Review

Assessment of publication bias for the surgeon scientist

S. S. Mahid[1], M. Qadan[1], C. A. Hornung[2] and S. Galandiuk[1]

[1]Price Institute of Surgical Research and Section of Colorectal Surgery, Department of Surgery, and [2]Department of Epidemiology and Population Health, University of Louisville School of Medicine, Louisville, Kentucky, USA

Correspondence to: Dr S. Galandiuk, Department of Surgery, University of Louisville, Louisville, Kentucky 40202, USA (e-mail: s0gala01@louisville.edu)

Background: **Publication bias occurs when statistically non-significant (negative) findings are not published. It can profoundly affect the results of systematic reviews and meta-analyses.**

Methods: **Qualitative and quantitative methods of detecting publication bias are described, including their advantages and disadvantages.**

Results and conclusion: **Accepted quality standards for the reporting of meta-analyses recommend assessment of publication bias, but currently there is no uniform standard for reporting. Quantitative methods are being used with increasing frequency. Authors should take steps to minimize publication bias, and use both qualitative and quantitative assessment methods to determine whether it is present.**

Paper accepted 18 April 2008

Published online in Wiley InterScience (www.bjs.co.uk). **DOI:** 10.1002/bjs.6302

Introduction

It is widely accepted that studies with a positive result are more likely to be submitted for publication and subsequently published than inconclusive or negative studies. A rare exception occurs when a very large negative study that fails to reject the null hypothesis is regarded as conclusive and approved for publication by both reviewers and journal editors. Nonetheless, the tendency for studies showing $P < 0{\cdot}050$ to be more likely to be published than those with $P > 0{\cdot}050$ is the essence of what is known as publication bias[1]. Such bias can affect a systematic review, with or without a meta-analysis, by showing a clinical intervention to be effective (statistically significantly related to a positive outcome) when the intervention in reality has little to no real impact on the outcome of interest.

Publication bias is not a new concept; it has been around for nearly 50 years. It was initially noted when Sterling[2] reviewed four journals and found that 97 per cent of the reports published in them reported statistically significant findings. This suggested the possibility that studies with statistically non-significant findings were not being published. Scherer and colleagues[3] have claimed that just over half of all conference abstracts are published in full, and that abstracts with significant results are more likely to be so treated than those with non-significant results.

Since the publication of 'Systematic reviews and meta-analyses for the surgeon scientist' in this journal[4], the authors have witnessed a tenfold increase in the assessment of publication bias in meta-analysis in peer-reviewed journals between 1998 and 2007. Such bias is not, however, assessed routinely in the surgical literature despite accepted reporting standards for systematic reviews with meta-analysis[5–7]. Unfamiliarity and difficulty in interpretation of the test findings may play a large part in this underutilization (*Table 1*)[8].

The Cochrane handbook does not encourage the use of quantitative methods to evaluate whether or not publication bias exists, because many of these tests are considered to have poor statistical power. However, the Cochrane group does acknowledge and accept the use of funnel plots to detect publication bias qualitatively. Nevertheless, quantitative methods are increasingly being employed by meta-analysts, partly because the widely accepted quality of reports of meta-analyses (QUOROM) statement requires the assessment of bias as part of its checklist[6]. QUOROM, like Cochrane, however, does not specify which quantitative techniques should be used. The Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines, which are also widely accepted, are more specific and stipulate that either the funnel plot or the fail safe method should be used to detect publication bias[7].

Table 1 Frequently used methods of assessing the presence of publication bias in meta-analysis

	Reporting method	Interpretation	Advantage	Disadvantage
Qualitative test				
Funnel plot	Visual interpretation	Asymmetry suggests the presence of publication bias	Easy to judge	Subjective interpretation Asymmetry does not always suggest publication bias. Need to explore systematically the reasons for asymmetry Inadequate number of samples means that there may not be enough large studies to form the apex of the predicted funnel Requires a relatively large number of component studies with varying sample size Data can be distributed in a funnel shape even when studies come from more than one population, if they have the same population mean but different variances
Quantitative tests				
Egger	Regression analysis. Graphical output	$P < 0{\cdot}050$, if there is evidence of publication bias. This is also confirmed by a confidence interval that does not include zero	Data sets that report dichotomous outcomes Measures of associations or risk can be analysed More sensitive than Begg's test	Sensitivity is generally low in meta-analyses based on fewer than 20 studies Overestimates the extent of publication bias with studies of small sample size
Begg and Mazumdar	Rank correlation test. Graphical output	$P < 0{\cdot}050$, if there is evidence of publication bias. This is also confirmed by a confidence interval that does not include zero	Distribution-free method Involves no modelling assumptions	Sensitivity is generally low in meta-analyses based on fewer than 20 studies Non-significant test does not exclude the presence of publication bias
Trim and fill method	Rank-based technique	Aims to correct the asymmetry observed in the funnel plot in the setting of non-significant findings	Can be used to estimate the no. of missing studies Estimates the treatment effect by adjusting for potential publication bias	Low specificity – it might detect missing studies even in the absence of bias Danger of overcorrecting non-existent bias in response to funnel plot asymmetry arising from nothing more than random variation
Fail safe number	$5n + 10$ n represents the no. of studies in the meta-analysis	No. of unpublished studies with non-statistical findings required to change a significant result in a meta-analysis to non-significant	Fail safe n is easy to interpret Used only when publication bias has been detected by other tests	Overemphasizes importance of statistical significance Highly dependent on the mean effect size assumed for unpublished studies Does not account for studies that report effect in the opposite direction to that observed in the meta-analysis

