

Chapter 10: Addressing reporting biases

Editors: Jonathan AC Sterne, Matthias Egger and David Moher on behalf of the Cochrane Bias Methods Group.

Copyright © 2008 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd under “The Cochrane Book Series” Imprint.

This extract is made available solely for use in the authoring, editing or refereeing of Cochrane reviews, or for training in these processes by representatives of formal entities of The Cochrane Collaboration. Other than for the purposes just stated, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the copyright holders.

Permission to translate part or all of this document must be obtained from the publishers.

This extract is from *Handbook* version 5.0.1. For guidance on how to cite it, see Section 10.5. The material is also published in Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* (ISBN 978-0470057964) by John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, Telephone (+44) 1243 779777; Email (for orders and customer service enquiries): cs-books@wiley.co.uk. Visit their Home Page on www.wiley.com.

Key points

- Only a proportion of research projects will be published in sources easily identifiable by authors of systematic reviews. Reporting biases arise when the dissemination of research findings is influenced by the nature and direction of results.
- The contribution made to the totality of the evidence in systematic reviews by studies with statistically non-significant results is as important as that from studies with statistically significant results.
- The convincing evidence for the presence of several types of reporting biases (outlined in this chapter) demonstrates the need to search comprehensively for studies that meet the eligibility criteria for a Cochrane review.
- Prospective trial registration, now a requirement for publication in many journals, has the potential to substantially reduce the effects of publication bias.
- Funnel plots can be used for reviews with sufficient numbers of included studies, but an asymmetrical funnel plot should not be equated with publication bias.
- Several methods are available to test for asymmetry in a funnel plot and recommendations are included in the chapter for selecting an appropriate test.

10.1 Introduction

The dissemination of research findings is not a division into published or unpublished, but a continuum ranging from the sharing of draft papers among colleagues, through presentations at meetings and published abstracts, to papers in journals that are indexed in the major bibliographic databases (Smith 1999). It has long been recognized that only a proportion of research projects ultimately reach publication in an indexed journal and thus become easily identifiable for systematic reviews.

Reporting biases arise when the dissemination of research findings is influenced by the nature and direction of results. Statistically significant, ‘positive’ results that indicate that an intervention works are more likely to be published, more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, more likely to be published in high impact journals and, related to the last point, more likely to be cited by others. The contribution made to the totality of the evidence in systematic reviews by studies with non-significant results is as important as that from studies with statistically significant results.

Table 10.1.a summarizes some different types of reporting biases. We consider these in more detail in Section 10.2, highlighting in particular the evidence supporting the presence of each bias. We discuss approaches for avoiding reporting biases in Cochrane reviews in Section 10.3, and address funnel plots and statistical methods for detecting potential biases in Section 10.4. Although for the purpose of discussing these biases we will sometimes denote statistically significant ($P < 0.05$) results as ‘positive’ results and statistically non-significant or null results as ‘negative’ results, such labels should not be used by Cochrane reviews authors.

Table 10.1.a: Definitions of some types of reporting biases

Type of reporting bias	Definition
Publication bias	The <i>publication</i> or <i>non-publication</i> of research findings, depending on the nature and direction of the results
Time lag bias	The <i>rapid</i> or <i>delayed</i> publication of research findings, depending on the nature and direction of the results
Multiple (duplicate) publication bias	The <i>multiple</i> or <i>singular</i> publication of research findings, depending on the nature and direction of the results
Location bias	The publication of research findings in journals with different <i>ease of access</i> or <i>levels of indexing</i> in standard databases, depending on the nature and direction of results.
Citation bias	The <i>citation</i> or <i>non-citation</i> of research findings, depending on the nature and direction of the results
Language bias	The publication of research findings <i>in a particular language</i> , depending on the nature and direction of the results
Outcome reporting bias	The <i>selective reporting</i> of some outcomes but not others, depending on the nature and direction of the

10.2 Types of reporting biases and the supporting evidence

10.2.1 Publication bias

In a 1979 article, “The ‘file drawer problem’ and tolerance for null results”, Rosenthal described a gloomy scenario where “the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant (e.g. $p > 0.05$) results” (Rosenthal 1979). The file drawer problem has long been suspected in the social sciences: a review of psychology journals found that of 294 studies published in the 1950s, 97.3% rejected the null hypothesis at the 5% level ($P < 0.05$) (Sterling 1959). The study was updated and complemented with three other journals (*New England Journal of Medicine*, *American Journal of Epidemiology*, *American Journal of Public Health*) (Sterling 1995). Little had changed in the psychology journals (95.6% reported significant results) and a high proportion of statistically significant results (85.4%) was also found in the general medical and public health journals. Similar results have been reported in many different areas such as emergency medicine (Moscati 1994), alternative and complementary medicine (Vickers 1998, Pittler 2000) and acute stroke trials (Liebeskind 2006).

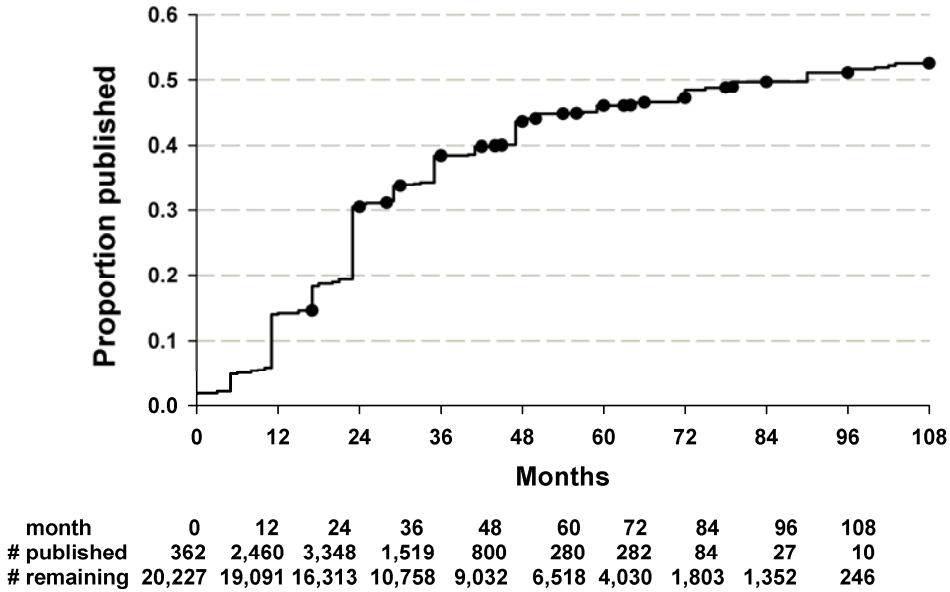
It is possible that studies suggesting a beneficial intervention effect or a larger effect size are published, while a similar amount of data pointing in the other direction remains unpublished. In this situation, a systematic review of the published studies could identify a spurious beneficial intervention effect, or miss an important adverse effect of an intervention. In cardiovascular medicine, investigators who, in 1980, found an increased death rate among patients with acute myocardial infarction treated with a class 1 anti-arrhythmic dismissed it as a chance finding and did not publish their trial at the time (Cowley 1993). Their findings would have contributed to a more timely detection of the increased mortality that has since become known to be associated with the use of class I anti-arrhythmic agents (Teo 1993, CLASP Collaborative Group 1994).

Studies empirically examining the existence of publication bias can be viewed in two categories: indirect and direct evidence. Surveys of published results, such as those described above, can provide only indirect evidence of publication bias, as the proportion of all hypotheses tested for which the null hypothesis is truly false is unknown. There is also substantial direct evidence of publication bias. Roberta Scherer and colleagues recently updated a systematic review which summarizes 79 studies describing subsequent full publication of research initially presented in abstract or short report form (Scherer 2007). The data from 45 studies that included data on time to publication are summarized in [Figure 10.2.a](#). Only about half of abstracts presented at conferences were later published in full (63% for randomized trials), and subsequent publication was associated with positive results, (Scherer 2007).

Additional direct evidence is available from a number of cohort studies of proposals submitted to ethics committees and institutional review boards (Easterbrook 1991, Dickersin 1992, Stern 1997, Decullier 2005, Decullier 2007), trials submitted to licensing authorities (Bardy 1998), analyses of trials registries (Simes 1987) or from cohorts of trials funded by specific funding agencies (Dickersin 1993). For each cohort of research proposals the principal investigators were contacted several years later to determine the publication status of each completed study. In all these studies publication was more likely if the intervention effects were large and statistically significant.

Hopewell et al. recently completed a methodology review of such studies, limited to those that considered clinical trials separately (Hopewell 2008). The percentages of full publication as journal articles in the five studies included in the review ranged from 36% to 94% (Table 10.2.a). Positive results were consistently more likely to have been published than negative results; the odds of publication were approximately four times greater if results were statistically significant (OR = 3.90, 95% CI 2.68 to 5.68) as shown in Figure 10.2.b. Other factors such as the study size, funding source, and academic rank and sex of primary investigator were not consistently associated with the probability of publication or were not possible to assess separately for clinical trials (Hopewell 2008).

Figure 10.2.a: Cumulative full publication of results initially presented as abstracts from 45 studies reporting time to publication that followed up research presented at meetings and conferences



N = 20,227 abstracts
 Circles show points where data censored because reports stopped follow-up.

