BMJ

RESEARCH

Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications

Santiago G Moreno, research student,[1] Alex J Sutton, professor of medical statistics,[1] Erick H Turner, assistant professor,[2] Keith R Abrams, professor of medical statistics,[1] Nicola J Cooper, senior research fellow,[1] Tom M Palmer, research associate,[3] A E Ades, professor of public health science[4]

[1]Department of Health Sciences, University of Leicester, Leicester LE1 7RH

[2]Department of Psychiatry, Oregon Health and Science University, Portland Veterans Affairs Medical Center, Portland, Oregon, USA

[3]MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol

[4]Department of Community Based Medicine, University of Bristol

Correspondence to: S G Moreno
sgm8@le.ac.uk

ABSTRACT

Objective To assess the performance of novel contour enhanced funnel plots and a regression based adjustment method to detect and adjust for publication biases.

Design Secondary analysis of a published systematic literature review.

Data sources Placebo controlled trials of antidepressants previously submitted to the US Food and Drug Administration (FDA) and matching journal publications.

Methods Publication biases were identified using novel contour enhanced funnel plots, a regression based adjustment method, Egger's test, and the trim and fill method. Results were compared with a meta-analysis of the gold standard data submitted to the FDA.

Results Severe asymmetry was observed in the contour enhanced funnel plot that appeared to be heavily influenced by the statistical significance of results, suggesting publication biases as the cause of the asymmetry. Applying the regression based adjustment method to the journal data produced a similar pooled effect to that observed by a meta-analysis of the FDA data. Contrasting journal and FDA results suggested that, in addition to other deviations from study protocol, switching from an intention to treat analysis to a per protocol one would contribute to the observed discrepancies between the journal and FDA results.

Conclusion Novel contour enhanced funnel plots and a regression based adjustment method worked convincingly and might have an important part to play in combating publication biases.

INTRODUCTION

In 2008 Turner et al published a study in the *New England Journal of Medicine* showing that the scientific journal literature on antidepressants was biased towards "favourable" results.[1] The authors compared the results in journal based reports of trials with data on the corresponding trials submitted to the US Food and Drug Administration (FDA) when applying for licensing. The discrepancies observed in the journal based reports were due to publication biases. Although the term publication bias has been used historically to refer to the suppression of whole studies based on (the lack of) statistical significance or "interest level," a range of mechanisms can distort the published literature. These include, in addition to the suppression of whole studies, selective reporting of outcomes or subgroups; data "massaging," such as the selective exclusion of patients from the analysis; and biases regarding timelines.[2] A good umbrella term for all these is dissemination biases[3 4]; in keeping with common usage we refer to them as publication biases. If such biases are present, any decision making based on the literature could be misleading,[5 6] not least through obtaining inflated clinical effects from meta-analysis.[7]

The FDA dataset is assumed to be an unbiased (but not the complete) body of evidence in the specialty of antidepressants and so is regarded a gold standard data source owing to the legal requirements of submitting evidence in its entirety to the FDA and its careful monitoring for deviations from protocol.[8-10] A gold standard dataset will not, however, be available in most contexts. In the absence of a gold standard, meta-analysts have had to rely on analytical methods to both detect and adjust for publication biases. This has been an active area of methodology development over the past decades, with much written on approaches to deal with publication biases in a meta-analysis context.[2] These include graphical diagnostic approaches and formal statistical tests to detect the presence of publication bias, and statistical approaches to modify effect sizes to adjust a meta-analysis estimate when the presence of publication bias is suspected.[2] While the performance of many of these methods has been evaluated using simulation studies, concerns remain as to whether the simulations reflect real life situations and therefore whether their perceived performance is representative of what would happen if they were used in practice. Understandably this has led to caution in the use of the methods, particularly for those that adjust effect sizes for publication biases[6]; but ultimately this is what is required for rational decision making if publication biases exist.

We consider what we believe are currently the best methods for identifying and adjusting for publication biases—both of which have been described only recently. Specifically, we consider a funnel plot (a scatter plot of effect size versus associated standard error) enhanced by contours separating areas of statistical significance from non-significance.[11] These contours help distinguish publication biases from other factors that lead to asymmetry in the funnel plot. The method used to adjust a meta-analysis for publication bias is based on a regression line fitted to the funnel plot.[12] The adjusted effect size is obtained by extrapolating the regression line to predict the effect size that would be seen in a hypothetical study of infinite size—that is, which has an effect size with zero associated standard error. For comparison and completeness we consider established methods to deal with publication bias. These are the regression based Egger's test for funnel asymmetry,[13] and the trim and fill method,[14] which adjusts a meta-analysis for publication bias by imputing studies to rectify any asymmetry in the funnel plot.