One of the many advantages of conducting systematic reviews is that they can be used to generate or test hypotheses. However, they are prone to several other limitations, including inadequate or missing information, weakness of individual study designs, incomplete reporting of outcomes and heterogeneity of the intervention effects between studies. The effect of selecting a subset of studies based on outcome distorts the estimated effect of the intervention and possibly misleads clinicians who are trying to make evidence-based clinical decsions[9,10]. The problem of publication bias may be most likely, perhaps greatest, when there are financial incentives to publish positive results and suppress submission for publication of inconclusive studies[11].

Those who undertake meta-analyses and systematic reviews should account for publication bias in their methodology. One method of minimizing bias is to perform a comprehensive search for unpublished studies[12].

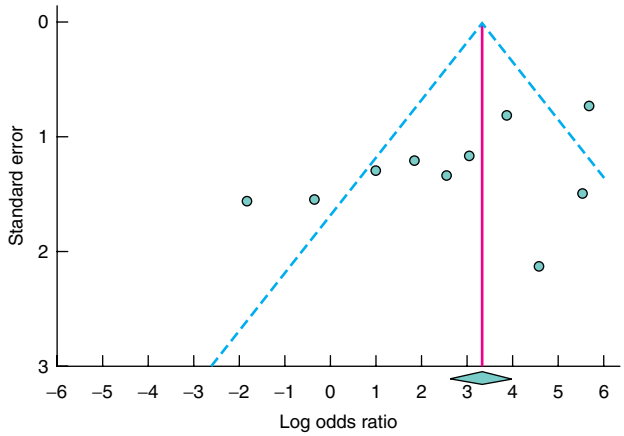
Analytical tools, such as a funnel plot, may also be used to assess the degree qualitatively and quantitatively. What follows is a description of the most commonly used techniques of evaluating publication bias. The advantages and common pitfalls of each technique are highlighted.

Qualitative assessment of publication bias

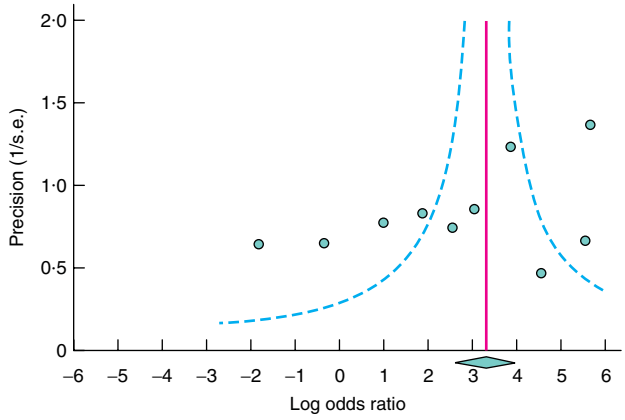
Funnel plot

The funnel plot is the most frequently used method of assessing the existence and potential impact on the estimated effect parameter[13,14]. The Central Limit Theorem and normal theory predict that a sample statistic (mean, log odds ratio and regression coefficient) will tend to be a normal distribution as the sample size increases. A plot of the published studies with their effect parameter, such as the log odds ratio on the *x*-axis and standard error on the *y*-axis, yields an inverted funnel[15]. In the absence of publication bias, the distribution of studies will be funnel shaped and symmetrical (or nearly so) around the true population parameter. In the presence of publication bias, there will be gaps or vacant regions within the funnel. Such gaps usually appear in the lower left or lower right corner of the funnel, suggesting that smaller or statistically non-significant studies have not been published[4,16].

