

Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy

Fujian Song,^{a,b} Khalid S Khan,^{b,c} Jacqueline Dinnes^{b,d} and Alex J Sutton^e

Background	Despite the great possibility of publication bias in studies of diagnostic test research, empirical studies about publication bias have mainly focused on studies of treatment effect.
Methods	A sample of 28 meta-analyses of diagnostic accuracy was selected from the Database of Abstracts of Reviews of Effectiveness (DARE). Methods used to deal with publication and related biases in these meta-analyses were examined. Asymmetry of funnel plot of estimated test accuracy against corresponding precision for each meta-analysis was assessed by three statistical methods: rank correlation method, regression analysis, and Trim and Fill method.
Results	In reviews of diagnostic accuracy, there was a general lack of consideration of appropriate literature searching to minimize publication bias, and the impact of possible publication bias has not been systematically assessed. The results of the three different statistical methods consistently showed that in a large proportion of the 28 meta-analyses evaluated, the smaller studies were associated with a greater diagnostic accuracy. Exploratory analyses found that the fewer the literature databases searched, the greater the funnel plot asymmetry in meta-analyses. Funnel plot asymmetry tended to be greater in meta-analyses that included smaller number of primary studies. Our data revealed no consistent relationship between funnel plot asymmetry and language restriction in reviews.
Conclusions	Further research is required to explain why smaller studies tended to report greater test accuracy in a large proportion of meta-analyses of diagnostic tests. In systematic reviews of diagnostic studies, literature search should be sufficiently comprehensive and possible impact of publication bias should be assessed.
Keywords	Meta-analysis, diagnostic accuracy, publication bias, funnel plot
Accepted	1 August 2001

Clinical and public health decisions should be made based on findings from scientific research. The evidence that is available for decision-making is generally that which has been published. However, publication bias has been recognized as a serious problem in medical and health-related research.^{1,2} If those studies that are published are a biased sample of all studies that have been conducted, then the validity of any inferences that can be drawn is threatened.

The accessibility of research findings is dependent not only on whether a study is published but also on when, where and in what format this occurs. Dissemination bias is a term broader than publication bias, in order to include publication bias as well as other related biases due to the time, type and language of publication, multiple publication, selective citation of references, database index bias, and biased media attention.² However, publication bias poses the most serious difficulty in locating relevant studies.

The concern about dissemination bias (or publication and related biases) has led to efforts to deal with such bias. Suggested methods include prospective registration of research, comprehensive searches for relevant studies, and graphical and statistical methods for detecting publication bias. In meta-analysis, funnel plot and related statistical analyses are the most commonly used methods for assessing the possible existence of publication bias.^{2–4}

Theoretical and empirical evidence indicates that publication bias may be exacerbated where similar studies may produce

^a Department of Public Health and Epidemiology, University of Birmingham, UK.

^b NHS Centre for Reviews and Dissemination, University of York, UK.

^c Academic Department of Obstetrics & Gynaecology, University of Birmingham, UK.

^d Wessex Institute for Health Research and Development, University of Southampton, UK.

^e Department of Epidemiology and Public Health, University of Leicester, UK.

Correspondence: Dr Fujian Song, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. E-mail: f.song@bham.ac.uk

very different results and such studies can be easily conducted and abandoned. Thus, it has been argued that publication bias poses a particular problem for studies of test accuracy.^{5,6} This is because data regarding test accuracy is often collected as part of routine clinical care, and in the absence of any registration process for attempted evaluations. Hence, one might expect reviews of such studies to exaggerate the summary estimate of test accuracy if publication is related to the positivity of results. Despite the great potential for publication bias in diagnostic test research, empirical studies about publication bias have mainly focused on studies of treatment effect.^{7,8}

In this paper, we examined methods used to deal with dissemination bias in a sample of meta-analyses of studies of diagnostic tests. We also used funnel plot and related statistical methods to empirically assess the risk of publication bias in this literature.