The dataset from Turner et al provides a unique opportunity to evaluate the performance of these analytical methods against a gold standard. We present the results of applying the diagnostic and adjustment methods to the journal published results and compare the findings with those obtained through (gold standard) analysis of the data submitted to the FDA.

METHODS
A full description of the dataset, how it was obtained, and the references to the trials associated with it have been published previously.[1] Briefly, Turner et al identified the cohort of all phase II and phase III short term double blind placebo controlled trials used for the licensing of antidepressant drugs between 1987 and 2004 by the FDA. Seventy four trials registered with the FDA and involving 12 drugs and 12 564 patients were identified. To compare drug efficacy reported by the published literature with that of the FDA gold standard, Turner et al collected data on the primary outcome from both sources. Once the primary outcome data were extracted from the FDA trial registry, they searched the published scientific literature for publications matching the same trials. When a match was identified, they extracted data on the article's apparent primary efficacy outcome. Because studies reported their outcomes on different scales, they expressed all effect sizes as standardised mean differences using Hedges' g scores (accompanied by corresponding variances).[15] Among the 74 studies registered with the FDA, 23 (31%), accounting for 3449 participants, were not published. Overall, larger effects were derived from the journal data than from the FDA data. Among the 38 studies with results viewed by the FDA as statistically significant, only one was unpublished. Conversely, inconclusive studies were, with three exceptions, either not published (22 studies) or published in conflict with the FDA findings (11 studies). Moreover, 94% of published studies reported a positive significant result for their primary outcome,

compared with 51% according to the FDA. Data for the analysis were extracted from the previous paper (table C in the appendix),[1] in which two studies were combined, making a total of 73 studies in our assessment.

Analysis
We applied two novel methods to the journal dataset: the contour enhanced funnel plot[11 16] to detect publication biases, and a regression based adjustment method[12] to adjust for them. For completeness and comparison we also applied to the dataset the most established and commonly used methods to deal with publication biases—namely, Egger's regression test[13] for detecting bias, and the trim and fill adjustment method (fixed effects linear estimator).[14 17-19] The trim and fill method is an iterative non-parametric technique that uses rank based data augmentation to adjust for publication bias by imputing studies estimated to be missing from the dataset. We use fixed effect models for the primary analysis in this paper; we also reanalysed the data using random effects models as a sensitivity analysis. Stata v.9.2 was used for all the analyses.

Contour enhanced funnel plots
In its simplest form a funnel plot is a scatter plot of study effect sizes (x axis) against their estimated standard errors (y axis).[20] When no bias is present such a plot should be symmetrical, with increasing variability in effect sizes being observed in the less precise studies towards the bottom of the plot, producing a funnel shape. Asymmetry in this plot may indicate that publication biases are present through the lack of observed data points in a region of the plot.[20] Asymmetry alone does not necessarily imply publication biases exist, however, since alternative explanations for the asymmetry may be present.[21] For example, confounding factors (that is, any unmeasured variable associated with both study precision and effect size) may distort the appearance of the plot. It has been observed that certain aspects of trial quality may influence the estimates of effect size,[22-25] and empirical evidence suggests that small studies are, on average, of lower quality and this could induce asymmetry on a funnel plot.[26] Mechanisms such as this lead to what have been termed small study effects,[21 26-28] and their presence will also make funnel plots asymmetrical.

With a view to disentangling genuine publication biases from other causes of funnel asymmetry, the funnel plot can be enhanced by including contours that partition it into areas of statistical significance and non-significance[11 16] based on the standard Wald test, marking traditionally perceived milestones of significance—for example, the 1%, 5%, and 10% levels.[29] In this way the level of statistical significance of every study's effect estimate is identified. Since there is evidence that publication biases are related to these milestones,[30 31] this can aid interpretation of the funnel plot—that is, if studies seem to be missing in areas of statistical non-significance, then this adds credence to the notion that the asymmetry is due to publication biases. In such cases

an attempt should be made to adjust for such biases (in the absence of being able to obtain gold standard data unaffected by publication biases, such as data from regulatory authorities like the FDA). Conversely, if the parts of the funnel where studies are perceived to be missing are found in areas of higher statistical significance, the cause of asymmetry is more likely to be due to factors other than publication biases.