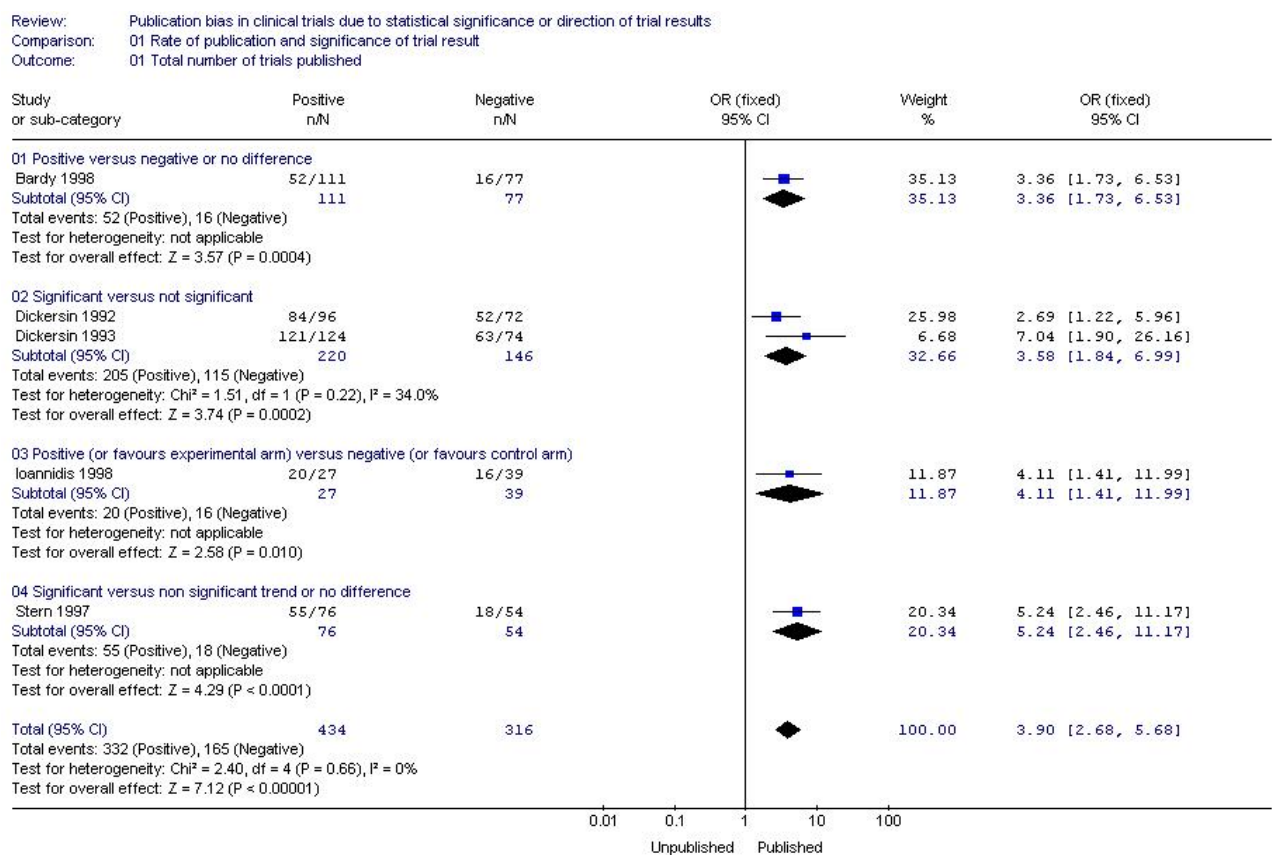
Table 10.2.a: Publication status of five cohorts of research projects approved by ethics committees or institutional review boards which had been completed and analysed at the time of follow up. (Adapted from Hopewell et al. (Hopewell 2008).)

	Johns Hopkins University, Baltimore	National Institutes of Health, U.S.A.	Royal Prince Alfred Hospital, Sydney	National Agency for Medicine, Finland	National Institutes of Health, U.S.A., Multi-centre trials in HIV/AIDS
Reference	Dickersin 1992	Dickersin 1993	Stern 1997	Bardy 1998	Ioannidis 1998
Period of approval	1980	1979	1979-88	1987	1986-1996
Year of follow-up	1988	1988	1992	1995	1996

Number approved	168	198	130	188	66
Published	136 (81%)	184 (93%)	73 (56%)	68 (36%)	36 (54%)
Positive*	84/96 (87%)	121/124 (98%)	55/76 (72%)	52/111 (47%)	20/27 (75%)
Negative*	52/72 (72%)	63/74 (85%)	3/15 (20%)	5/44 (11%)	16/39 (41%)
Inconclusive/null (if assessed separately)	Not assessed	Not assessed	15/39 (38%)	11/33 (33%)	Not assessed

*-Definitions differed by study.

Figure 10.2.b: Publication bias in clinical trials due to statistical significance or direction of trial results (adapted from Hopewell et al. (Hopewell 2008)).



10.2.1.1 Time lag bias

Studies continue to appear in print many years after approval by ethics committees. Hopewell and colleagues reviewed studies examining time to publication for results of clinical trials (Hopewell 2007a). The two studies included in this review (Stern 1997, Ioannidis 1998) found that about half of all trials were published and that those with positive results were published, on average, approximately 2-3 years earlier than trials with null or negative results.

Among proposals submitted to the Royal Prince Alfred Hospital Ethics Committee in Sydney, Australia, an estimated 85% of studies with significant results as compared to 65% of studies with null results had been published after 10 years (Stern 1997). The median time to publication was 4.7 years for studies with significant results and 8.0 years for studies with negative/null results. Similarly, trials

conducted by multi-centre trial groups in the field of HIV infection in the United States appeared on average 4.3 years after the start of patient enrolment if results were statistically significant but took 6.5 years to be published if the results were negative (Ioannidis 1998). A recent study has found similar results (Decullier 2005). The fact that a substantial proportion of studies remain unpublished even a decade after the study had been completed and analysed is troubling as potentially important information remains hidden from systematic reviewers and consumers.

Ioannidis and colleagues also found that trials with positive and negative results differed little in the time they took to complete follow-up (Ioannidis 1998). Rather, the time lag was attributable to differences in the time from completion to publication (Ioannidis 1998). These findings indicate that time lag bias may be introduced in systematic reviews even in situations when most or all studies will eventually be published. Studies with positive results will dominate the literature and introduce bias for several years until the negative, but equally important, results finally appear. Furthermore, rare adverse events are likely to be found later in the research process than short-term beneficial effects.

10.2.1.2 Who is responsible for publication bias?

Studies with negative results could remain unpublished because authors fail to write manuscripts and submit them to journals, because such studies are peer reviewed less favourably, or because editors simply do not want to publish negative results. The peer review process is sometimes unreliable and susceptible to subjectivity, bias and conflict of interest (Peters 1982, Godlee 1999). Experimental studies in which test manuscripts were submitted to peer reviewers or journals showed that peer reviewers are more likely to referee favourably if results were in accordance with their own views (Mahoney 1977, Epstein 1990, Ernst 1994). For example, when a selected group of authors was asked to peer review a fictitious paper on transcutaneous electrical nerve stimulation (TENS) they were influenced by their own findings and preconceptions. Other studies have shown no association between publication of submitted manuscripts and study outcomes (Abbot 1998, Olson 2002), suggesting that although peer reviewers may hold strong beliefs which will influence their assessments, there is no general bias for or against positive findings.

A number of studies have directly asked authors why they had not published their findings. The most frequent answer was that they were not interesting enough to merit publication (e.g. journals would be unlikely to accept the manuscripts) (Easterbrook 1991, Dickersin 1992, Stern 1997, Weber 1998, Decullier 2005) or the investigators did not have enough time to prepare a manuscript (Weber 1998, Hartling 2004). Rejection of a manuscript by a journal was rarely mentioned as a reason for not publishing. Selective submission of papers by authors rather than selective recommendation by peer reviewers and selective acceptance by editors thus appears to be the dominant contributor to publication bias. In addition, Dickersin et al. examined the time from manuscript submission (to the journal *JAMA*) to full publication and found no association between this time and any study characteristics examined, including statistical significance of the study results (Dickersin 2002). Thus, time-lag bias may also be the result of delayed submission of manuscripts for publication by authors rather than by delayed publication by journals.

10.2.1.3 The influence of external funding and commercial interests

External funding has been found to be associated with publication independently of the statistical significance of the results (Dickersin 1997). Funding by government agencies was significantly associated with publication in three cohorts of proposals submitted to ethics committees (Easterbrook 1991, Dickersin 1992, Stern 1997) whereas pharmaceutical industry sponsored studies were less likely to be published in two studies (Easterbrook 1991, Dickersin 1992). Indeed, a large proportion of clinical trials submitted by drug companies to licensing authorities remain unpublished (Hemminki 1980, Bardy 1998).

In a systematic review, Lexchin et al. identified 30 studies published between 1966 and 2002 that examined whether funding of drug studies by the pharmaceutical industry was associated with outcomes that are favourable to the funder. They found that research funded by drug companies was less likely to be published than research funded by other sources, and that studies sponsored by pharmaceutical companies were more likely to have outcomes favouring the sponsor than were studies with other sponsors (Lexchin 2003). Other studies have since examined these associations and have found similar results (Bhandari 2004, Heres 2006). Heres et al., in a study of head-to-head comparisons of antipsychotics, found that the overall outcome of the trials favoured the drug manufactured by the industry sponsor in 90% of studies considered, and that some similar studies reported opposing conclusions, each supporting the product of the study sponsor (Heres 2006).

The implication is that the pharmaceutical industry tends to discourage the publication of negative studies that it has funded. For example, a manuscript reporting on a trial comparing the bioequivalence of generic and brand levothyroxine products, which had failed to produce the results desired by the sponsor of the study, Boots Pharmaceuticals, was withdrawn because Boots took legal action against the university and the investigators. The actions of Boots, recounted in detail by one of the editors of *JAMA*, Drummond Rennie (Rennie 1997), meant that publication of the paper (Dong 1997) was delayed by about seven years. In a national survey of life-science faculty members in the United States, 20% reported that they had experienced delays of more than six months in publication of their work and reasons for not publishing included “to delay the dissemination of undesired results” (Blumenthal 1997). Delays in publication were associated with involvement in commercialization and academic-industry research relationship, as well as with male sex and higher academic rank of the investigator (Blumenthal 1997).

10.2.2 Other reporting biases

While publication bias has long been recognized and much discussed, other factors can contribute to biased inclusion of studies in meta-analyses. Indeed, among published studies, the probability of identifying relevant studies for meta-analysis is also influenced by their results. These biases have received much less consideration than publication bias, but their consequences could be of equal importance.