Methods

Identification of meta-analyses of test accuracy

Systematic reviews included in the Database of Abstracts of Reviews of Effectiveness (DARE) up to December 1999 were examined to identify systematic reviews of diagnostic test. The DARE database, produced by the National Health Service (NHS) Centre for Reviews and Dissemination in York, includes systematic reviews which are identified by regular and extensive literature searches, including: electronic searching of a large number of bibliographic databases, hand searching of key major medical journals, and by scanning grey literature since 1994. To be included in the DARE database, a review has to meet at least four of the following six criteria: well-defined review question, appropriate literature search, explicit criteria for including primary studies, validity assessment, presentation of study details, and appropriate synthesis of study results.⁹

Two independent reviewers assessed the relevance of identified systematic reviews of diagnostic studies. Systematic reviews of diagnostic test were deemed relevant if they contained meta-analyses with five or more primary studies of test accuracy and provided sufficient data for 2×2 tables of a binary test result versus a reference standard. Reviews addressing test development and diagnostic effectiveness or cost effectiveness were excluded. Disagreements about selection were resolved by consensus or using arbitration by a third reviewer.

Data extraction

From each included systematic review, one author extracted details of the literature search used, language restrictions in study selection, and assessment of risk of publication bias. Data were also extracted to generate 2×2 tables for estimating accuracy of diagnostic tests. Since a systematic review may assess more than one diagnostic test versus the same reference test, this generated more than one meta-analysis data set in several reviews. All extracted data were double-checked by a second author.

Statistical analyses

A commonly used method for detecting publication bias employs a graphical plot of estimates of effect versus some measure of their precision for each of the primary studies in a meta-analysis. The plot is known as the 'funnel' plot because studies

of smaller size will have a wider distribution of results than studies of larger size, due to a higher degree of random variation.³ If the possibility of publication is greater for larger and more positive studies, the smaller negative studies may not appear in the literature. This may lead to asymmetry in the funnel. Thus an asymmetric funnel plot in meta-analysis suggests the existence of publication bias, although other alternative explanations should also be explored.

There is no universally accepted measure of test accuracy in meta-analyses of screening and diagnostic data.¹⁰ Studies of test accuracy often summarize their findings as sensitivity, specificity and predictive values. Sensitivity and specificity are often inversely related depending on varying test thresholds.¹¹ Therefore, it may not be appropriate to combine estimates of sensitivity or specificity separately in meta-analysis.^{5,11} To estimate test accuracy and its variance we used a measure that combines the number of true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*) in a 2×2 table into a single statistic, *d*, as shown below:¹¹

$$d = \sqrt{3} (\log TP + \log TN - \log FP - \log FN) / \pi$$

$$\text{Variance } (d) = 3(1/TP + 1/FP + 1/FN + 1/TN) / \pi^2$$

The interpretation of *d* values is similar to that of the log odds ratio (or diagnostic odds ratio), which measures the discriminative ability of a diagnostic test by considering both sensitivity and specificity simultaneously. Diagnostic *d* values that are significantly greater than zero indicate discriminative ability. The larger the *d* values the more accurate the test.

In this paper, we assessed the possibility of publication bias in meta-analyses of test accuracy by examining asymmetry of funnel plot of estimates of diagnostic *d* against corresponding precision. The informal examination for asymmetry of funnel plot is subjective so that the same graph may be interpreted differently by different observers.¹² Therefore, we formally evaluated asymmetry of funnel plots by using three statistical approaches described below.

Begg's method

The adjusted rank correlation method was used to assess the correlation between test accuracy estimates and their variances.¹³ The deviation of Spearman's *rho* values from zero provides an estimate of funnel plot asymmetry. Positive values indicate a trend towards higher levels of test accuracy in studies with smaller sample sizes.

Egger's test

To determine whether there is an association between test accuracy estimates and their precision, regression analysis was conducted by using the following equation:⁷ $SND = a + b \times SE(d)^{-1}$. In this regression equation *SND* is the standard normal deviate defined as diagnostic *d* divided by its standard error *SE(d)*; *a* is the intercept and *b* the slope. The intercept value (*a*) provides an estimate of asymmetry of funnel plot, with positive values (*a* > 0) indicating a trend towards higher levels of test accuracy in studies with smaller sample sizes. Unweighted regression analyses was conducted in this paper.

Trim & Fill method

A rank-based data augmentation technique was used to estimate the number of missing studies and to produce an adjusted estimate of test accuracy by imputing suspected missing

studies.¹⁴ Both random and fixed effect models were used to assess the impact of model choice on publication bias. The adjusted estimates indicate if the estimates of test accuracy based on simple meta-analysis are biased due to funnel plot asymmetry. We computed the difference between unadjusted and adjusted accuracy estimates. Positive values indicate a trend towards overestimation of diagnostic accuracy due to missing studies.