Regression based adjustment

The regression based adjustment method fits a regression line of best fit to the data presented on a funnel plot.[32] An adjusted pooled estimate of effect is obtained by predicting, from the regression line, the pooled effect size for an ideal study of infinite size (hence with zero standard error), which would be located at the top of a funnel plot; since it is hypothesised that there would be no bias in studies of that size. This idea has been discussed in the literature[33-35] (and additionally, such metaregressions are commonly used to test for the presence of publication bias),[13] but only recently has the notion been formally evaluated.[12] In that evaluation the performance of several different regression models was considered over an extensive range of meta-analytical and publication bias scenarios. The best models were shown to consistently outperform the established trim and fill method. One of these, the quadratic version of the original Egger's regression test,[13] is implemented here. This assumes a linear trend between the effect size and its variance (rather than its standard error, as assumed in the original Egger's test). Other models considered in the simulation study were designed for binary outcomes exclusively and are not considered here.

RESULTS

Figure 1A displays a contour enhanced funnel plot of the studies submitted to the FDA, with the corresponding fixed effect meta-analysis pooled estimate providing a weighted average of effect sizes across trials (*g* score 0.31, 95% confidence interval 0.27 to 0.35). This funnel plot is reasonably symmetrical (Egger's test P=0.10), which is consistent with the hypothesis that the FDA is an unbiased and appropriate gold standard data source.

The contour enhanced funnel plot for the journal data (fig 1B) is different and highly asymmetrical (Egger's test P<0.001). A meta-analysis of these data results in a higher average effect size (*g* score 0.41, 0.37 to 0.45). Most of the study estimates now lie above (but many close to) the right contour line, indicating a statistically significant benefit at the 5% level, with few studies located below this 5% contour line—that is, not reaching significance at the 5% level. Crucially, the area where studies seem to be "missing" is contained within the area where non-significant studies would be located; inside the triangle defined by P=0.10 contour boundaries. This adds further credence to the hypothesis that the observed asymmetry is caused by publication biases. Hence, even without the availability of the corresponding funnel plot for the FDA data
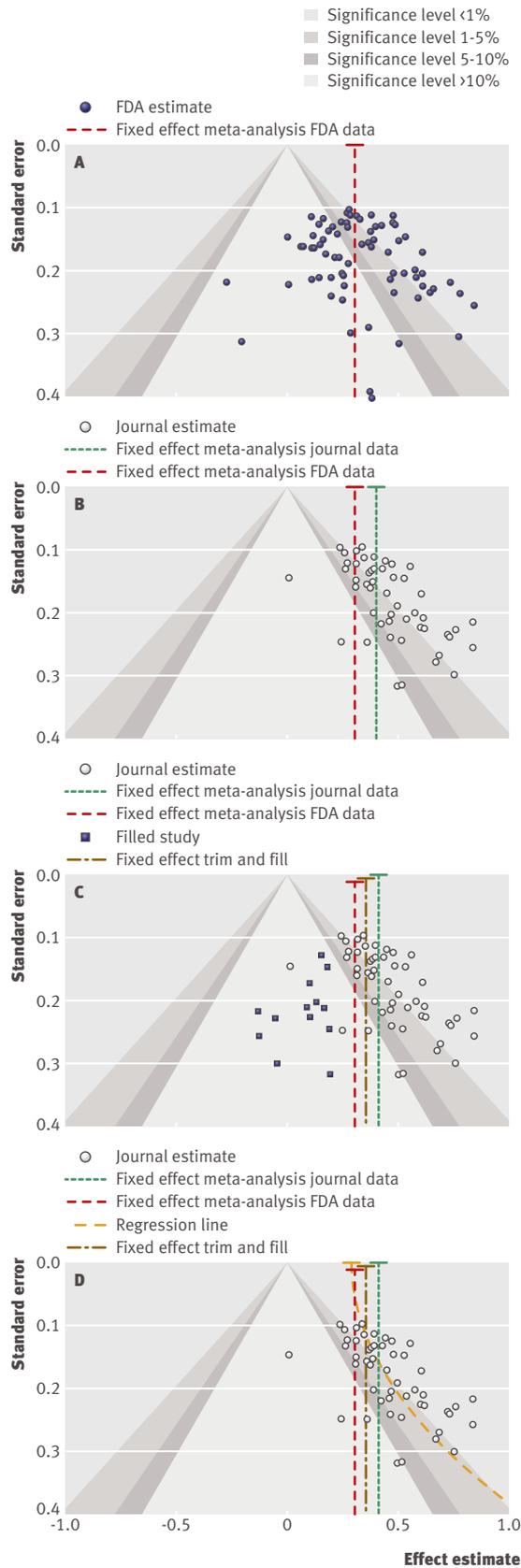
(fig 1A), a contour enhanced funnel plot has convincingly identified publication biases as a major problem for the journal data.