10.2.2.1 Duplicate (multiple) publication bias

In 1989, Gøtzsche found that, among 244 reports of trials comparing non-steroidal anti-inflammatory drugs in rheumatoid arthritis, 44 (18%) were redundant, multiple publications, which overlapped substantially with a previously published article. Twenty trials were published twice, ten trials three times and one trial four times (Gøtzsche 1989). The production of multiple publications from single studies can lead to bias in a number of ways (Huston 1996). Most importantly, studies with significant results are more likely to lead to multiple publications and presentations (Easterbrook 1991), which makes it more likely that they will be located and included in a meta-analysis. It is not always obvious that multiple publications come from a single study, and one set of study participants may be included in an analysis twice. The inclusion of duplicated data may therefore lead to overestimation of intervention effects, as was demonstrated for trials of the efficacy of ondansetron to prevent postoperative nausea and vomiting (Tramèr 1997).

Other authors have described the difficulties and frustration caused by redundancy and the ‘disaggregation’ of medical research when results from a multi-centre trial are presented in several publications (Huston 1996, Johansen 1999). Redundant publications often fail to cross-reference each other (Bailey 2002, Barden 2003) and there are examples where two articles reporting the same trial do not share a single common author (Gøtzsche 1989, Tramèr 1997). Thus, it may be difficult or impossible for review authors to determine whether two papers represent duplicate publications of one

study or two separate studies without contacting the authors, which may result in biasing a meta-analysis of this data.

10.2.2.2 Location bias

Research suggests that various factors related to the accessibility of study results are associated with effect sizes in trials. For example, in a series of trials in the field of complementary and alternative medicine, Pittler and colleagues examined the relationship between trial outcome, methodological quality and sample size with characteristics of the journals of publication of these trials (Pittler 2000). They found that trials published in low or non-impact factor journals were more likely to report significant results than those published in high-impact mainstream medical journals and that the quality of the trials was also associated with the journal of publication. Similarly, some studies suggest that trials published in English language journals are more likely to show strong significant effects than those published in non-English language journals (Egger 1997b), however this has not been shown consistently (Moher 2000, Jüni 2002, Pham 2005); see Section 10.2.2.4.

The term 'location bias' is also used to refer to the accessibility of studies based on variable indexing in electronic databases. Depending on the clinical question, choices regarding which databases to search may bias the effect estimate in a meta-analysis. For example, one study found that trials published in journals that were not indexed in MEDLINE might show a more beneficial effect than trials published in MEDLINE-indexed journals (Egger 2003). Another study of 61 meta-analyses found that, in general, trials published in journals indexed in EMBASE but not in MEDLINE reported smaller estimates of effect than those indexed in MEDLINE, but that the risk of bias may be minor, given the lower prevalence of the EMBASE unique trials (Sampson 2003). As above, these findings may vary substantially with the clinical topic being examined.

A final form of location bias is regional or developed country bias. Research supporting the evidence of this bias suggests that studies published in certain countries may be more likely than others to produce research showing significant effects of interventions. Vickers and colleagues demonstrated the potential existence of this bias (Vickers 1998).

10.2.2.3 Citation bias

The perusal of the reference lists of articles is widely used to identify additional articles that may be relevant although there is little evidence to support this methodology. The problem with this approach is that the act of citing previous work is far from objective and retrieving literature by scanning reference lists may thus produce a biased sample of studies. There are many possible motivations for citing an article. Brooks interviewed academic authors from various faculties at the University of Iowa and asked for the reasons for citing each reference in one of the authors' recent articles (Brooks 1985). Persuasiveness, i.e. the desire to convince peers and substantiate their own point of view, emerged as the most important reason for citing articles. Brooks concluded that authors advocate their own opinions and use the literature to justify their point of view: "Authors can be pictured as intellectual partisans of their own opinions, scouring the literature for justification" (Brooks 1985).

In Gøtzsche's analysis of trials of non-steroidal anti-inflammatory drugs in rheumatoid arthritis, trials demonstrating a superior effect of the new drug were more likely to be cited than trials with negative results (Gøtzsche 1987). Similar results were shown in an analysis of randomized trials of hepatobiliary diseases (Kjaergard 2002). Similarly, trials of cholesterol lowering to prevent coronary heart disease were cited almost six times more often if they were supportive of cholesterol lowering (Ravnskov 1992). Over-citation of unsupportive studies can also occur. Hutchison et al. examined reviews of the effectiveness of pneumococcal vaccines and found that unsupportive trials were more likely to be cited than trials showing that vaccines worked (Hutchison 1995).

Citation bias may affect the 'secondary' literature. For example, the *ACP Journal Club* aims to summarize original and review articles so that physicians can keep abreast of the latest evidence. However, Carter et al. found that trials with a positive outcome were more likely to be summarized, after controlling for other reasons for selection (Carter 2006). If positive studies are more likely to be cited, they may be more likely to be located and, thus, more likely to be included in a systematic review, thus biasing the findings of the review.

10.2.2.4 Language bias

Reviews have often been exclusively based on studies published in English. For example, among 36 meta-analyses reported in leading English-language general medicine journals from 1991 to 1993, 26 (72%) had restricted their search to studies reported in English (Grégoire 1995). This trend may be changing, with a recent review of 300 systematic reviews finding approximately 16% of reviews limited to trials published in English; systematic reviews published in paper-based journals were more likely than Cochrane reviews to report limiting their search to trials published in English (Moher 2007). In addition, of reviews with a therapeutic focus, Cochrane reviews were more likely than non-Cochrane reviews to report having no language restrictions (62% vs. 26%) (Moher 2007).

Investigators working in a non-English speaking country will publish some of their work in local journals (Dickersin 1994). It is conceivable that authors are more likely to report in an international, English-language journal if results are positive whereas negative findings are published in a local journal. This was demonstrated for the German-language literature (Egger 1997b).

Bias could thus be introduced in reviews exclusively based on English-language reports (Grégoire 1995, Moher 1996). However, the research examining this issue is conflicting. In a study of 50 reviews that employed comprehensive literature searches and included both English and non-English-language trials, Jüni et al. reported that non-English trials were more likely to produce significant results at $P < 0.05$, while estimates of intervention effects were, on average, 16% (95% CI 3% to 26%) more beneficial in non-English-language trials than in English-language trials (Jüni 2002). Conversely, Moher and colleagues examined the effect of inclusion or exclusion of English language trials in two studies of meta-analyses and found, overall, that the exclusion of trials reported in a language other than English did not significantly affect the results of the meta-analyses (Moher 2003). These results were similar when the analysis was limited to meta-analyses of trials of conventional medicines. When the analyses were conducted separately for meta-analyses of trials of complementary and alternative medicines, however, the effect size of meta-analyses significantly decreased by excluding reports in languages other than English (Moher 2003).

The extent and effects of language bias may have diminished recently because of the shift towards publication of studies in English. In 2006, Galandi et al. reported a dramatic decline in the number of randomized trials published in German-language healthcare journals: with fewer than two randomized trials published per journal and year after 1999 (Galandi 2006). While the potential impact of studies published in languages other than English in a meta-analysis may be minimal, it is difficult to predict in which cases this exclusion may bias a systematic review. Review authors may want to search without language restrictions and decisions about including reports from languages other than English may need to be taken on a case-by-case basis.

10.2.2.5 Outcome reporting bias

In many studies, a range of outcome measures is recorded but not all are reported (Pocock 1987, Tannock 1996). The choice of outcomes that are reported can be influenced by the results, potentially making published results misleading. For example, two separate analyses (Mandel 1987, Cantekin

1991) of a double-blind placebo-controlled trial assessing the efficacy of amoxicillin in children with non-suppurative otitis media reached opposite conclusions mainly because different ‘weight’ was given to the various outcome measures that were assessed in the study. This disagreement was conducted in the public arena, since it was accompanied by accusations of impropriety against the team producing the findings favourable to amoxicillin. The leader of this team had received substantial fiscal support, both in research grants and as personal honoraria, from the manufacturers of amoxicillin (Rennie 1991). It is a good example of how reliance upon the data chosen to be presented by the investigators can lead to distortion (Anonymous 1991). Such ‘outcome reporting bias’ may be particularly important for adverse effects. Hemminki examined reports of clinical trials submitted by drug companies to licensing authorities in Finland and Sweden and found that unpublished trials gave information on adverse effects more often than published trials (Hemminki 1980). Since then several other studies have shown that the reporting of adverse events and safety outcomes in clinical trials is often inadequate and selective (Ioannidis 2001, Melander 2003, Heres 2006). A group from Canada, Denmark and the UK recently pioneered empirical research into the selective reporting of study outcomes (Chan 2004a, Chan 2004b, Chan 2005). These studies are described in Chapter 8 (Section 8.13), along with a more detailed discussion of outcome reporting bias.

10.3 Avoiding reporting biases

10.3.1 Implications of the evidence concerning reporting biases

The convincing evidence for the presence of reporting biases, described in Section 10.2, demonstrates the need to search comprehensively for studies that meet the eligibility criteria for a Cochrane review. Review authors should ensure that multiple sources are searched; for example, a search of MEDLINE alone would not be considered sufficient. Sources and methods for searching are described in detail in Chapter 6. Comprehensive searches do not necessarily remove bias, however. Review authors should bear in mind, for example, that study reports may selectively present results; that reference lists may selectively cite sources; and that duplicate publication of results can be difficult to spot. Furthermore, the availability of study information may be subject to time-lag bias, particularly in fast-moving research areas. We now discuss two further means of reducing, or potentially avoiding, reporting biases: the inclusion of unpublished studies, and the use of trial registries.

10.3.2 Including unpublished studies in systematic reviews

Publication bias clearly is a major threat to the validity of any type of review, but particularly of unsystematic, narrative reviews. Obtaining and including data from unpublished trials appears to be one obvious way of avoiding this problem. Hopewell and colleagues conducted a review of studies comparing the effect of the inclusion or exclusion of ‘grey’ literature (defined here as reports that are produced by all levels of government, academics, business and industry in print and electronic formats but that are not controlled by commercial publishers) in meta-analyses of randomized trials (Hopewell 2007b). They included five studies (Fergusson 2000, McAuley 2000, Burdett 2003, Hopewell 2004), all of which showed that published trials had an overall greater intervention effect than grey trials. A meta-analysis of three of these studies suggested that, on average, published trials showed a 9% larger intervention effect than grey trials (Hopewell 2007b).

