discussion

### Summary of evidence

The number of meta-analyses in diagnostic research is increasing annually (Figure 2). From this review it seems that, not only are investigators assessing the quality of primary studies more often (Figure 4), but the quality of reporting of the meta-analyses is also improving. A large part of this is likely to result from the publishing of a number of guidelines over the last decade. The introduction of STARD [10], for primary studies on diagnostic tests, was mirrored by QUOROM [13] and then later PRISMA [14], for systematic reviews and meta-analyses. These have undoubtedly helped standardise the quality of reporting.

The need for high quality studies in health care, which include meta-analyses of diagnostic tests, has probably

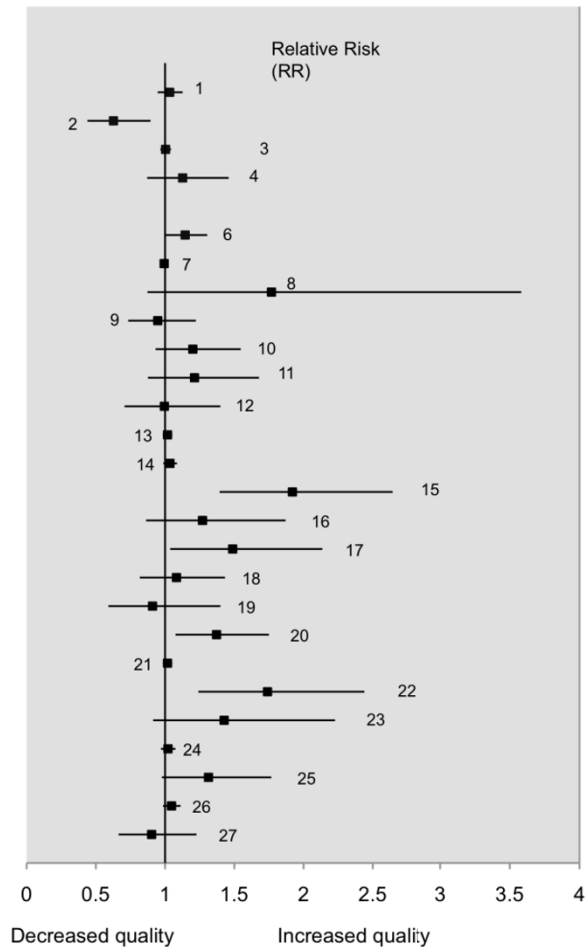
never been greater as the emphasis on evidence-based decision-making increases. A corollary of this is a drive for more complete and transparent reporting of how a review has been designed and conducted, so that stakeholders may make informed decisions on the validity of the findings [20].

Although the quality of reporting of randomised controlled trials has recently been reported by other authors [20], we are not aware of such an assessment being made on the quality of reporting of published systematic reviews of diagnostic test studies. There was no restriction on the type of journal, test, or target disorders being sought. With a view of capturing those reviews, which were unequivocally meta-analyses that had followed systematic review principles, the inclusion criteria were perhaps tighter than reported in some other reviews [26].

Yet, despite the tighter inclusion criteria, this review found their reporting was, in general, far from adequate and having a number of shortcomings. Using the PRISMA statement [14] it was shown that, although there has been some improvement in certain aspects of reporting over the last decade, for nine PRISMA items