For the journal dataset, the trim and fill method imputed a total of 18 "missing" studies (all in the region of non-statistical significance indicated by squares in figure 1C). This agrees reasonably well with the truth, as 23 studies identified through the FDA registry were not identified in the journal literature. The application of the trim and fill method reduced the average effect size to 0.35 (95% confidence interval 0.31 to 0.39), which is about halfway between the FDA and journal estimates (all three estimates are presented in figure 1C).

The fitted line corresponding to the regression based adjustment method is plotted in figure 1D (orange dashed line). The adjusted estimate is obtained by extrapolating the line to where the standard error is 0 (at the top of figure 1D). This produces an adjusted average effect size of 0.29 (95% confidence interval 0.23 to 0.35), which is close to the estimate produced by the meta-analysis of the FDA data (0.31, 0.27 to 0.35).

The situation is complicated by the fact that among the FDA non-significant studies that were published in medical journals, most were published as if they were significant. This is investigated in figure 2A by linking the effect sizes from each study where estimates were available from both data sources (69% (n=50) of all the trials), using arrows indicating the magnitude and direction of change between FDA and published effect sizes. The effect size differed between FDA and journal analyses in 62% (n=31) of the 50 trials by at least a *g* score of 0.01. Of these, the journal published effects were larger in 77% (n=24) of the studies (arrow pointing to right). As expected, a meta-analysis of these data produces a higher average effect size for the journal data (*g* score=0.41, 95% confidence interval 0.37 to 0.45) compared with the matched FDA data (0.37, 0.33 to 0.41). About eight studies in figure 2 achieve statistical significance at the 5% level when published in medical journals, contradicting their non-significant FDA submission, whereas no journal publication revokes statistical significance previously reported to the FDA. This suggests that reporting biases within published studies are directed towards the realisation of statistical significance. Similarly, 96% (n=21) of the 22 unpublished studies (in journals) were non-significant when submitted to the FDA (fig 2B); which again supports the hypothesis of the presence of publication biases. The fixed effect meta-analysis estimate for these 22 unpublished studies (0.15, 95% confidence interval 0.08 to 0.22) was far lower than the one for published studies (0.41, 0.37 to 0.45; fig 2B), adding further support that serious publication biases are present in the journal data.

A reanalysis of the data using random effects models produced similar results to the fixed effect (proportion of total variability explained by heterogeneity (I^2) was 16% for the FDA data and 0% for the journal data).[36] Details are available on request from the first author.
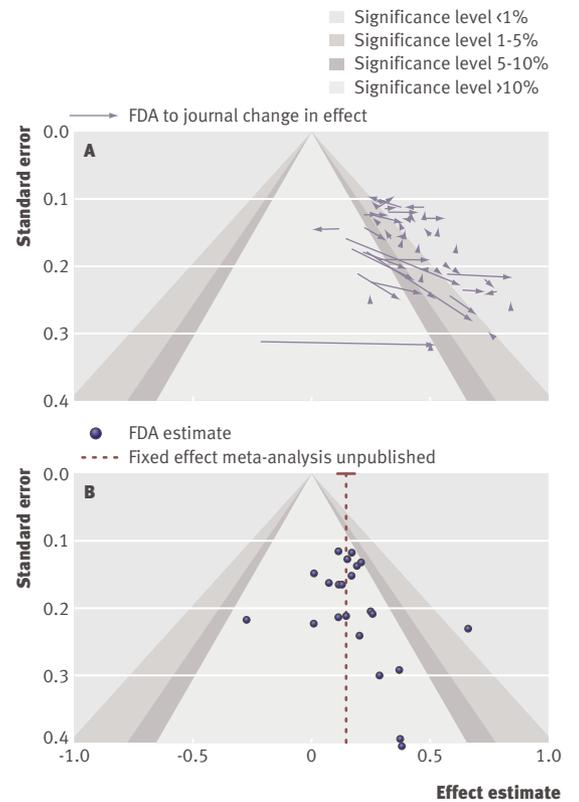
Fig 1 | Contour enhanced funnel plots (95% CI at top). (A) Studies submitted to Food and Drug Administration (FDA). (B) Studies published in journals. (C) Implementation of trim and fill method on journal data. (D) Implementation of regression adjustment model on journal data (adjusted effect at top where SE is 0)