The inclusion of data from unpublished studies can itself introduce bias. The studies that can be located may be an unrepresentative sample of all unpublished studies. Unpublished studies may be of lower methodological quality than published studies: a study of 60 meta-analyses that included published and unpublished trials found that unpublished trials were less likely to conceal intervention allocation adequately and to blind outcome assessments (Egger 2003). In contrast, Hopewell and colleagues found no difference in the quality of reporting of this information (Hopewell 2004).

A further problem relates to the willingness of investigators of located unpublished studies to provide data. This may depend upon the findings of the study, more favourable results being provided more readily. This could again bias the findings of a systematic review. Interestingly, when Hetherington et al., in a massive effort to obtain information about unpublished trials in perinatal medicine, approached 42,000 obstetricians and paediatricians in 18 countries they identified only 18 unpublished trials that had been completed for more than two years (Hetherington 1989).

A questionnaire assessing the attitudes toward inclusion of unpublished data was sent to the authors of 150 meta-analyses and to the editors of the journals that published them (Cook 1993). Researchers and editors differed in their views about including unpublished data in meta-analyses. Support for the use of unpublished material was evident among a clear majority (78%) of meta-analysts while journal editors were less convinced (47%) (Cook 1993). This study was recently repeated, with a focus on the inclusion of grey literature in systematic reviews, and it was found that acceptance of inclusion of grey literature has increased and, although differences between groups remain (systematic review authors: 86%, editors: 69%), they may have decreased compared with the data presented by Cook et al. (Tetzlaff 2006).

Reasons for reluctance to include grey literature included the absence of peer-review of unpublished literature. It should be kept in mind, however, that the refereeing process has not always been a successful way of ensuring that published results are valid (Godlee 1999). The team involved in preparing a Cochrane review should have at least a similar level of expertise with which to appraise unpublished studies as a peer reviewer for a journal. On the other hand, meta-analyses of unpublished data from interested sources are clearly a cause for concern.

10.3.3 Trial registries and publication bias

In September 2004 a number of major medical journals belonging to the International Committee of Medical Journal Editors (ICMJE) announced they would no longer publish trials that were not registered at inception (Abbasi 2004). All trials that began enrolment of participants after September 2005 had to be registered in a public trials registry at or before the onset of enrolment to be considered for publication in those journals. The ICMJE described 'acceptable' registers: electronically searchable, freely accessible to the public, open to all registrants, and managed by a non-profit organization. Similarly, the ICMJE asks clinical trialists to adhere to a minimum dataset proposed by the World Health Organization.

If this long-overdue initiative is successful, it has the potential to substantially reduce the effects of publication bias. However this would depend on review authors identifying all relevant trials by searching online trial registries, and also on the results of unpublished trials identified via registries being made available to them. Initiatives to mandate the registration of trial results are currently in the early stages, are evolving quickly, and should have an impact on the accessibility of these data. While there is emerging evidence suggesting that some of the data fields requested in the registries are incomplete (Zarin 2005), this is likely to improve over time. The extent to which trial registration will facilitate the work of Cochrane review authors is unclear at present. For advice on searching trial registries, see Chapter 6 (Section 6.2.3).

10.4 Detecting reporting biases

10.4.1 Funnel plots

A funnel plot is a simple scatter plot of the intervention effect estimates from individual studies against some measure of each study's size or precision. In common with forest plots, it is most common to plot the effect estimates on the horizontal scale, and thus the measure of study size on the

vertical axis. This is the opposite of conventional graphical displays for scatter plots, in which the outcome (e.g. intervention effect) is plotted on the vertical axis and the covariate (e.g. study size) is plotted on the horizontal axis.

The name 'funnel plot' arises from the fact that precision of the estimated intervention effect increases as the size of the study increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot should approximately resemble a symmetrical (inverted) funnel. This is illustrated in Panel A of [Figure 10.4.a](#), in which the effect estimates in the larger studies are close to the true intervention odds ratio of 0.4.

If there is bias, for example because smaller studies without statistically significant effects (shown as open circles in [Figure 10.4.a](#), Panel A) remain unpublished, this will lead to an asymmetrical appearance of the funnel plot with a gap in a bottom corner of the graph (Panel B). In this situation the effect calculated in a meta-analysis will tend to overestimate the intervention effect (Egger 1997a, Villar 1997). The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

Funnel plots were first used in educational research and psychology, with effect estimates plotted against total sample size (Light 1984). It is now usually recommended that the standard error of the intervention effect estimate be plotted, rather than the total sample size, on the vertical axis (Sterne 2001). This is because statistical power of a trial is determined by factors in addition to sample size, such as the number of participants experiencing the event for dichotomous outcomes, and the standard deviation of responses for continuous outcomes. For example, a study with 100,000 participants and 10 events is less likely to show a statistically significant intervention effect than a study with 1000 participants and 100 events. The standard error summarizes these other factors. Plotting standard errors on a reversed scale places the larger, or most powerful, studies towards the top of the plot. Another potential advantage of using standard errors is that a simple triangular region can be plotted, within which 95% of studies would be expected to lie in the absence of both biases and heterogeneity. These regions are included in [Figure 10.4.a](#). Funnel plots of effect estimates against their standard errors (on a reversed scale) can be created using RevMan. A triangular 95% confidence region based on a fixed-effect meta-analysis can be included in the plot, and different plotting symbols allow studies in different subgroups to be identified.

Publication bias need not lead to asymmetry in funnel plots. In the absence of any intervention effect, selective publication based on the P value alone will lead to a symmetrical funnel plot in which studies on the extreme left or right are more likely to be published than those in the middle. This could bias the estimated between-study heterogeneity variance.

Ratio measures of intervention effect (such as odds ratios and risk ratios) should be plotted on a logarithmic scale. This ensures that effects of the same magnitude but opposite directions (for example odds ratios of 0.5 and 2) are equidistant from 1.0. For outcomes measured on a continuous (numerical) scale (e.g. blood pressure, depression score) intervention effects are measured as mean differences or standardized mean differences, which should therefore be used as the horizontal axis in funnel plots. So far as we are aware, no empirical investigations have examined choice of axes for funnel plots for continuous outcomes. For mean differences, the standard error is approximately proportional to the inverse of the square root of the number of participants, and therefore seems an uncontroversial choice for the vertical axis.

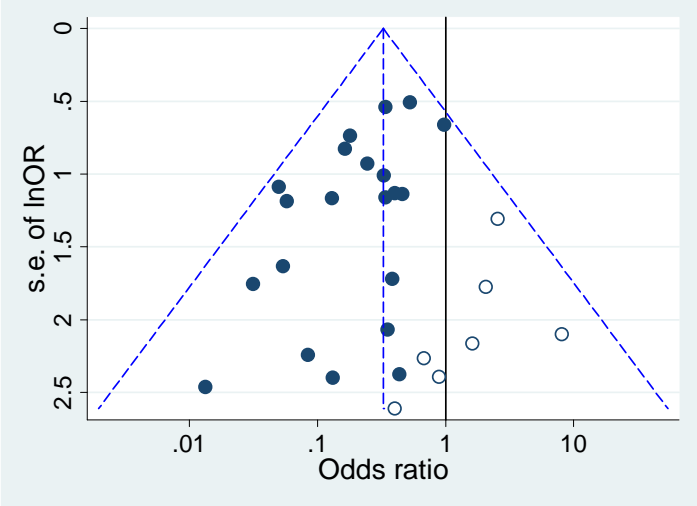
Some authors have argued that visual interpretation of funnel plots is too subjective to be useful. In particular, Terrin et al. found that researchers had only a limited ability to correctly identify funnel plots from meta-analyses subject to publication bias (Terrin 2005).

A further, important, problem with funnel plots is that some effect estimates (e.g. odds ratios and standardized mean differences) are naturally correlated with their standard errors, and can produce spurious asymmetry in a funnel plot. We discuss this problem in more detail in Section 10.4.3.

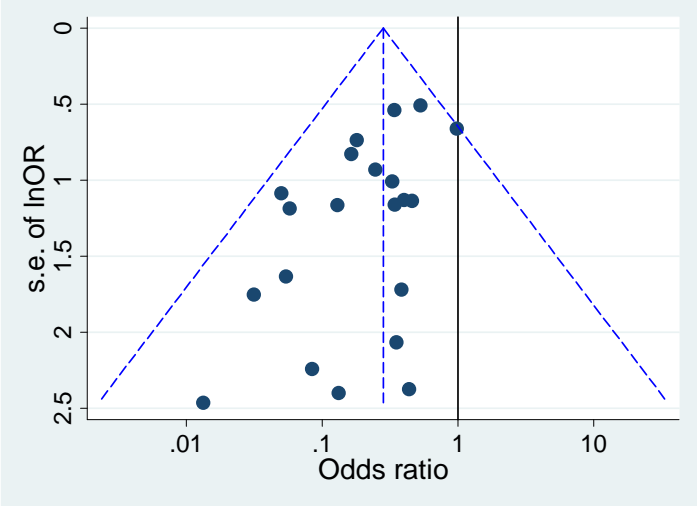
Figure 10.4.a: Hypothetical funnel plots

Panel A: symmetrical plot in the absence of bias. Panel B: asymmetrical plot in the presence of reporting bias. Panel C: asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates.