Results

The DARE database contained 1110 abstracts of reviews at the time of this study. The initial scanning of DARE abstracts generated 95 systematic reviews that were potentially relevant. After independent examination of the full publication, 59 systematic reviews of test accuracy were identified. Of these, 20 systematic reviews of test accuracy met the criteria for inclusion.^{15–34} Among these reviews there were a total of 28 meta-analyses, which were assessed for asymmetry of funnel plot.

Literature search and consideration of publication bias

As shown in Table 1, the most commonly employed search strategy used the MEDLINE database (all 20 systematic reviews) supplemented with checking the reference lists of retrieved studies (all but three). Only six reviews included searched an additional electronic database (EMBASE, SCI, or Current Contents); four contacted experts or authors; and two conducted some form of hand searching to identify relevant studies. Only two reviews stated that no language restrictions were employed; 12 had language restriction (11 to English only) and six did not specify this issue. None of the reviews assessed publication bias as part of data synthesis, although four considered the possibility of its existence in the discussion section of the paper.

Asymmetry in funnel plots

The rank correlation analysis (Begg's method) of funnel plots indicated that smaller studies tend to report better test performance in general. Of the 28 meta-analyses, 23 showed a positive correlation between test accuracy and variance, though this relationship was statistically significant in only six meta-analyses (Figure 1). Similarly the linear regression analysis (Egger's method) also indicated a general trend towards asymmetry of funnel plots with smaller studies tending to have better test performance. Of the 28 meta-analyses, 25 showed positive intercept values with statistical significance in 12 analyses. The Trim & Fill analysis using random effects model suggested that 17 of the 28 meta-analyses had missing studies. In seven meta-analyses, the estimated number of missing studies was greater than three, which suggested that the number of missing trials was significant.¹⁴ The test accuracy estimates would have been more conservative had the missing studies been identified and included in the review. This general trend was not altered by use of a fixed effect model.

According to the results of the Trim & Fill analysis, the value of diagnostic d has been overestimated by 21–30% in two meta-analyses, by 11–20% in seven meta-analyses, and by 5–10% in six meta-analyses. Given a diagnostic d of 2.0, a specificity of 90.0% corresponds to a sensitivity of 80.7%. If this diagnostic d is overestimated by 25% (from 2.0 to 2.5), the corresponding

sensitivity will be increased from 80.7% to 91.2%. However, the adjusted results produced by the Trim & Fill method are not intended to give a 'better' estimate of test performance, but they can be used as a form of sensitivity analysis for estimating the likely impact of bias on the meta-analysis.^{8,14}

Funnel plot asymmetry and efforts to minimize publication bias

Table 2 shows the results of exploratory analyses of funnel plot asymmetry and efforts to minimize publication bias in reviews. These results should be interpreted with caution because they were based on a small number of meta-analyses.

There was a general trend that the fewer the literature databases searched, the greater the funnel plot asymmetry in meta-analyses, according to Begg's method and Egger's method. For example, the average Spearman's rho was 0.37, 0.26, and 0.09, respectively, in meta-analyses that searched literature from one or two databases, three databases, and four or more databases. However, this trend did not exist according to the results by the Trim & Fill method.

Our data revealed no consistent association between funnel plot asymmetry and language restriction in reviews (Table 2). On the other hand, results from all the three methods suggested that funnel plot asymmetry tended to be greater in meta-analysis that included smaller number of primary studies (Table 2).

Discussion

In systematic reviews, a thorough literature search is crucial to identify all relevant studies, published or not. The search of electronic databases alone is seldom sufficient and should be supplemented by checking references of relevant studies and contacting experts or organizations.³⁵ Moreover, systematic reviews should not restrict inclusion of studies according to language of publication due to possible language bias.³⁶ However, in the sample of reviews assessed, there is a general lack of consideration of appropriate literature searching to minimize possible publication and related biases. Tentative results in our study (Table 2) indicate that comprehensive literature search may reduce publication and related biases.