Although the funnel plot is effective at demonstrating the presence of publication bias[16], plot asymmetry can result from other factors. These include differences in the methodological quality of the studies[17], differences in underlying risk in the various populations studied, choice of effect measure and/or measure of precision, and chance itself[18]. When smaller studies of poorer methodological quality are included, they may show a positive and larger effect than well powered and better designed studies[19]. An example of methodological bias that can produce funnel plot asymmetry was reported in a meta-analysis of observational studies of symptomatic hepatobiliary iminodiacetic acid (HIDA)-positive patients without evidence of gallstones[20]. The funnel plot using data from that analysis is asymmetrical about its base, suggesting the presence of publication bias with an absence of studies in the lower left quadrant of the plot (*Fig. 1a*). Symptomatic outcome following surgery is plotted against the standard error in *Fig. 1a* and against precision (*Fig. 1b*). Precision is denoted by the reciprocal of the standard error (1/s.e.), that is, as the standard error increases, precision decreases, and vice versa. This results in a curvilinear plot with an emphasis on larger studies, while smaller studies are concentrated at the bottom. Generally, the standard error is used on the vertical axis as this



a Standard error by log odds ratio



b Precision by log odds ratio

Fig. 1 **a** Asymmetrical funnel plot with standard error on the vertical axis. Dashed diagonal lines represent the 95 per cent confidence interval around the overall effect estimate, which is indicated by the vertical line. There were ten studies in this analysis, each shown by a circle. The absence of negative or null studies in the lower left corner indicates the presence of publication bias. The axis has been inverted to place the largest studies at the top of the graph. There is an emphasis on the smaller studies, where bias is more likely. **b** Asymmetrical curvilinear plot with precision (1/s.e.) on the vertical axis. The 95 per cent confidence interval lines are curved. There is an emphasis on larger studies, with smaller studies being concentrated at the bottom

represents the best choice in most cases. This results in a symmetrical funnel plot bounded by the 95 per cent confidence interval in the absence of publication bias[21]. Plotting precision on the vertical axis is usually employed in studies that compare trials with large variations in sample size[22].

Subjectivity in the visual interpretation of a funnel plot leaves it open to a wide range of potential errors[15,23].

Terrin and co-workers[23] reported the ability of medical researchers to identify publication bias based on visual inspection of 16 funnel plots each containing ten studies. This has been shown to be only slightly better than chance (53 per cent accuracy). Publication bias in this study was created by allowing studies of lower P value or larger sample size to have a greater probability of inclusion[23].

This highlights the importance of using additional methods that can reduce the subjective element of visual interpretation. In the following sections, tests used to evaluate visual asymmetry quantitatively are discussed. These tests theoretically require at least 30 studies to be included in the analysis in order to have sufficient statistical power (typically 0·80)[24]. Unfortunately, many meta-analyses involving randomized controlled trials include fewer than 30 studies, rendering these tests underpowered[19].

Quantitative assessment of publication bias

Egger's test

The Egger regression asymmetry test relies on regression[18], and its plot suggests publication bias more often than the Begg approach (see below)[25]. The Egger test detects funnel plot asymmetry by determining whether the intercept deviates significantly from zero in a regression of the standardized effect estimates against their precision. The test gives $P < 0{\cdot}050$ if there is evidence of publication bias. This is also confirmed by a confidence interval that does not include zero. Using the data set illustrated in *Fig. 1*, the Egger test gave $P = 0{\cdot}437$, with the confidence interval including zero, indicating the absence of publication bias.

Begg and Mazumdar adjusted rank correlation test

The Begg and Mazumdar adjusted rank correlation test is a direct statistical analogue of the visual funnel plot[25]. The presence of bias is determined by assessing whether or not there is a significant correlation between the effect estimates and their variances. Many commercially available meta-analysis computer programs include this test and provide graphical output in which the regression asymmetry graph plots the standardized effect estimates (odds ratio) against precision (1/s.e.), along with the variance weighted regression line as well as the confidence interval about the intercept. The Begg and Mazumdar test gives $P < 0{\cdot}050$ if there is evidence of publication bias. Failure of the confidence interval to include zero denotes asymmetry in the funnel plot, and this indicates the presence of possible publication bias in the included studies. Using the data set illustrated in *Fig. 1*, Begg's test did not indicate the presence of publication bias ($P = 0{\cdot}089$).