#	2001-2004	2005-2008	RR (95% CI)
1	54/59	138/146	1.03 (0.95-1.13)
2	29/59	45/146	0.63 (0.44-0.89)
3	58/59	144/146	1.00 (0.97-1.04)
4	33/59	92/146	1.13 (0.87-1.46)
5	0/59	0/146	
6	48/59	136/146	1.14 (1.01-1.30)
7	59/59	145/146	0.99 (0.98-1.01)
8	8/59	35/146	1.77 (0.87-3.58)
9	35/59	82/146	0.95 (0.73-1.22)
10	33/59	98/146	1.20 (0.93-1.55)
11	26/59	78/146	1.21 (0.88-1.68)
12	26/59	64/146	0.99 (0.71-1.40)
13	58/59	146/146	1.02 (0.98-1.05)
14	57/59	146/146	1.04 (0.99-1.09)
15	24/59	114/146	1.92 (1.39-2.64)
16	21/59	66/146	1.27 (0.86-1.87)
17	22/59	81/146	1.49 (1.04-2.14)
18	31/59	83/146	1.08 (0.82-1.43)
19	20/59	45/146	0.91 (0.59-1.40)
20	33/59	112/146	1.37 (1.08-1.75)
21	58/59	146/146	1.02 (0.98-1.05)
22	23/59	99/146	1.74 (1.24-2.44)
23	17/59	60/146	1.43 (0.91-2.23)
24	57/59	144/146	1.02 (0.97-1.07)
25	28/59	91/146	1.31 (0.98-1.77)
26	56/59	145/146	1.05 (0.98-1.11)
27	30/59	67/146	0.90 (0.66-1.23)



**Figure 6 Sensitivity analysis.** The HTA reports have been removed from the sample to check robustness of results. The numbered items (#) correspond to the PRISMA item numbers (see table 1). In item #5, the relative risk was undefined, but, Fisher's exact test demonstrated no significant difference, ( $p = 1.00$ ), between the two periods. Overall there was no change in the significance of results in any of the 27 PRISMA items. RR (95% CI) denotes the relative risk with the associated 95% confidence interval.

less than half the meta-analyses were fully compliant. Some of these may reflect inadequacies in the reporting process, rather than flaws in the design or conduct of the review. Nonetheless, these latter flaws cannot be discounted [27-30]. For example, in 70% of the meta-analyses heterogeneity was reported as being present, yet it was investigated in less than half: it is unlikely that this difference is explained entirely by deficiencies in reporting.

Despite these deficiencies it should be borne in mind that, in addition to the spate of guidelines aimed at improving the quality of reporting [2,10,13], there have been a number of developments in the statistical methodology used in meta-analyses of diagnostic test accuracy studies [3,4,31-33]. Increased dissemination of these methods should lead to increased precision of the summary estimates on a test's performance, which in some instances, should enhance the validity of the reported findings.

### Limitations

There are limitations to this review. Like other recent reviews on the quality of reporting [20], the review process used here, predominately consisted of study selection and data extraction by a single reviewer. A second reviewer performed independent verification of the study selection and data extraction process on a 10% random sample. Although inter-observer agreement demonstrated 'good to excellent agreement' for selection and data extraction, this method is still more likely to yield errors than the preferred method of complete, independent replication of both steps by the two reviewers.

The definition of a systematic review is open to interpretation. Chalmers and Altman described a systematic review as a review, which had been prepared using a 'systematic approach to minimising biases and random errors', with the different components of the process

being documented in the 'methods section' [34]. The Cochrane Handbook for Systematic Reviews on Interventions states that systematic reviews possess a number of 'key characteristics' [35]. The eligibility criteria used here were in line with these key characteristics and also coincide with those items of PRISMA that help define a systematic review. It follows that a review's inclusion guarantees it being at least partially compliant with four items of PRISMA. Thus, the PRISMA statement captures more than the quality of reporting in systematic reviews, it captures the essence of what defines a systematic review.

Clearly, if some of the eligibility criteria are relaxed this would not only lead to additional studies being included it would increase the likelihood of poorer compliance with PRISMA. However, it then raises the question of whether these additional studies are truly systematic reviews if they do not possess some of the 'key characteristics' described by, perhaps, the leading authority on systematic reviews [35]. As a result, the criteria used here ensure that there was no ambiguity on the type of reviews included. It also serves to reinforce the conclusion that the quality of reporting in diagnostic systematic reviews is far from adequate if deficiencies in reporting are still being observed in a 'more selective' and therefore 'higher quality' group of studies.