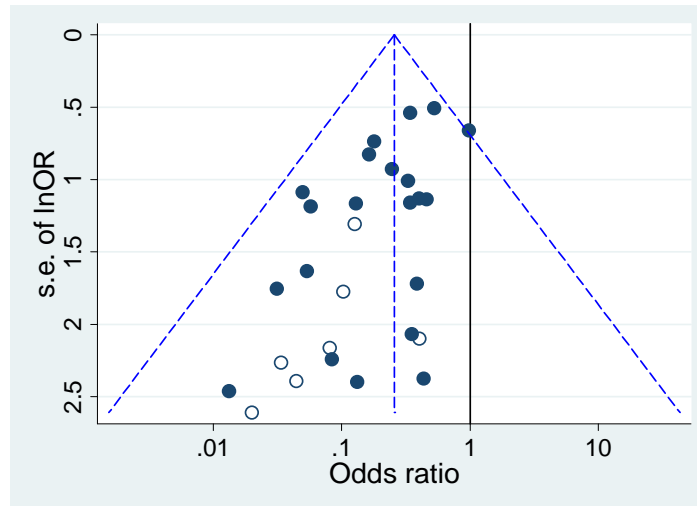
Panel A



Panel B



Panel C



10.4.2 Different reasons for funnel plot asymmetry

Although funnel plot asymmetry has long been equated with publication bias (Light 1984, Begg 1988), the funnel plot should be seen as a generic means of displaying *small-study effects* – a tendency for the intervention effects estimated in smaller studies to differ from those estimated in larger studies (Sterne 2000). Small-study effects may be due to reasons other than publication bias (Egger 1997a, Sterne 2000). Some of these are shown in Table 10.4.a.

Differences in methodological quality are an important potential source of funnel plot asymmetry. Smaller studies tend to be conducted and analysed with less methodological rigour than larger studies (Egger 2003). Trials of lower quality also tend to show larger intervention effects (Schulz 1995). Therefore trials that would have been ‘negative’, if conducted and analysed properly, may become ‘positive’ (Figure 10.4.a, Panel C).

True heterogeneity in intervention effects may also lead to funnel plot asymmetry. For example, substantial benefit may be seen only in patients at high risk for the outcome which is affected by the intervention and these high risk patients are usually more likely to be included in early, small studies (Davey Smith 1994, Glasziou 1995). In addition, small trials are generally conducted before larger trials are established and in the intervening years standard treatment may have improved (resulting in smaller intervention effects in the larger trials). Furthermore, some interventions may have been implemented less thoroughly in larger trials and may, therefore, have resulted in smaller estimates of the intervention effect (Stuck 1998). Finally, it is of course possible that an asymmetrical funnel plot arises merely by the play of chance. Terrin et al. have suggested that the funnel plot is inappropriate for heterogeneous meta-analyses, drawing attention to the premise that the studies come from a single underlying population given by the originators of the funnel plot (Light 1984, Terrin 2003).

A proposed enhancement (Peters 2008) to the funnel plot is to include contour lines corresponding to perceived ‘milestones’ of statistical significance ($P = 0.01, 0.05, 0.1$ etc). This allows the statistical significance of study estimates, and areas in which studies are perceived to be missing, to be considered. Such ‘contour-enhanced’ funnel plots may help review authors to differentiate asymmetry due to publication bias from that due to other factors. For example if studies appear to be missing in areas of statistical non-significance (see Figure 10.4.b, Panel A for an example) then this adds credence to the possibility that the asymmetry is due to publication bias. Conversely, if the supposed missing studies are in areas of higher statistical significance (see Figure 10.4.b, Panel B for an example), this would suggest the cause of the asymmetry may be more likely to be due to factors other

than publication bias (see Table 10.4.a). If there are no statistically significant studies then publication bias may not be a plausible explanation for funnel plot asymmetry (Ioannidis 2007b).

In interpreting funnel plots, systematic review authors thus need to distinguish the different possible reasons for funnel plot asymmetry listed in Table 10.4.a. Knowledge of the particular intervention, and the circumstances in which it was implemented in different studies, can help identify true heterogeneity as a cause of funnel plot asymmetry. There remains a concern that visual interpretation of funnel plots is inherently subjective. Therefore, we now discuss statistical tests for funnel plot asymmetry, and the extent to which they may assist in the objective interpretation of funnel plots. When review authors are concerned that small study effects are influencing the results of a meta-analysis, they may want to conduct sensitivity analyses in order to explore the robustness of the meta-analysis' conclusions to different assumptions about the causes of funnel plot asymmetry: these are discussed in Section 10.4.4.

Table 10.4.a: Possible sources of asymmetry in funnel plots

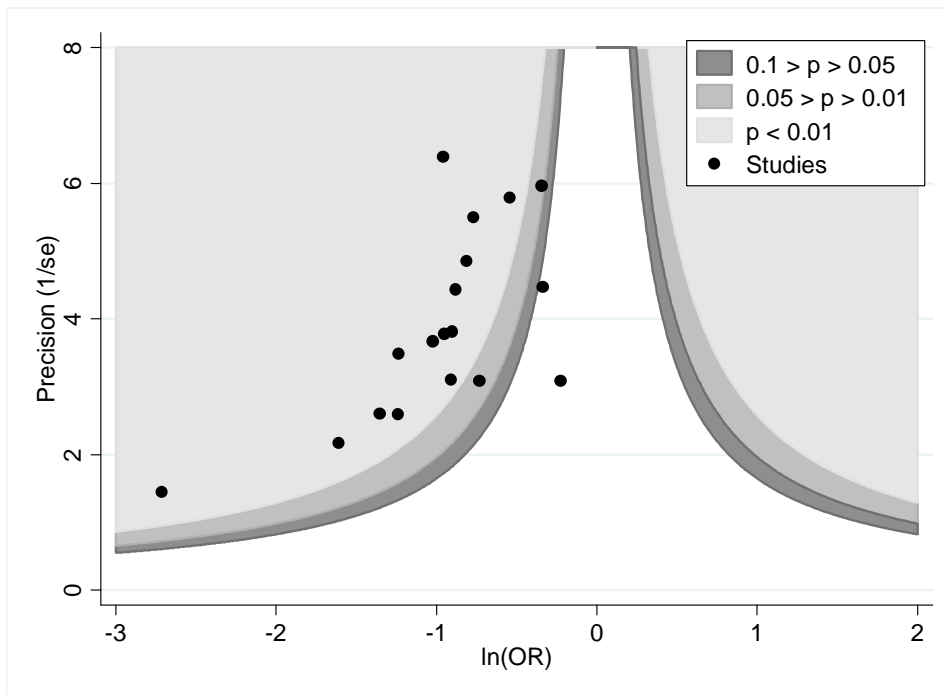
Adapted from Egger et al. (Egger 1997a)).

<p>1. Selection biases:</p> <ul style="list-style-type: none"> • Publication bias: <ul style="list-style-type: none"> ○ Delayed publication (also known as ‘time-lag’ or ‘pipeline’) bias. ○ Location biases: <ul style="list-style-type: none"> ▪ Language bias; ▪ Citation bias; ▪ Multiple publication bias. • Selective outcome reporting.
<p>2. Poor methodological quality leading to spuriously inflated effects in smaller studies:</p> <ul style="list-style-type: none"> • Poor methodological design; • Inadequate analysis; • Fraud.
<p>3. True heterogeneity:</p> <ul style="list-style-type: none"> • Size of effect differs according to study size (for example, due to differences in the intensity of interventions or differences in underlying risk between studies of different sizes).
<p>4. Artefactual:</p> <ul style="list-style-type: none"> • In some circumstances (see Section 10.4.3), sampling variation can lead to an association between the intervention effect and its standard error.
<p>5. Chance.</p>

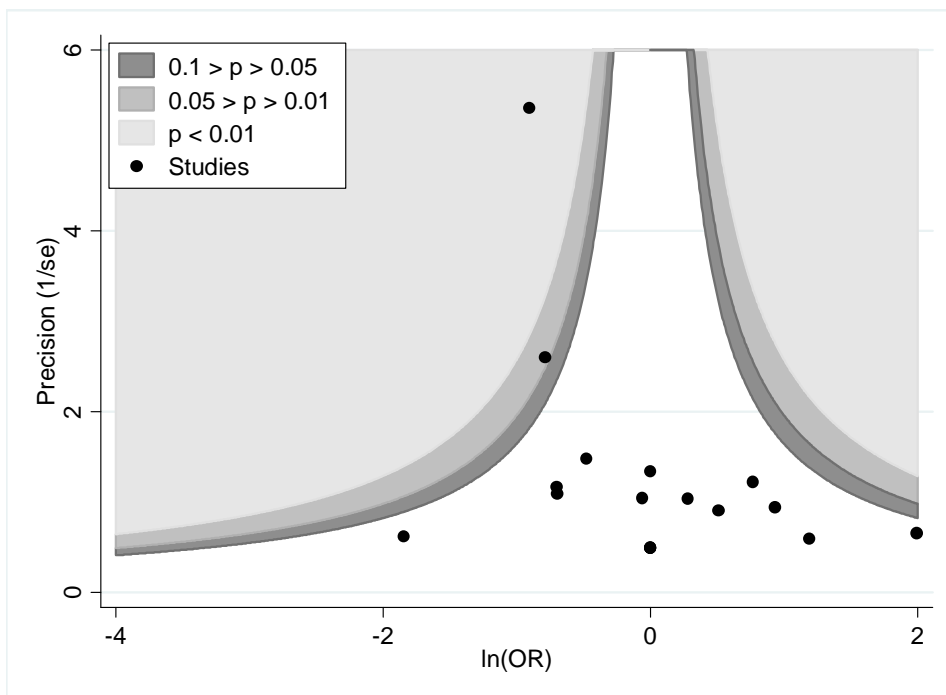
Figure 10.4.b: Contour-enhanced funnel plots

Panel A: there is a suggestion of missing studies on the right-hand-side of the plot, broadly in the area of non-significance (i.e. the white area where $P > 0.1$) for which publication bias is a plausible explanation. Panel B: there is a suggestion of missing studies on the bottom left-hand-side of the plot. Since the majority of this area contains regions of high statistical significance (i.e. indicated by darker shading), this reduces the plausibility that publication bias is the underlying cause of this funnel asymmetry.