DISCUSSION

The application of two novel approaches to identify and adjust for publication biases in a dataset derived from a journal publication, where a gold standard dataset exists, produced encouraging results. Firstly, detection of publication biases was convincing using a contour enhanced funnel plot. Secondly, the regression based method produced a corrected average effect size, which was close to that obtained from the FDA dataset (and closer than that obtained by the trim and fill method).

This assessment does, however, have limitations. Firstly, the findings relate to a single dataset and thus are not necessarily generalisable to other examples. Specifically, all the trials were sponsored by the pharmaceutical industry and we make the assumption that the FDA data are completely unbiased. Furthermore, the methods under evaluation were designed primarily for the assessment of efficacy outcomes and they might not be appropriate for safety outcomes—for example, there may be incentives to suppress statistically significant safety outcomes (rather than non-significant ones). This is an area that requires more research.

Debate is ongoing about the usefulness of funnel plots and related tests for the identification of publication biases. Although their use is widely advocated[2 37] some question their validity,[27 38-41] including in this

Fig 2 | Contour enhanced funnel plots displaying discrepancy between Food and Drug Administration (FDA) data and journal data. (A) Arrows joining effect results from same studies when both were available from FDA and journals. (B) Estimates of effect only available from FDA (not journal published studies)

journal.[42] We think the analysis presented here provides strong evidence that they do have a useful role.

Recently there has been a lot of research into refining tests for funnel plot asymmetry,[13 26 43-45] and while we support the formalisation of such an assessment, none of the tests (nor trim and fill or the regression adjustment method) considers the statistical significance of the available study estimates. For this reason we think the consideration of the contours on the funnel plot to be an essential component of distinguishing publication biases from other causes of funnel plot asymmetry. We make no claim that the contours can distinguish between the different mechanisms for publication bias—for example, whether it is missing whole studies, selectively reported outcomes, or "massaged" data that have led to the distorted funnel plot. (Because we have the FDA data, we do go on to disentangle this (fig 2) but generally this will not be possible.) But we do not think this is an important limitation because all these biases have the same effect in a meta-analysis—that is, they are all assumed to be related to statistical significance and they all result in an exaggeration of the pooled effect. There is empirical evidence to support this notion for the effect of reporting biases within published clinical trials in general[46-48] and for trials on antidepressants in particular.[1 49 50] Potential mechanisms that are known to induce this include: (a) selectivity in which outcomes are reported or labelled as primary in journal publications; (b) post hoc searches for statistical significance using numerous hypothesis tests—that is, data dredging or fishing; and (c) selectivity in the analysis methods applied to the data for journal publication. Regarding the last point, the FDA makes its recommendations based on the intention to treat principle,[51 52] whereas only half the journal publications are analysed and reported using this approach.[53-56] The usual alternative—the per protocol approach to analysis—excludes dropouts and nonadherents (or patients with protocol deviations in general) and aims to estimate drug efficacy, which will tend to inflate effect sizes compared with the intention to treat approach, which estimates effectiveness.[57-60] An estimate from a per protocol analysis will generally have less precision than for the associated intention to treat analyses owing to the removal of patients with protocol deviations,[61 62] which would result in a shift downwards along the y axis of a funnel plot. This is consistent with what is observed in figure 2A, where most arrows are in a downward (as well as right moving) direction. How much such a mechanism commonly contributes to funnel plot asymmetry would be worthy of further investigation.