This review is an analysis of the quality of reporting of meta-analyses of diagnostic test studies published over the previous decade. The issue of actual quality of the reviews, as opposed to the quality of reporting, is more difficult to assess, since the PRISMA statement is not a direct measure of quality. Furthermore it is generic checklist aimed at improving the reporting of all types of systematic reviews and does not contain some of the more specific nuances of diagnostic test reviews.

However, it does contain items on the reporting of the process that lie behind the synthesis of a systematic review, and the robustness of this process certainly contributes to the overall quality. There may still be inaccuracies between the reporting and how the review was actually conducted, as has been shown with some primary studies [27-30], nevertheless, it seems unreasonable not to consider this as providing some measure of the overall quality of the meta-analyses included in this review.

The included reviews represent a diverse group of meta-analyses evaluating a range of diagnostic tests over different settings and target disorders. It is unlikely that quality of reporting will be completely independent of variation in these and so the effect heterogeneity has on the results needs to be considered.

## Conclusion

In summary, this review demonstrates that the quality of reporting of meta-analyses has measurably improved

over the previous decade. Unfortunately, there are still many deficiencies, identified in the reporting of meta-analyses of diagnostic test studies, which have been highlighted. These need to be addressed by future investigators, if informed judgements on the validity and reproducibility of the findings of their reviews are to be made.

## Additional material

**Additional file 1: Appendix 1 - Search algorithms.**

**Additional file 2: Appendix 2 - Set of included studies.**

**Additional file 3: Appendix 3 - PRISMA results for individual reviews.**

**Additional file 4: Appendix 4 - Evidence Tables.**

## Acknowledgements

During the period of this research BHW was in receipt of a fellowship from the Medical Research Council. We would like to thank all reviewers who have commented on this manuscript.

## Author details

<sup>1</sup>Biostatistics, University of Manchester, Jean McFarlane Building, Oxford Road, Manchester, M13 9PL, United Kingdom. <sup>2</sup>Centre for Social Ethics and Policy, University of Manchester, Williamson Building, Oxford Road, Manchester, M13 9PL, United Kingdom.

## Authors' contributions

Both BHW and MQ selected, extracted and appraised the data. BHW wrote the first draft. MQ commented on, and edited all drafts. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 13 December 2010 Accepted: 9 December 2011

Published: 9 December 2011

## References

1. Buntinx F, Aertgeerts B, Macaskill P: **Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic tests.** In *The evidence base for clinical diagnosis: theory and methods of diagnostic research*. 2 edition. Edited by: Knottnerus JA, Buntinx F. London: BMJ books; 2009:180-212.
2. Loefflang M, Deeks JJ, Gatsonis C, Bossuyt PMM: **Systematic reviews of diagnostic test accuracy.** *Ann Intern Med* 2008, **149**:889-897.
3. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH: **Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.** *J Clin Epidemiol* 2005, **58**:982-90.
4. Rutter CM, Gatsonis CA: **A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations.** *Stat Med* 2001, **20**:2865-2884.
5. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ: **Publication and related biases.** *Health Technol Assess* 2000, **4**(10).
6. Egger M, Dickersin K, Smith GD: **Chapter 3, Problems and limitations in conducting systematic reviews.** In *Systematic reviews in healthcare, BMJ books*. 2 edition. Edited by: Egger M, Smith GD, Altman DG. London; 2001:43-68.
7. Higgins JPT, Thompson SG: **Quantifying heterogeneity in a meta-analysis.** *Stat Med* 2002, **21**(11):1539-1559.
8. Thompson SG: **Why and how sources of heterogeneity should be investigated.** In *Systematic reviews in healthcare, BMJ books* Edited by: Egger M, Smith GD, Altman DG, 2 2001, 157-175.