Panel A



Panel B



10.4.3 Tests for funnel plot asymmetry

A test for funnel plot asymmetry (small study effects) formally examines whether the association between estimated intervention effects and a measure of study size (such as the standard error of the intervention effect) is greater than might be expected to occur by chance. For outcomes measured on a continuous (numerical) scale this is reasonably straightforward. Using an approach proposed by Egger et al. (Egger 1997a), we can perform a linear regression of the intervention effect estimates on their standard errors, weighting by $1/(\text{variance of the intervention effect estimate})$. This looks for a straight-

line relationship between intervention effect and its standard error. Under the null hypothesis of no small study effects (e.g. Panel A in [Figure 10.4.a](#)) such a line would be vertical. The greater the association between intervention effect and standard error (e.g. as in Panel B in [Figure 10.4.a](#)), the more the slope would move away from vertical. Note that the weighting is important to ensure the regression estimates are not dominated by the smaller studies.

When outcomes are dichotomous, and intervention effects are expressed as odds ratios, the approach proposed by Egger et al. (Egger 1997a) corresponds to a linear regression of the log odds ratio on its standard error, weighted by the inverse of the variance of the log odds ratio (Sterne 2000). This has been by far the most widely used and cited approach to testing for funnel plot asymmetry. Unfortunately, there are statistical problems with this approach, because the standard error of the log odds ratio is mathematically linked to the size of the odds ratio, even in the absence of small study effects (Irwig 1998) (see Deeks et al. for an algebraic explanation of this phenomenon (Deeks 2005)). This can cause funnel plots plotted using log odds ratios (or odds ratios on a log scale) to appear asymmetric and can mean that P values from the test of Egger et al. are too small, leading to false-positive test results. These problems are especially prone to occur when the intervention has a large effect, there is substantial between-study heterogeneity, there are few events per study, or when all studies are of similar sizes.

A number of authors have therefore proposed alternative tests for funnel plot asymmetry: these are summarized in [Table 10.4.b](#). Because it is impossible to know the precise mechanism for publication bias, simulation studies (in which the tests are evaluated on a large number of computer-generated datasets) are required to evaluate the characteristics of the tests under a range of assumptions about the mechanism for publication bias (Sterne 2000, Macaskill 2001, Harbord 2006, Peters 2006, Schwarzer 2007). The most comprehensive study (in terms of scenarios examined, simulations carried out and the range of tests compared) was reported by Rücker et al. (Rücker 2008). Results of this and the other published simulation studies inform the recommendations on testing for funnel plot asymmetry below. Although simulation studies provide useful insights, they inevitably evaluate circumstances that differ from a particular meta-analysis of interest, so their results must be interpreted carefully.

Most of this methodological work has focused on intervention effects measured as odds ratios. While it seems plausible to expect that corresponding problems will arise for intervention effects measured as risk ratios or standardized mean differences, further investigations of these situations are required.

There is ongoing debate over the representativeness of the parameter values used in the simulation studies, and the mechanisms used to simulate publication bias and small study effects, which are often chosen with little explicit justification. Some potentially useful variations on the different tests remain unexamined. Therefore it is not possible to make definitive recommendations on choice of tests for funnel plot asymmetry. Nevertheless, we can identify three tests that should be considered by review authors wishing to test for funnel plot asymmetry.

None of the tests described here is implemented in RevMan, and consultation with a statistician is recommended for their implementation.

Table 10.4.b: Proposed tests for funnel plot asymmetry

N_{tot} is the total sample size, N_E and N_C are the sizes of the experimental and control intervention groups, S is the total number of events across both groups and $F = N_{\text{tot}} - S$. Note that only the first three of these tests (Begg 1994, Egger 1997a, Tang 2000) can be used for continuous outcomes.

Reference	Basis of test
(Begg 1994)	Rank correlation between standardized intervention effect and its standard error.
(Egger 1997a)	Linear regression of intervention effect estimate against its standard error, weighted by the inverse of the variance of the intervention effect estimate.
(Tang 2000)	Linear regression of intervention effect estimate on $1/\sqrt{N_{\text{tot}}}$, with weights N_{tot} .
(Macaskill 2001)*	Linear regression of intervention effect estimate on N_{tot} , with weights $S \times F/N_{\text{tot}}$.
(Deeks 2005)*	Linear regression of log odds ratio on $1/\sqrt{\text{ESS}}$ with weights ESS, where effective sample size $\text{ESS} = 4N_E \times N_C / N_{\text{tot}}$.
(Harbord 2006)*	Modified version of the test proposed by Egger et al., based on the 'score' (O-E) and 'score variance' (V) of the log odds ratio.
(Peters 2006)*	Linear regression of intervention effect estimate on $1/N_{\text{tot}}$, with weights $S \times F/N_{\text{tot}}$.
(Schwarzer 2007)*	Rank correlation test, using mean and variance of the non-central hypergeometric distribution.
(Rücker 2008)	Test based on arcsine transformation of observed risks, with explicit modelling of between-study heterogeneity.

* Test formulated in terms of odds ratios, but may be applicable to other measures of intervention effect.

10.4.3.1 Recommendations on testing for funnel plot asymmetry

For all types of outcome:

- As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the meta-analysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry.
- Tests for funnel plot asymmetry should not be used if all studies are of similar sizes (similar standard errors of intervention effect estimates). However, we are not aware of evidence from simulation studies that provides specific guidance on when study sizes should be considered 'too similar'.
- Results of tests for funnel plot asymmetry should be interpreted in the light of visual inspection of the funnel plot. For example, do small studies tend to lead to more or less beneficial intervention effect estimates? Are there studies with markedly different intervention effect estimates (outliers), or studies that are highly influential in the meta-analysis? Is a small P value caused by one study alone? Examining a contour-enhanced funnel plot, as outlined in Section 10.4.1, may further help interpretation of a test result.
- When there is evidence of small-study effects, publication bias should be considered as only one of a number of possible explanations (see Table 10.4.a). Although funnel plots, and tests for funnel plot asymmetry, may alert review authors to a problem which needs considering, they do not provide a solution to this problem.
- Finally, review authors should remember that, because the tests typically have relatively low power, even when a test does not provide evidence of funnel plot asymmetry, bias (including publication bias) cannot be excluded.

For continuous outcomes with intervention effects measured as mean differences:

- The test proposed by (Egger 1997a) may be used to test for funnel plot asymmetry. There is currently no reason to prefer any of the more recently proposed tests in this situation, although their relative advantages and disadvantages have not been formally examined. While we know of no research specifically on the power of the approach in the continuous case, general considerations suggest that the power will be greater than for dichotomous outcomes, but that use of the method with substantially fewer than 10 studies would be unwise.

For dichotomous outcomes with intervention effects measured as odds ratios:

- The tests proposed by Harbord et al. (Harbord 2006) and Peters et al. (Peters 2006) avoid the mathematical association between the log odds ratio and its standard error (and hence false-positive test results) that occurs for the test proposed by Egger et al. when there is a substantial intervention effect, while retaining power compared with alternative tests. However, false-positive results may still occur in the presence of substantial between-study heterogeneity.
- The test proposed by Rucker et al. (Rucker 2008) avoids false-positive results both when there is a substantial intervention effect and in the presence of substantial between-study heterogeneity. As a rule of thumb, when the estimated between-study heterogeneity variance of log odds ratios, tau-squared, is more than 0.1, only the version of the arcsine test including random-effects (referred to as 'AS+RE' by Rucker et al.) has been shown to work reasonably well. However it is slightly conservative in the absence of heterogeneity, and its interpretation is less familiar because it is based on an arcsine transformation. (Note that although this recommendation is based on the magnitude of tau-squared other factors, including the sizes of the different studies and their distribution, influence a test's performance. We are not currently able to incorporate these other factors in our recommendations).
- When the heterogeneity variance tau-squared is less than 0.1, one of the tests proposed by Harbord 2006, Peters 2006 or Rucker 2008 can be used. (Test performance generally deteriorates as tau-squared increases).
- As far as possible, review authors should specify their testing strategy in advance (noting that test choice may be dependent on the degree of heterogeneity observed). They should apply only one test, appropriate to the context of the particular meta-analysis, from the above-recommended list and report only the result from their chosen test. Application of two or more tests is undesirable since the most extreme (largest or smallest) P value from a set of tests does not have a well-characterized interpretation.

For dichotomous outcomes with intervention effects measured as risk ratios or risk differences, and continuous outcomes with intervention effects measured as standardized mean differences:

- Potential problems in funnel plots have been less extensively studied for these effect measures than for odds ratios, and firm guidance is not yet available.
- Meta-analyses of risk differences are generally considered less appropriate than meta-analyses using a ratio measure of effect (see Chapter 9, Section 9.4.4.4). For similar reasons, funnel plots using risk differences should seldom be of interest. If the risk ratio (or odds ratio) is constant across studies, then a funnel plot using risk differences will be asymmetrical if smaller studies have higher (or lower) baseline risk.

Based on a survey of meta-analyses published in the *Cochrane Database of Systematic Reviews*, these criteria imply that tests for funnel plot asymmetry should be used in only a minority of meta-analyses (Ioannidis 2007b).

Tests for which there is insufficient evidence to recommend use

The following comments apply to all intervention measures. The test proposed by Begg and Mazumdar (Begg 1994) has the same statistical problems but lower power than the test of Egger et al., and is therefore not recommended. The test proposed by Tang and Liu (Tang 2000) has not been evaluated in simulation studies, while the test proposed by Macaskill et al. (Macaskill 2001) has lower power than more recently proposed alternatives. The test proposed by Schwarzer et al. (Schwarzer 2007) avoids the mathematical association between the log odds ratio and its standard error, but has low power relative to the tests discussed above.

In the context of meta-analyses of intervention studies considered in this chapter, the test proposed by Deeks et al. (Deeks 2005) is likely to have lower power than more recently proposed alternatives. This test was not designed as a test for publication bias in systematic reviews of randomized trials: rather it is aimed at meta-analyses of diagnostic test accuracy studies, where very large odds ratios and very imbalanced studies cause problems for other tests.