Few methods for specifically addressing outcome[63 64] and subgroup reporting biases[65] exist, and further development of analytical methods to specifically tackle aspects of reporting biases within studies is encouraged. Nevertheless, it is reassuring that the methods used in this article to address publication and related biases generally seem to work well in the presence of multiple types of publication biases. We no longer advocate the use of the trim and fill method because of problems identified through simulation studies.[12 40 66] The regression adjustment method, which is easy to carry out,[67] consistently outperformed the trim and fill method in an extensive simulation study[12] (as well as within this particular dataset).

We consider technical issues relating the influence of choice of outcome metric on the robustness of the results, and analyses methods used within the assessments. Firstly, the Hedges' g score outcome metric was used throughout the analysis. This includes a correction for small sample size. An alternative metric, without the correction, is the Cohen's d score, which could also have been used. However this would have negligible influence on the funnel plots presented here since the correction is still modest even for the smallest trials (n=25). An additional consideration is that the contours on the funnels are constructed assuming normality of the effect size since they are based on the Wald test. We acknowledge that this may not be exactly the statistical test used in the original analyses for some of the trials. For example, for trials with small sample sizes, a t test may have been used. However, as the Wald and t test statistics converge as the sample size increases, this is only going to affect the assessment of the most imprecise trials at the bottom of the funnel, and all our findings are clearly robust to this.

The 73 randomised controlled trials considered here correspond to 12 different antidepressants. Despite this, there was little statistical heterogeneity in both datasets and so we carried out fixed effect analyses for simplicity (and findings are consistent if random effects are used). There is an ever present tension in meta-analysis between "lumping and splitting" studies, and an argument could be made for allowing for specific differences in drug treatment by stratifying them and carrying out 12 separate analyses. Challenges would arise if attempting to detect and adjust for publication biases in each of the analyses independently owing to the difficulty of interpreting funnel plots with small numbers of studies and the limited power of statistical methods.[26] We agree with the suggestions of Shang et al,[68] in their assessment of biases in the homoeopathy trial literature (which has some commonalities with the analysis presented here), that it is advantageous to "borrow strength" from a large number of trials and provide empirical information to assist reviewers and readers in the interpretation of findings from small meta-analyses that focus on a specific intervention. Furthermore, investigations of extensions of the existing statistical methods that would formalise such ideas for borrowing strength to produce stratum specific tests and estimates of bias are under way.

Given the apparent biases in the journal based literature for these placebo controlled trials on antidepressants, we are concerned about the validity of the findings of a recent high profile network meta-analysis[69] of non-placebo controlled trials on antidepressants as no assessment of potential publication biases seemed to be carried out.[70]

Undoubtedly the best solution to publication biases is to prevent them from occurring in the first place.[2]

WHAT IS ALREADY KNOWN ON THIS TOPIC

Publication biases exaggerate clinical effects resulting in potentially erroneous clinical decision making

While most of the attention has focused on the non-publication of whole studies, the problem of reporting biases within published studies is receiving increased attention

WHAT THIS STUDY ADDS

Mechanisms including suppression of whole studies, selective outcome reporting, and data "massaging" (for example, selective exclusion of patients from the analysis) may act simultaneously, but may be motivated by underlying statistical significance

Contour enhanced funnel plots and a regression based adjustment method to identify and adjust for multiple publication biases using real data where a gold standard exists showed promising results

Using a gold standard data source, such as the FDA trial registry database, is one way of achieving this. However, this is still a long way off from becoming a reality for many analyses. Hence we often have to rely on analytical methods to deal with the problem, and we believe that the contour enhanced funnel plot and the regression based adjustment method provide important developments in the toolkit to combat publication biases.

1 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;358:252-60.

2 Rothstein HR, Sutton AJ, Borenstein M. *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester: Wiley, 2005.

3 Bax L, Ikeda N, Fukui N, Yaju Y, Tsuruta H, Moons KGM. More than numbers: the power of graphs in meta-analysis. *Am J Epidemiol* 2009;169:249-55.

4 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4:1-115.

5 Egger M, Smith GD. Misleading meta-analysis [editorial]. *BMJ* 1995;310:752-4.

6 Sterne JAC, Egger M, Davey Smith G. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101-5.

7 Egger M, Davey Smith G, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;315:1533-7.

8 Ioannidis JPA. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14.

9 Chan A-W. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med* 2008;5:e230.

10 Turner EH. A taxpayer-funded clinical trials registry and results database. *PLoS Med* 2004;1:e60.

11 Peters J, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol* 2008:991-6.

12 Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol* 2009;9:2.