Trim and fill

The trim and fill technique was first described by Duval and Tweedie.[26] formally to evaluate bias in funnel plots in the setting of non-significant findings in a particular area of research. This technique attempts to correct the funnel plot asymmetry using the existing data. In this rank-based technique, the number of unmatched trials producing asymmetry on one side of the funnel is estimated. These trials are then trimmed from the funnel, leaving a symmetrical remainder from which the true centre of the funnel is estimated (filling) by standard meta-analytical procedures. The trimmed trials are then replaced and their missing counterparts imputed. This then allows an adjusted overall confidence interval to be calculated. If the results become non-significant as a result of the correction, this should not be accepted as the true value but rather as an indication that the results should be assessed more carefully in the likely presence of publication bias. Using this method, the presence of publication bias was detected in the data set used in *Fig. 1*.

Fail safe number

The fail safe number is the number of unpublished studies that would be needed to nullify a significant effect suggesting publication bias by raising the observed P value to $\geq 0{\cdot}050$[27]. If a large number of unpublished studies are required to render a statistically significant result non-significant, it is probable that they do not exist. The fail safe number is used only when publication bias has been detected by other tests. Limitations of this method include an overemphasis on statistical significance rather than on biological significance. The method also does not take into account the unlocated studies that report effects opposite to the results observed in the included studies, resulting in a smaller 'true' fail safe number, and implying that fewer studies would be required to overturn a significant result reported in a meta-analysis[28]. For the data in *Fig. 1*, the fail safe number is 138, representing the additional negative or inconclusive studies that would make the results non-significant. In other words, a further 138 studies would be needed to show that surgery for symptomatic HIDA scan-positive patients without gallstones is not beneficial compared with medical treatment.

Discussion

Publication bias occurs especially when research remains unpublished[29–31]. There are three main sources that contribute to this: researchers themselves, journal editors and the peer review process, and commercial funding sources. Researchers are often reluctant to submit a report when the results are not statistically significant[29]. They may erroneously believe that there will be no interest in inconclusive results or that the failure of their study to achieve statistically significant results reflects negatively on their research ability. Investigators may, however, be more likely to submit such work as an abstract or oral presentation at a meeting. Publication bias occurs when the results are not submitted to and published in a scientific journal, or when their appearance as an abstract in the proceedings of a meeting is not discovered and so they are not included in systematic reviews with meta-analysis.

Journal editors and the peer review process are also relevant in this context. Bias occurs when reviewers are disinclined to recommend acceptance of an article because of inconclusive results, or because of the possibility that methodological or design flaws exist that they cannot pinpoint and which may be responsible for the study not achieving statistical significance. They may fear that if they recommend publication and such methodological deficiencies are identified by another reviewer or reader, their prestige in the eyes of the editor and the scientific community will be diminished[32]. Editors, meanwhile, are probably not eager to publish inconclusive studies, even those that are methodologically sound, given the space limitations and competition between journals for significant research findings. Such significant findings are more likely to be published rapidly (pipeline bias)[31], to be published in journals with a higher impact factor[30], to be published in multiple forms (multiple publication bias)[33] and, finally, to be quoted more often than inconclusive findings[34].

Studies funded by commercial interests that fail to find statistically significant effects of the treatment or intervention are less likely to be published than similarly funded studies with positive results[1,29,30]. Commercial interests that are likely to thwart publication of inconclusive studies are probably also interested in thwarting presentation of inconclusive findings at professional scientific meetings. This means that such studies are not discoverable, even by thorough searching of the 'grey' literature (meeting proceedings, symposiums), making recognition of the resulting publication bias (nearly) impossible. This is an important reason for mandatory registration of all clinical trials.

Although the extent of publication bias is unknown[35], its presence in the era of evidence-based practice is important. Failure to identify bias can lead to an erroneous interpretation of systematic reviews and is a threat to the credibility of meta-analyses[30,36,37]. The consequences of bias are exemplified by the clinical trials of selective serotonin-reuptake inhibitors in children[38]. Nearly one-fifth of children aged less than 18 years experience at least one episode of major clinical depression. With such a large potential market, this attracted some large pharmaceutical companies and led to the development of paroxetine hydrochloride. Only when it was discovered that this antidepressant led to higher than expected rates of deliberate self-harm did the manufacturers release unpublished data indicating that the drug was ineffective at adequately treating childhood depression[39]. It later emerged that, of the 15 studies submitted to the Food and Drug Administration (FDA) in the USA, only three suggested the drug had any beneficial effect. This created a furore in the lay media, resulting in the development of a clinical registry for all clinical trials from their inception, as well as a number of other measures[40]. Despite the creation of this registry, there is evidence that publication bias remains a major issue in evidence-based medicine. In a recent report[41] of 74 FDA-registered studies involving 12 antidepressant drugs, nearly all trials (94 per cent) produced positive results, based on the published literature. Analysis of FDA data, however, showed that only 51 per cent (including unpublished studies) of all the trials conducted were in fact positive.