10.4.4 Sensitivity analyses

When review authors find evidence of small-study effects, they should consider sensitivity analyses examining how the results of the meta-analysis change under different assumptions relating to the reasons for these effects. We stress the exploratory nature of such analysis, due to the inherent difficulty in adjusting for publication bias and lack of research into the performance of such methods applied conditionally based on the results of tests for publication bias considered in Section 10.4.3. This area is relatively underdeveloped; the following approaches have been suggested.

10.4.4.1 Comparing fixed and random-effects estimates

In the presence of heterogeneity, a random-effects meta-analysis weights the studies relatively more equally than a fixed-effect analysis. It follows that in the presence of small-study effects such as those displayed in Figure 10.2.a, in which the intervention effect is more beneficial in the smaller studies, the random-effects estimate of the intervention effect will be more beneficial than the fixed-effect estimate. Poole and Greenland summarized this by noting that “random-effects meta-analyses are not always conservative” (Poole 1999). This issue is also discussed in Chapter 9 (Section 9.5.4).

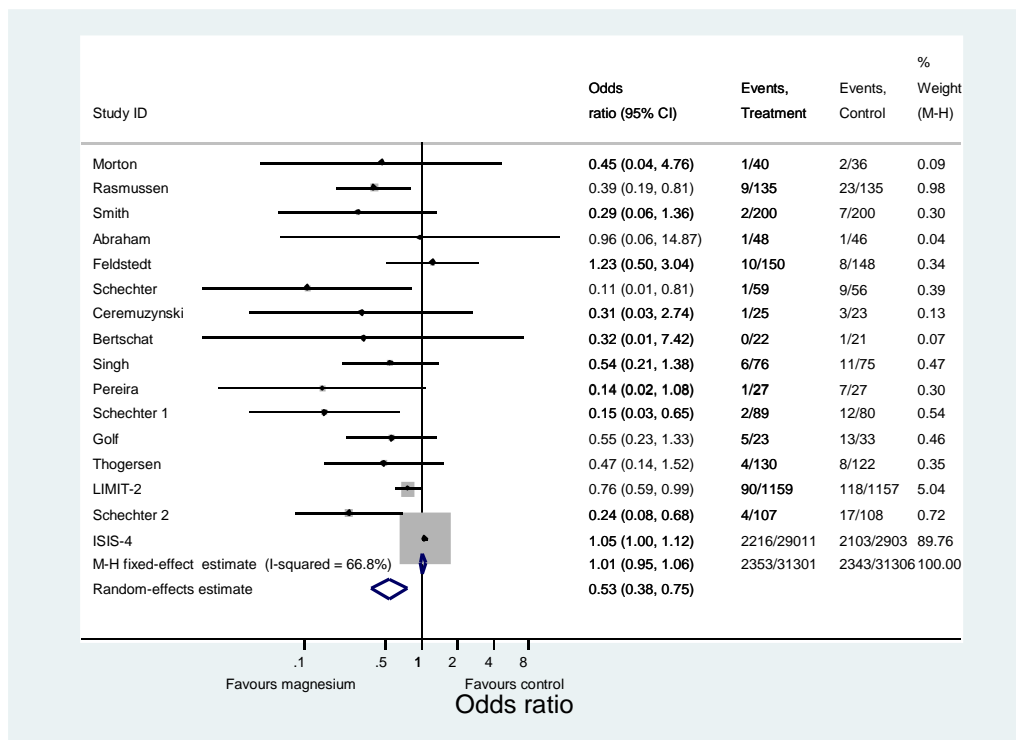
An extreme example of the differences between fixed- and random-effects analyses that can arise in the presence of small-study effects is shown in Figure 10.4.c, which displays both fixed- and random-effects estimates of the effect of intravenous magnesium on mortality following myocardial infarction. This is a well-known example in which beneficial effects of intervention were found in a meta-analysis of small studies, subsequently contradicted when the very large ISIS-4 study found no evidence that magnesium affected mortality.

Because there is substantial between-trial heterogeneity, the studies are weighted much more equally in the random-effects analysis than in the fixed-effect analysis. In the fixed-effect analysis the ISIS-4 trial gets 90% of the weight and so there is no evidence of a beneficial intervention effect. In the random-effects analysis the small studies dominate, and there appears to be clear evidence of a beneficial effect of intervention. To interpret the accumulated evidence, it is necessary to make a judgement about the likely validity of the combined evidence from the smaller studies, compared with that from the ISIS-4 trial.

We recommend that when review authors are concerned about the influence of small-study effects on the results of a meta-analysis in which there is evidence of between-study heterogeneity ($I^2 > 0$), they compare the fixed- and random-effects estimates of the intervention effect. If the estimates are similar,

then any small-study effects have little effect on the intervention effect estimate. If the random-effects estimate is more beneficial, review authors should consider whether it is reasonable to conclude that the intervention was more effective in the smaller studies. If the larger studies tend to be those conducted with more methodological rigour, or conducted in circumstances more typical of the use of the intervention in practice, then review authors should consider reporting the results of meta-analyses restricted to the larger, more rigorous studies. Formal evaluation of such strategies in simulation studies would be desirable. Note that formal statistical comparisons of the fixed- and random-effects estimates of intervention effect are not possible, and that it is still possible for small-study effects to bias the results of a meta-analysis in which there is no evidence of heterogeneity, even though the fixed- and random-effects estimates of intervention effect will be identical in this situation.

Figure 10.4.c: Comparison of fixed- and random-effects meta-analytic estimates of the effect of intravenous magnesium on mortality following myocardial infarction



10.4.4.2 Trim and fill

The ‘trim and fill’ method aims both to identify and correct for funnel plot asymmetry arising from publication bias (Taylor 1998, Duval 2000). The basis of the method is to (1) ‘trim’ (remove) the smaller studies causing funnel plot asymmetry, (2) use the trimmed funnel plot to estimate the true ‘centre’ of the funnel, then (3) replace the omitted studies and their missing ‘counterparts’ around the centre (filling). As well as providing an estimate of the number of missing studies, an adjusted intervention effect is derived by performing a meta-analysis including the filled studies.

The trim and fill method requires no assumptions about the mechanism leading to publication bias, provides an estimate of the number of missing studies, and also provides an estimated intervention effect ‘adjusted’ for the publication bias (based on the filled studies). However, it is built on the strong assumption that there should be a symmetric funnel plot, and there is no guarantee that the adjusted intervention effect matches what would have been observed in the absence of publication bias, since we cannot know the true mechanism for publication bias. Equally importantly, the trim and fill method does not take into account reasons for funnel plot asymmetry other than publication bias. Therefore,

'corrected' intervention effect estimates from this method should be interpreted with great caution. The method is known to perform poorly in the presence of substantial between-study heterogeneity (Terrin 2003, Peters 2007). Additionally, estimation and inferences are based on a dataset containing imputed intervention effect estimates. Such estimates, it can be argued, inappropriately contribute information that reduces the uncertainty in the summary intervention effect.

10.4.4.3 Fail-safe N

Rosenthal suggested assessing the potential for publication bias to have influenced the results of a meta-analysis by calculating the 'fail-safe N', the number of additional 'negative' studies (studies in which the intervention effect was zero) that would be needed to increase the P value for the meta-analysis to above 0.05 (Rosenthal 1979). However the estimate of fail-safe N is highly dependent on the mean intervention effect that is assumed for the unpublished studies (Iyengar 1988), and available methods lead to widely varying estimates of the number of additional studies (Becker 2005). The method also runs against the principle that in medical research in general, and systematic reviews in particular, one should concentrate on the size of the estimated intervention effect and the associated confidence intervals, rather than on whether the P value reaches a particular, arbitrary threshold, although related methods for effect sizes have also been proposed (Orwin 1983). Therefore this and related methods are not recommended for use in Cochrane reviews.

10.4.4.4 Other selection models

Other authors have proposed more sophisticated methods that avoid strong assumptions about the association between study P value and publication probability (Dear 1992, Hedges 1992). These methods can be extended to estimate intervention effects, corrected for the estimated publication bias (Vevea 1995). However they require a large number of studies so that a sufficient range of study P values is included. A Bayesian approach in which the number and outcomes of unobserved studies are simulated has also been proposed as a means of correcting intervention effect estimates for publication bias (Givens 1997). Recent work has examined the possibility of assessing robustness over a range of weight functions, thus avoiding the need for large numbers of studies (Vevea 2005). The complexity of the statistical methods, and the large number of studies needed, probably explain why selection models have not been widely used in practice.

10.4.4.5 Sensitivity analyses based on selection models

Copas developed a model in which the probability that a study is included in a meta-analysis depends on its standard error. Because it is not possible to estimate all model parameters precisely, he advocates sensitivity analyses in which the value of the estimated intervention effect is computed under a range of assumptions about the severity of the selection bias (Copas 1999). Rather than a single intervention effect estimated 'corrected' for publication bias, the reader can see how the estimated effect (and confidence interval) varies as the assumed amount of selection bias increases. Application of the method to epidemiological studies of environmental tobacco smoke and lung cancer suggests that publication bias may explain some of the association observed in meta-analyses of these studies (Copas 2000).

10.4.4.6 Testing for excess of studies with significant results

Ioannidis and Trikalinos propose a simple test that aims to evaluate whether there is an excess of studies that have formally statistically significant results (Ioannidis 2007a). The test compares the number of studies that have formally statistically significant results with the number of statistically significant results expected under different assumptions about the magnitude of the effect size. The simplest assumption is that the effect size is equal to the observed summary effect in the meta-analysis (but this may introduce an element of circularity). Other values for the underlying effect size, and different thresholds of significance, may be used. Hence, like the contour funnel plots described in Section 10.4.1, but unlike the regression tests, this method considers the distribution of the

significance of study results. However, unlike either the regression tests or contour funnel plots, the test does not make any assumption about small-study effects. An excess of significant results can reflect either suppression of whole studies or related selective/manipulative analysis and reporting practices that would cause similar excess.

The test has limited power, as do most other tests, when there are few studies and when there are few studies with significant results. Because the test has not been rigorously evaluated through simulation in comparison with alternative tests and under different scenarios, we currently do not recommend the test as an alternative to those described in Section 10.4.3.