Because even a thorough literature search may not completely eliminate the risk of publication bias, formal assessment of such a risk should be incorporated into reviews' analyses, conclusions and recommendations.³⁷ However, a vast majority of the reviews in our sample did not assess the possibility of publication bias. The impact of the possible publication bias has not been considered even in meta-analyses that showed asymmetric funnel plots (statistically significant in 6 of 28 according to Begg's methods and in 12 of 28 according to Egger's method).

The results of the three different statistical methods we used in the paper consistently showed that in a large proportion of the 28 meta-analyses evaluated, the smaller studies were associated with greater test accuracy. However, the number of meta-analyses in our sample is small ($n = 28$), and our findings need to be confirmed by further research in a large number of meta-analyses. Moreover, this paper has other methodological limitations that are discussed below.

First, the included meta-analyses of test accuracy were obtained from one database (DARE). Although the DARE database is

Table 1 Characteristics of the included meta-analyses of test accuracy

Systematic review	Diagnostic problem	Meta-analysis No.	Description of test versus reference standard	No. of studies	No. of patients	Literature search	Language restriction	Consideration of publication bias
Badgett <i>et al.</i> , 1997 ¹⁵	Left-sided heart failure	1	Clinical examination versus increased filling pressure	10	2560	MEDLINE	English	No
Bax <i>et al.</i> , 1997 ¹⁶	Chronic stable left ventricular dysfunction	2	F18-FDG PET versus echocardiography, radionuclide ventriculography or contrast ventriculography	12	775	MEDLINE Reference lists Hand-searching	English	No
		3	TL201 Rest-Redistribution imaging versus echocardiography, radionuclide ventriculography	8	758			
		4	Low dose dobutamine echocardiography versus echocardiography, radionuclide ventriculography	16	2420			
Bonis <i>et al.</i> , 1997 ¹⁷	Hepatitis C	5	Alanine aminotransferase (ALT) versus liver histology (strict definition)	12	446	MEDLINE Reference lists	No language restriction	Narrative discussion
DaSilva <i>et al.</i> , 1995 ¹⁸	Neonatal sepsis	6	C-reactive protein (quantitative) versus positive blood culture; postmortem confirmation of diagnosis	9	1234	MEDLINE EMBASE Reference lists	No language restriction	No
de Vries <i>et al.</i> , 1996 ¹⁹	Peripheral arterial disease	7	Duplex ultrasonography versus angiography	14	4906	MEDLINE Reference lists	English	Narrative discussion
Fahey <i>et al.</i> , 1995 ²⁰	Cervical disease	8	Pap smear versus histology	59	17 421	MEDLINE Reference lists	English	No
Faron <i>et al.</i> , 1998 ²¹	Pregnancy	9	Fetal fibronectin (high risk; single test) versus pre-term delivery	12	1601	MEDLINE Current contents correspondence Author Conference proceedings Contact experts	Not specified	Narrative discussion
Hallan & Asberg, 1997 ²²	Acute appendicitis	10	C-reactive protein versus histology	23	3551	MEDLINE SCI Reference lists	Not specified	No
Huicho <i>et al.</i> , 1996 ²³	Infectious diarrhoea	11	Fecal leukocytes versus stool culture	19	7132	MEDLINE Reference lists Contact experts	Not specified	No
Kearon <i>et al.</i> , 1998 ²⁴	Deep venous thrombosis	12	Venous ultrasonography (symptomatic patients) versus venography	16	2399	MEDLINE Reference lists Personal files	Not specified	No
Liedberg <i>et al.</i> , 1995 ²⁵	Temporomandibular disc imaging	13	Venous ultrasonography (asymptomatic patients) versus venography	14	1808			
		14	Sagittal arthrography versus surgery, autopsy, histology, imaging, macroscopy	8	366	MEDLINE Reference lists Dissertations	English	No