Several proposals have been put forward to document and record all unpublished studies[42]. It is unethical to enrol patients into a clinical study with the argument that their participation will advance science only to withhold the results from the scientific community[43]. Researchers have an ethical responsibility to their study participants, and moral responsibility to their colleagues, sponsors and the scientific community at large to ensure dissemination of study findings, irrespective of outcome[44].

The surgeon scientist should appreciate that an inconclusive study can be of great value, provided that it is scientifically sound. It can generate important hypotheses by stimulating thinking about the unexpectedly inconclusive results. It is also worth noting that there is no consensus about how publication bias is best detected and that a thorough assessment is not conducted routinely. Funnel plot analysis remains a minimum requirement. If systematic reviews with meta-analyses are to provide the highest level of evidence to aid clinicians in decision making, their authors must be confident about the results.

These should reflect a comprehensive search of all relevant healthcare databases, including the grey literature, to minimize publication bias. Ideally, qualitative and (when possible) quantitative tests for the presence of publication bias should be performed.

Acknowledgements

S.S.M. holds the John W. and Caroline Price Research Fellowship, and M.Q. holds the Joint Royal College of Surgeons of Edinburgh/James and Emmeline Ferguson Research Fellowship.

References

1 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**: 1385–1389.
2 Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significances. *Am Stat Assoc J* 1959; **54**: 30–34.
3 Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994; **272**: 158–162.
4 Mahid SS, Hornung CA, Minor KS, Turina M, Galandiuk S. Systematic reviews and meta-analysis for the surgeon scientist. *Br J Surg* 2006; **93**: 1315–1324.
5 Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; **274**: 1935–1938.
6 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**: 1896–1900.
7 Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**: 2008–2012.
8 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000; **320**: 1574–1577.
9 Craig JC, Irwig LM, Stockler MR. Evidence-based medicine: useful tools for decision making. *Med J Aust* 2001; **174**: 248–253.
10 Hagdrup N, Falshaw M, Gray RW, Carter Y. All members of primary care team are aware of importance of evidence based medicine. *BMJ* 1998; **317**: 282.
11 Friedberg M, Saffran B, Stinson TJ, Nelson W, Bennett CL. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA* 1999; **282**: 1453–1457.
12 Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; **309**: 1286–1291.
13 Beard SM, Holmes M, Price C, Majeed AW. Hepatic resection for colorectal liver metastases: a cost-effectiveness analysis. *Ann Surg* 2000; **232**: 763–776.
14 Gurusamy K, Junnarkar S, Farouk M, Davidson BR. Meta-analysis of randomized controlled trials on the safety and effectiveness of day-case laparoscopic cholecystectomy. *Br J Surg* 2008; **95**: 161–168.
15 Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000; **53**: 477–484.
16 Egger M, Smith GD. Misleading meta-analysis. *BMJ* 1995; **311**: 753–754.
17 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**: 408–412.
18 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–634.
19 Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000; **53**: 1119–1129.
20 Mahid SS, Jafri NS, Brangers CB, Minor KS, Hornung CA, Galandiuk S. A meta-analysis of the results of surgery in HIDA scan positive symptomatic patients without gallstones. *Arch Surg* 2008; (in press).
21 Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001; **54**: 1046–1055.
22 Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC *et al*. Large trials *vs* meta-analysis of smaller trials: how do their results compare? *JAMA* 1996; **276**: 1332–1338.
23 Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005; **58**: 894–901.
24 Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006; **333**: 597–600.
25 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**: 1088–1101.
26 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; **56**: 455–463.
27 Rosenthal R. Effect size estimation, significance testing, and the file-drawer problem. *J Parapsychol* 1992; **56**: 57–58.
28 Evans S. Statistician's comment. *BMJ* 1996; **312**: 125.
29 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992; **267**: 374–378.
30 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**: 867–872.

31 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; **315**: 640–645.
32 Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *Prog Clin Biol Res* 1983; **132E**: 57–68.
33 Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; **315**: 635–640.
34 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992; **305**: 15–19.
35 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000; **4**: 1–115.
36 Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci* 1993; **703**: 135–146.
37 Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; **279**: 281–286.
38 Couzin J. Psychopharmacology. Volatile chemistry: children and antidepressants. *Science* 2004; **305**: 468–470.
39 Couzin J. Drug research. Legislators propose a registry to track clinical trials from start to finish. *Science* 2004; **305**: 1695.
40 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986; **4**: 1529–1541.
41 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**: 252–260.
42 Horton R. Medical editors trial amnesty. *Lancet* 1997; **350**: 756.
43 Altman DG. The scandal of poor medical research. *BMJ* 1994; **308**: 283–284.
44 Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990; **263**: 1405–1408.