9. Deville W, Butinx F, Bouter LM, Montori VM, de Vet HCW, van der Windt DAWM, Bezemer PR: **Conducting systematic reviews of diagnostic studies: didactic guidelines.** *BMC Medical Research Methodology* 2002, **2**:9.
10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC: **Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative.** *BMJ* 2003, **326**(7379):41-4.
11. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ: **Chapter 9: Assessing methodological quality.** In *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration* Edited by: Deeks JJ, Bossuyt PM, Gatsonis C 2009, Available from: <http://srdta.cochrane.org/>.
12. Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PM, Kleijnen J: **Development and validation of methods for assessing the quality of diagnostic accuracy studies.** *Health Technol Assess* 2004, **8**(25).
13. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: **Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses.** *Lancet* 1999, **354**:1896-900.
14. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration.** *PLOS Medicine* 2009, **6**(7):e1000100.
15. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J: **Challenges in Systematic Reviews of Diagnostic Technologies.** *Ann Intern Med* 2005, **142**:1048-1055.
16. Willis and Quigley: **Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review.** *BMC Medical Research Methodology* 2011, **11**:27.
17. Golder S, McIntosh HM, Duffy S, Glanville J: **Centre for Reviews and Dissemination and UK Cochrane Centre Search Filters Design Group. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE.** *Health Information and Libraries Journal* 2006, **23**:3-12.
18. Whiting P, Westwood M, Burke M, Sterne J, Glanville J: **Systematic reviews of test accuracy should search a range of databases to identify primary studies** *Journal of Clinical Epidemiology.* 2008, **61**:357-364.
19. Morris JA, Gardner MJ: **Epidemiological Studies.** In *Statistics with confidence.* 2 edition. Edited by: Altman D, Machin D, Bryant TN, Gardner M. London: BMJ books; 2001:57-67.
20. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG: **The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed.** *BMJ* 2010, **340**:c723.
21. Armitage P, Berry G, Matthews JNS: **Statistical methods in medical research.** Oxford, UK: Blackwell; 4 2002, 134-137.
22. Cohen J: **A Coefficient of Agreement for Nominal Scales.** *Educ Psychol Meas* 1960, **20**:37-46.
23. Fleiss JL, Levin B, Paik MC: **Statistical methods for rates and proportions.** New Jersey: John Wiley & Sons; 3 2003, 598-622.
24. Landis JR, Koch GG: **The Measurement of Observer Agreement for Categorical Data.** *Biometrics* 1977, **33**(1):159-174.
25. Stevens A, Milne R, Burls A: **Health technology assessment: history and demand.** *J Pub Hlth Med* 2003, **25**:98-101.
26. Schmid C, Chung M, Chew P, Lau J: **[O13] Survey of diagnostic test meta-analyses.** *12th Cochrane Colloquium Ottawa*; 2004.
27. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schunemann HJ, Garg AX, Busse JW, Heels-Ansdell D, Ghali WA, Manns BJ, Guyatt GH: **An observational study found that authors of randomised controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods.** *J Clin Epidemiol* 2004, **57**:1232-6.
28. Daniels S, Kumar A, Scott C, Swann S, Djulbegovic B: **Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group.** *BMJ* 2004, **328**:22-4.
29. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG: **Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols.** *BMJ* 2008, **337**:a2299.
30. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC: **Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study.** *BMJ* 2005, **330**:1049.
31. Moses LE, Shapiro D, Littenberg B: **Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.** *Stat Med* 1993, **12**(14):1293-316.
32. Littenberg B, Moses LE: **Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method.** *Med decis Making* 1993, **13**:313-321.
33. Van Houwelingen HC, Zwinderman KH, Stijnen T: **A bivariate approach to meta-analysis.** *Statist Med* 1993, **12**:2273-2284.
34. Chalmers I, Altman D: **Foreword.** In *Systematic Reviews.* Edited by: Chalmers I, Altman D. London: BMJ Publishing Group; 1995.
35. Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD: **Chapter 1: Introduction.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008).* *The Cochrane Collaboration* Edited by: Higgins JPT, Green S 2008, Available from <http://www.cochrane-handbook.org>.

#### Pre-publication history

The pre-publication history for this paper can be accessed here:  
<http://www.biomedcentral.com/1471-2288/11/163/prepub>

doi:10.1186/1471-2288-11-163

**Cite this article as:** Willis and Quigley: The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Medical Research Methodology* 2011 **11**:163.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