Table 1 Continued

Systematic review	Meta-analysis No.	Diagnostic problem	Description of test versus reference standard	No. of studies	No. of patients	Literature search	Language restriction	Consideration of publication bias
Mackenzie <i>et al.</i> , 1996 ²⁶	15	Knee injury	MRI of anterior cruciate ligament versus arthroscopy	15	2076	MEDLINE EMBASE RSNA index to imaging literature Reference lists	English	No
	16		MRI of medial meniscus versus arthroscopy	14	2104			
	17		MRI of lateral meniscus versus arthroscopy	13	2095			
Mitchell <i>et al.</i> , 1998 ²⁷	18	Squamous intraepithelial lesions	Colposcopic impression versus colposcopic biopsy	8	6131	MEDLINE	English	No
Mol <i>et al.</i> , 1997 ²⁸	19	Tubal pathology in subfertile patients	Serum chlamydial antibodies versus laparoscopy (occlusion, hydrosalpinx or adhesions)	13	1725	MEDLINE EMBASE Reference lists	Not specified	No
	20		Serum chlamydial antibodies versus laparoscopy (occlusion, hydrosalpinx or adhesions) and/or HSG	8	992			
Orr <i>et al.</i> , 1995 ²⁹	21	Appendicitis	Ultrasound versus appendicitis	17	3358	MEDLINE Reference lists Hand-searching	English	Narrative discussion
Scheidtler <i>et al.</i> , 1997 ³⁰	22	Lymph node metastases in cervical cancer	Lymphanography versus histology	17	1408	MEDLINE Reference lists Contact experts	English	No
	23		CT versus histology	19	1040			
	24		MRI versus histology	10	837			
Stegman-Igra <i>et al.</i> , 1997 ³¹	25	Vascular catheter-related bloodstream infection	Semi-quantitative catheter culture versus clinical definition	13	3280	MEDLINE Reference lists	English	No
Storgaard <i>et al.</i> , 1994 ³²	26	Alcoholism	Michigan Alcoholism Screening Test versus 'defined diagnostic criteria'	17	4521	MEDLINE Reference lists	Not specified	No
Swart <i>et al.</i> , 1995 ³³	27	Tubal pathology	Hysterosalpingography versus laparoscopy for tubal patency	12	2641	MEDLINE Reference lists	English, French, German, Dutch	No
Wells <i>et al.</i> , 1995 ³⁴	28	DVT—asymptomatic patients following surgery	Ultrasound versus contrast venography	11	1616	MEDLINE Current contents Reference lists Hand-searching	English	No

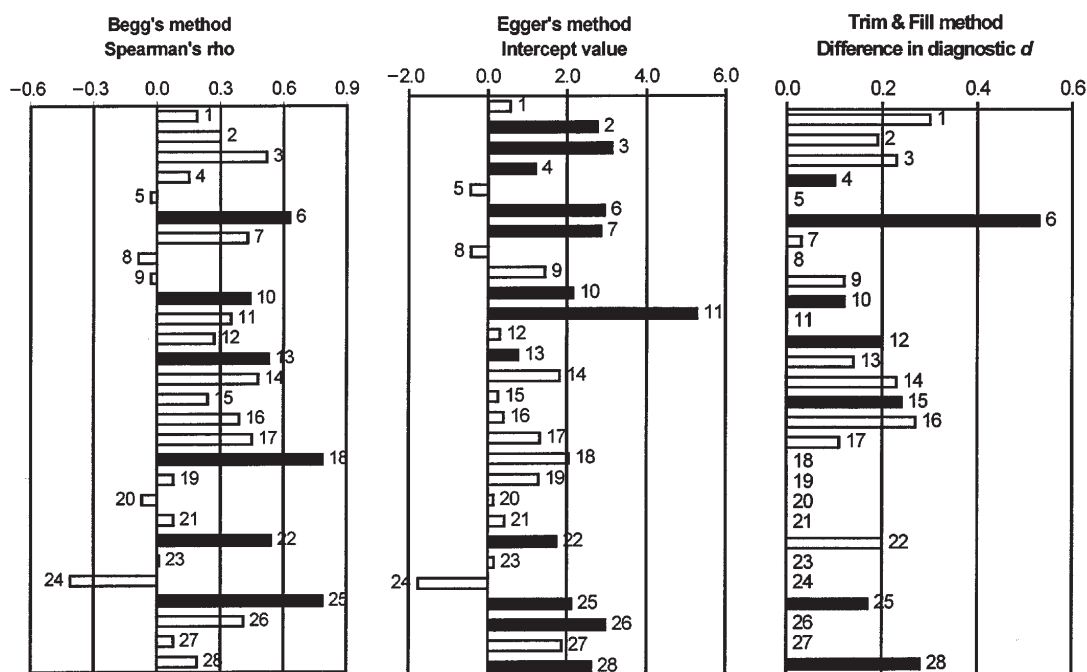


Figure 1 Results of statistical assessment of funnel plot asymmetry in 28 meta-analyses of test accuracy

Meta-analysis number corresponds to numbers shown in Table 1.