13 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.

14 Duval S, Tweedie RL. Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63.

15 Hedges LV. Estimating effect size from a series of independent experiments. *Psychol Bull* 1982;92:490-9.

16 Palmer TM, Peters JL, Sutton AJ, Moreno SG. Contour-enhanced funnel plots for meta-analysis. *Stata J* 2008;8:242-54.

17 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320:1574-7.

18 Duval S, Tweedie RL. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* 2000;95:89-98.

19 Steichen TJ. METATRIM: stata module to perform nonparametric analysis of publication bias. *Stata Tech Bull* 2000;61:8-14.

20 Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046-55.

21 McMahon B, Holly L, Harrington R, Roberts C, Green J. Do larger studies find smaller effects? The example of studies for the prevention of conduct disorder. *Eur Child Adolesc Psychiatry* 2008;17:432-7.

22 Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol* 2002;31:1-5.

23 Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513-24.

24 Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847-57.

25 Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.

26 Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119-29.

27 Ioannidis JPA. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract* 2008;14:951-7.

28 Jüni P, Nüesch E, Reichenbach S, Rutjes A, Scherrer M, Bürgi E, et al. Overestimation of treatment effects associated with small sample size in osteoarthritis research. (Abstracts of the 16th Cochrane Colloquium). *German J Qual Health Care* 2008;102:7-99.

29 Gerber AS, Malhotra N. Publication bias in empirical sociological research: do arbitrary significance levels distort published results? *Sociol Methods Res* 2008;37:3-30.

30 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.

31 Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281-6.

32 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559-73.

33 Steichen TJ. METABIAS: tests for publication bias in meta-analysis. *Stata Tech Bull* 1998;7:125-33.

34 DuMouchel W, Normand SL. Computer-modeling and graphical strategies for meta-analysis. In: Stangl DK, Berry DA, eds. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker, 2000:157.

35 Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. *Stat Med* 2008;27:4267-78.

36 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.

37 Sterne JAC, Egger M, Moher D, eds. Chapter 10: addressing reporting biases. In: Higgins JPT, Green S, ed. *Cochrane handbook for systematic reviews of intervention*. Version 5.0.0 (updated Feb 2008). Oxford: Cochrane Collaboration, 2008 (available from www.cochrane-handbook.org).

38 Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ* 2007;176:1091-6.

39 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.

40 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22:2113-26.

41 Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000;53:477-84.

42 Lau J, Ioannidis JPA, Terrin N, Schmid C, Olkin I. The case of the misleading funnel plot. *BMJ* 2006;333:597-600.

43 Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25:3443-57.

44 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295:676-80.

45 Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med* 2008;27:746-63.

46 Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials. Comparison of protocols to published articles. *JAMA* 2004;291:2457-65.

47 Chan A-W, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;17:735-40.

48 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.

49 Hotopf M, Barbui C. Bias in the evaluation of antidepressants. *Epidemiol Psichiatr Soc* 2005;14:55-7.

50 Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007;297:468-70.

51 Heritier SR, Gebski VJ, Keech AC. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. *Med J Aust* 2003;179:438-40.

52 Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med* 1999;18:1903-42.

53 Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.

54 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.

55 Hotopf M, Lewis G, Normand C. Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology. *J Epidemiol Community Health* 1997;51:354-8.

56 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.

57 Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. A simple and valid tool distinguished efficacy from effectiveness studies. *J Clin Epidemiol* 2006;59:1040-8.

58 Bollini P, Pampallona S, Tibaldi G, Kupelnick B, Munizza C. Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials. *Br J Psychiatry* 1999;174:297-303.

59 Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. *Pharmacoeconomics* 1999;15:423-34.

60 Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002;359:781-5.

61 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007;60:663-9.

62 Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-4.

63 Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007;8:9.

64 Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-70.

65 Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-36.

66 Peters JL, Sutton JA, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med* 2007;26:4544-62.

67 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;18:2693-708.

68 Shang A, Huwiler-Müntener K, Nartey L, Jüni P, Dörig S, Sterne JAC, et al. Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *Lancet* 2005;366:726-32.

69 Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JPT, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373:746-58.

70 Turner EH, Moreno SG, Sutton AJ. Concerns about reported rank-order of antidepressant efficacy. *Lancet* 2009;373:1760.