Black bars indicate statistical significance in Begg's method and Egger's method ($P < 0.1$); they indicate that the estimated number of missing studies is >3 in the Trim & Fill method.

Table 2 Efforts to minimize publication bias and the weighted average of measurement of funnel plot asymmetry (weighted by the number of studies included in each meta-analysis)

	No. of meta-analyses	Begg's method Spearman's rho	Egger's method Intercept a	Trim & Fill method Difference in d
No. of databases searched				
1–2 databases	7	0.372	1.825	0.066
3 databases	15	0.264	1.744	0.113
≥ 4 databases	6	0.093	0.381	0.108
Language restrictions				
English only	17	0.215	0.884	0.108
Unspecified	8	0.292	2.393	0.077
Less or no restriction	3	0.190	1.318	0.145
No. of studies included in meta-analysis				
< 10 studies	5	0.473	2.037	0.226
10–14 studies	13	0.299	1.821	0.119
≥ 15 studies	10	0.182	1.111	0.067

based on the results of extensive literature searches conducted since 1994, it is not exhaustive and some eligible meta-analyses may have been missed. Since systematic reviews included in the DARE database have to pass certain methodological criteria, the problems identified based on our sample might become worse if all relevant meta-analyses published in the whole literature had been considered.

Second, an asymmetric funnel plot can only reveal the existence of 'small-study effects'³⁸ that may be due to publication bias or other factors such as variations in test procedures, patients,

reference standards, study design quality, and chance.³⁹ For example, diagnostic studies with methodological shortcomings tend to overestimate the accuracy of a diagnostic test.⁴⁰ However, in this paper we are not able to separate the impact of publication bias from other factors on the observed asymmetry in funnel plots, because of lack of data and other resources. Future research is required to explain why smaller studies tended to report better test accuracy.

Finally, different methods may provide different statistical results for funnel plot asymmetry in a meta-analysis, although

Table 3 Funnel plot asymmetry in meta-analyses of diagnostic accuracy and treatment effectiveness

Study	Type of meta-analyses	Source of meta-analyses	Method	Proportion of asymmetric funnel plots
Present study	Diagnostic accuracy	DARE	Egger's method	43% (12/28) ^a
			Trim & Fill method	61% (17/28) ^b
Egger <i>et al.</i> 1997 ⁷	Effectiveness	Cochrane reviews	Egger's method	13% (5/38) ^a
		General medical literature	Egger's method	38% (14/37) ^a
Sutton <i>et al.</i> 2000 ⁸	Effectiveness	Cochrane reviews	Trim & Fill method	54% (26/48) ^b

^a Proportion of meta-analyses with significant intercept values.

^b Proportion of meta-analyses with missing studies.

they showed a similar trend (Figure 1). All three statistical methods used have limitations. For example, Begg's rank correlation method has low statistical power in detecting existing associations between effect estimates and their precision.^{7,13} Both Egger's regression method and the Trim & Fill method may be related to a great false-positive rate in detecting significant asymmetry of funnel plots under certain circumstances.^{38,41,42} Thus there may be no statistical method that is a superior remedy for assessing publication bias than another, and any method used in meta-analysis for detecting publication bias is by nature indirect and exploratory.²

The systematic reviews analysed in this paper generally contained meta-analyses with a small number of primary studies. The median number of studies per meta-analysis was 13 (range 9–59), and in our analysis the power of the statistical methods, in particular the rank correlation test, is perhaps not sufficient despite the use of a more liberal threshold for statistical significance ($P < 0.10$). It seems more appropriate to emphasize the direction and the size of estimates of funnel plot asymmetry, rather than the statistical significance. Although statistical significance was only observed in a small number, a negative relationship between test accuracy and its precision was observed in an overwhelming majority of the included meta-analyses (Figure 1). Hence, we are able to identify a general trend towards higher estimates of test accuracy among smaller diagnostic studies.

The appearance of the funnel plot may depend on the measures used to summarize the results of primary studies.⁴³ However, it is unlikely that the overall picture observed in our study will change if the test accuracy is measured by using different methods. Diagnostic d and odds ratio are the two methods available to obtain one single estimate of diagnostic

accuracy for individual studies. In fact, use of log odds ratio instead of diagnostic d will produce the identical results of statistical assessment of funnel plot asymmetry. In addition, variance and the inverse of standard errors have been used to measure the precision by different methods in this study, and the overall conclusion is consistent according to the results of the different methods.

Our finding of funnel plot asymmetry in meta-analyses of test accuracy is generally consistent with other empirical studies.^{7,8} Interestingly, comparison of our results with those of other empirical studies shows higher funnel plot asymmetry in meta-analyses of test accuracy than effectiveness literature (Table 3). While this comparison is not statistically and methodologically robust, it does provide credence to the hypothesis that publication bias may be more prevalent among studies of test accuracy.

In summary, future research is required to explain why smaller studies tend to report better test accuracy in so many meta-analyses. Although publication and related biases may not be the sole answer, the existence of such bias should be seriously considered. Ideally all primary studies of diagnostic tests should be prospectively registered in publicly accessible databases. All relevant sources of literature should be systematically searched, and the impact of possible publication bias should be assessed in systematic reviews of diagnostic studies. If the impact of publication bias cannot be ruled out, findings of literature reviews of diagnostic test accuracy should be interpreted with caution.

Acknowledgement

KSK and AS were funded by HEFC; JD and FS were funded by the Department of Health, England, UK.

KEY MESSAGES

- Empirical evidence from a sample of 28 meta-analyses of diagnostic accuracy suggested that smaller studies tended to report greater test accuracy.
- The observed 'small-study effects' may be at least partially due to publication and related biases.
- Exploratory analyses indicated that comprehensive literature search may be associated with less funnel plot asymmetry.

References

- ¹ Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Statist Soc A* 1988;**151**:419–63.
- ² Song F, Eastwood A, Gilbody S, Duley L, Sutton A. Publication and related biases. *Health Technol Assess* 2000;**4**:10.
- ³ Light R, Pillemer D. *Summing Up: The Science of Reviewing Research*. Cambridge, MA, and London: Harvard University Press, 1984.
- ⁴ Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases. In: Egger M, Davey Smith G, Altman DG

- (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Publishing Group, 2001, pp.189–208.
- 5 Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.
 - 6 Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests. A review of the concepts and methods. *Arch Pathol Lab Med* 1998;**122**:675–86.
 - 7 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315**:629–34.
 - 8 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. An empirical assessment of the impact of publication bias on meta-analyses. *Br Med J* 2000;**320**:1574–77.
 - 9 Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). *Inter J Technol Assess Health Care* 1999;**15**:671–78.
 - 10 Walter SD, Jadad AR. Meta-analysis of screening data: a survey of the literature. *Stat Med* 1999;**18**:3409–24.
 - 11 Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;**117**:167–78.
 - 12 Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140**:290–96.
 - 13 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;**50**:1088–101.
 - 14 Duval S, Tweedie R. Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;**50**:455–63.
 - 15 Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? *JAMA* 1997;**277**:1712–19.
 - 16 Bax JJ, Wijns W, Cornel JH, Visser FC, Boersma E, Fioretti PM. Accuracy of currently available techniques for prediction of functional recovery after revascularization in patients with left ventricular dysfunction due to chronic coronary artery disease: comparison of pooled data. *J Am Coll Cardiol* 1997;**30**:1451–60.
 - 17 Bonis PA, Ioannidis JP, Cappelleri JC, Kaplan MM, Lau J. Correlation of biochemical response to interferon alfa with histological improvement in hepatitis C: a meta-analysis of diagnostic test characteristics. *Hepatology* 1997;**26**:1035–44.
 - 18 Da Silva O, Ohlsson A, Kenyon C. Accuracy of leukocyte indices and c-reactive protein for diagnosis of neonatal sepsis: a critical review. *Pediatr Infect Dis J* 1995;**14**:362–66.
 - 19 de Vries SO, Hunink MG, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. *Acad Radiol* 1996;**3**:361–69.
 - 20 Fahey MT, Irwig L, Macaskill P. Meta-analysis of pap test accuracy. *Am J Epidemiol* 1995;**141**:680–89.
 - 21 Faron G, Boulvain M, Irion O, Barnard PM, Fraser WD. Prediction of preterm delivery by fetal fibronectin: a meta-analysis. *Obstet Gynecol* 1998;**92**:153–58.
 - 22 Hallan S, Asberg A. The accuracy of C-reactive protein in diagnosing acute appendicitis. *Scand J Clin Lab Invest* 1997;**57**:373–80.
 - 23 Huicho L, Campos M, Rivera J, Guerrant RL. Fecal screening tests in the approach to acute infectious diarrhea: a scientific overview. *Pediatr Infect Dis J* 1996;**15**:486–94.
 - 24 Kearon C, Julian JA, Newman TE, Ginsberg JS. Noninvasive diagnosis of deep venous thrombosis. *Ann Intern Med* 1998;**128**:663–77.
 - 25 Liedberg J, Panmekiate S, Petersson A, Rohlin M. Evidence-based evaluation of three imaging methods for the temporomandibular disc. *Dentomaxillofac Radiol* 1996;**25**:234–41.
 - 26 Mackenzie R, Palmer CR, Lomas DJ, Dixon AK. Magnetic resonance imaging of the knee: diagnostic performance statistics. *Clin Radiol* 1996;**51**:251–57.
 - 27 Mitchell MF, Schottenfeld D, Tortolero-Luna G, Cantor SB, Richards-Kortum R. Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol* 1998;**91**:626–31.
 - 28 Mol BWJ, Dijkman B, Wertheim P, Lijmer J, Vanderveen F, Bossuyt PMM. The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. *Fert Steril* 1997;**67**:1031–37.
 - 29 Orr RK, Porter D, Hartman D. Ultrasonography to evaluate adults for appendicitis: decision making based on meta-analysis and probabilistic reasoning. *Acad Emerg Med* 1995;**2**:644–50.
 - 30 Scheidler J, Hricak H, Yu K, Subak L, Segal M. Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *JAMA* 1997;**278**:1096–101.
 - 31 Siegman-Igra Y, Anglim AM, Shapiro DE, Adal KA, Strain BA, Farr BM. Diagnosis of vascular catheter-related bloodstream infection: a meta-analysis. *J Clin Microbiol* 1997;**35**:928–36.
 - 32 Storgaard H, Nielsen SD, Gluud C. The validity of the Michigan Alcoholism Screening Test (MAST). *Alcohol Alcohol* 1994;**29**:493–502.
 - 33 Swart P, Mol BWJ, Vanderveen F, Vanburden M, Redekop WK, Bossuyt PMM. The accuracy of hysterosalpingography in the diagnosis of tubal pathology: a meta-analysis. *Fertil Steril* 1995;**64**:486–91.
 - 34 Wells PS, Lensing AWA, Davidson BL, Prins MH, Hirsh J. Accuracy of ultrasound for the diagnosis of deep venous thrombosis in asymptomatic patients after orthopedic surgery: a meta-analysis. *Ann Intern Med* 1995;**122**:47–53.
 - 35 McManus R, Wilson S, Delaney B *et al*. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *Br Med J* 1998;**317**:1562–63.
 - 36 Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–29.
 - 37 Stroup DF, Berlin JA, Morton SC *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;**283**:2008–12.
 - 38 Sterne JAC, Gavaghan D, Egger M. Publication and related biases in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53**:1119–29.
 - 39 Egger M, Smith GD. Bias in location and selection of studies. *Br Med J* 1998;**316**:61–66.
 - 40 Lijmer J, Mol B, Heisterkamp S *et al*. Empirical evidence of design related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–66.
 - 41 Irwig L, Berry G. Graphical test is itself biased. *Br Med J* 1998;**316**:469–69.
 - 42 Sterne JAC. High false positive rate for trim and fill method. Electronic response to Sutton *et al*'s paper (*Br Med J* 2000;**320**:1574–77):<http://www.bmj.com/cgi/eletters/320/7249/1574#E1>
 - 43 Tang JL, Liu JL. Misleading funnel plot for detecting of bias in meta-analysis. *J Clin Epidemiol* 2000;**53**:477–84.