A novel feature of the test is that it can be applied across a large number of meta-analyses on the same research field to examine the extent of publication and selective reporting biases across a whole domain of clinical research. Again, further evaluation of this approach would be welcome.

10.4.5 Summary

Although there is clear evidence that publication and other reporting biases lead to over-optimistic estimates of intervention effects, overcoming, detecting and correcting for publication bias is problematic. Comprehensive searches are important, particularly to identify studies as well defined as randomized trials. However, comprehensive searching is not sufficient to prevent some substantial potential biases.

Publication bias should be seen as one of a number of possible causes of ‘small-study effects’ – a tendency for estimates of the intervention effect to be more beneficial in smaller studies. Funnel plots allow review authors to make a visual assessment of whether small-study effects may be present in a meta-analysis. For continuous (numerical) outcomes with intervention effects measured as mean differences, funnel plots and statistical tests for funnel plot asymmetry are valid. However for dichotomous outcomes with intervention effects expressed as odds ratios, the standard error of the log odds ratio is mathematically linked to the size of the odds ratio, even in the absence of small-study effects. This can cause funnel plots plotted using log odds ratios (or odds ratios on a log scale) to appear asymmetric and can mean that P values from the test of Egger et al. are too small. For other effect measures, firm guidance is not yet offered. Three statistical tests for small-study effects are recommended for use in Cochrane reviews, provided that there are at least 10 studies. However, none is implemented in RevMan and statistical support is usually required. Only one test has been shown to work when the between-study heterogeneity variance exceeds 0.1. Results from tests for funnel plot asymmetry should be interpreted cautiously. When there is evidence of small-study effects, publication bias should be considered as only one of a number of possible explanations. In these circumstances, review authors should attempt to understand the source of the small-study effects, and consider their implications in sensitivity analyses.

10.5 Chapter information

Editors: Jonathan AC Sterne, Matthias Egger and David Moher on behalf of the Cochrane Bias Methods Group.

This chapter should be cited as: Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

Contributing authors: James Carpenter, Matthias Egger, Roger Harbord, Julian Higgins, David Jones, David Moher, Jonathan Sterne, Alex Sutton and Jennifer Tetzlaff.

Acknowledgements: We thank Doug Altman, Jon Deeks, John Ioannidis, Jaime Peters and Gerta Rücker for helpful comments.

Declarations of interest: James Carpenter, Jon Deeks, Matthias Egger, Roger Harbord, David Jones, Jaime Peters, Gerta Rücker, Jonathan Sterne and Alex Sutton are all authors on papers proposing tests for funnel plot asymmetry.

Box 10.5.a: The Cochrane Bias Methods Group

The Bias Methods Group (BMG), previously the Reporting Bias Methods Group, was formally registered as a Methods Group in 2000. The BMG addresses a range of different forms of bias, such as publication bias, language bias, selective outcome reporting bias and biases arising from study design and conduct. A major initiative of the group, in collaboration with the Statistical Methods Group, was the development of the new guidance for assessing risk of bias of included studies in Cochrane reviews.

Activities of BMG members include:

- undertaking empirical research to examine whether, and in which circumstances, various biases may have a substantial impact on systematic reviews, including the preparation of Cochrane Methodology reviews;
- undertaking methodological research on how to identify and address potential biases in systematic reviews and meta-analyses;
- helping to complete and co-ordinate Methods systematic reviews pertinent to the Group's remit;
- providing advice to Cochrane entities; and
- offering training to both Cochrane and non-Cochrane systematic reviewers via formal and informal opportunities.

The BMG membership emailing list is used as a forum for discussion and dissemination of information. Cochrane newsletters and email distribution lists, such as the Cochrane Methods Group newsletter, *Cochrane News* and CCInfo, are also used for dissemination of group activities.

Funding: The BMG receives infrastructure funding as part of a commitment by the Canadian Institutes of Health Research (CIHR) and the Canadian Agency for Drugs and Technologies in Health (CADTH) to fund Canadian-based Cochrane entities. This supports dissemination activities, web hosting, travel, training, workshops and a full time Co-ordinator position.

Web site: www.chalmersresearch.com/bmg

10.6 References

Abbasi 2004

Abbasi K. Compulsory registration of clinical trials. *BMJ* 2004; 329: 637-638.

Abbot 1998

Abbot NC, Ernst E. Publication bias: direction of outcome is less important than scientific quality. *Perfusion* 1998; 11: 182-182.

Anonymous 1991

Anonymous. Subjectivity in data analysis. *The Lancet* 1991; 337: 401-402.

Bailey 2002

Bailey BJ. Duplicate publication in the field of otolaryngology-head and neck surgery. *Otolaryngology and Head and Neck Surgery* 2002; 126: 211-216.

Barden 2003

Barden J, Edwards JE, McQuay HJ, Moore RA. Oral valdecoxib and injected parecoxib for acute postoperative pain: a quantitative systematic review. *BMC Anesthesiology* 2003; 3: 1.

Bardy 1998

Bardy AH. Bias in reporting clinical trials. *British Journal of Clinical Pharmacology* 1998; 46: 147-150.

Becker 2005

Becker BJ. Failsafe *N* or file-drawer number. In: Rothstein HR, Sutton AJ, Borenstein M (editors). *Publication Bias in Meta-Analysis*. Chichester (UK): John Wiley & Sons, 2005.

Begg 1988

Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society Series A* 1988; 151: 419-463.

Begg 1994

Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088-1101.

Bhandari 2004

Bhandari M, Busse JW, Jackowski D, Montori VM, Schünemann H, Sprague S, Mears D, Schemitsch EH, Heels-Ansdell D, Devereaux PJ. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *Canadian Medical Association Journal* 2004; 170: 477-480.

Blumenthal 1997

Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS. Withholding research results in academic life science. Evidence from a national survey of faculty. *JAMA* 1997; 277: 1224-1228.

Brooks 1985

Brooks TA. Private acts and public objects: an investigation of citer motivations. *Journal of the American Society for Information Science* 1985; 36: 223-229.

Burdett 2003

Burdett S, Stewart LA, Tierney JF. Publication bias and meta-analyses: a practical example. *International Journal of Technology Assessment in Health Care* 2003; 19: 129-134.

Cantekin 1991

Cantekin EI, McGuire TW, Griffith TL. Antimicrobial therapy for otitis media with effusion ('secretory' otitis media). *JAMA* 1991; 266: 3309-3317.

Carter 2006

Carter AO, Griffin GH, Carter TP. A survey identified publication bias in the secondary literature. *Journal of Clinical Epidemiology* 2006; 59: 241-245.

Chan 2004a

Chan AW, 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: 2457-2465.

Chan 2004b

Chan AW, Krléža-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal* 2004; 171: 735-740.

Chan 2005

Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005; 330: 753.

CLASP Collaborative Group 1994

CLASP Collaborative Group. CLASP: a randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *The Lancet* 1994; 343: 619-629.

Cook 1993

Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McIlroy W, Oxman AD. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993; 269: 2749-2753.

Copas 1999

Copas J. What works?: selectivity models and meta-analysis. *Journal of the Royal Statistical Society Series A* 1999; 162: 95-109.

Copas 2000

Copas JB, Shi JQ. Reanalysis of epidemiological evidence on lung cancer and passive smoking. *BMJ* 2000; 320: 417-418.

Cowley 1993

Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcinide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *International Journal of Cardiology* 1993; 40: 161-166.

Davey Smith 1994

Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. *BMJ* 1994; 308: 72-74.

Dear 1992

Dear KBG, Begg CB. An approach to assessing publication bias prior to performing a meta-analysis. *Statistical Science* 1992; 7: 237-245.

Decullier 2005

Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ* 2005; 331: 19.

Decullier 2007

Decullier E, Chapuis F. Oral presentation bias: a retrospective cohort study. *Journal of Epidemiology and Community Health* 2007; 61: 190-193.

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005; 58: 882-893.

Dickersin 1992

Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 1992; 263: 374-378.

Dickersin 1993

Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials* 1993; Doc No 50.

Dickersin 1994

Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286-1291.

Dickersin 1997

Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Education and Prevention* 1997; 9: 15-21.

Dickersin 2002

Dickersin K, Olson CM, Rennie D, Cook D, Flanagin A, Zhu Q, Reiling J, Pace B. Association between time interval to publication and statistical significance. *JAMA* 2002; 287: 2829-2831.

Dong 1997

Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL, Greenspan FS. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism [see comments]. *JAMA* 1997; 277: 1205-1213.

Duval 2000

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.

Easterbrook 1991

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *The Lancet* 1991; 337: 867-872.

Egger 1997a

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-634.

Egger 1997b

Egger M, Zellweger Z, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *The Lancet* 1997; 350: 326-329.

Egger 2003

Egger M, Jüni P, Bartlett C, Hohenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment* 2003; 7: 1.

Epstein 1990

Epstein WM. Confirmational response bias among social work journals. *Science, Technology and Human Values* 1990; 15: 9-37.

Ernst 1994

Ernst E, Resch KL. Reviewer bias: A blinded experimental study. *Journal of Laboratory and Clinical Medicine* 1994; 124: 178-182.

Fergusson 2000

Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPO meta-analyses. *International Journal of Technology Assessment in Health Care* 2000; 16: 1109-1119.

Galanti 2006

Galanti D, Schwarzer G, Antes G. The demise of the randomised controlled trial: bibliometric study of the German-language health care literature, 1948 to 2004. *BMC Medical Research Methodology* 2006; 6: 30.

Givens 1997

Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis: a Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science* 1997; 12: 221-250.

Glasziou 1995

Glasziou PP, Iriwig LM. An evidence based approach to individualising treatment. *BMJ* 1995; 311: 1356-1359.

Godlee 1999

Godlee F, Dickersin K. Bias, subjectivity, chance, and conflict of interest in editorial decisions. In: Godlee F, Jefferson T (editors). *Peer Review in Health Sciences*. London (UK): BMJ Books, 1999.

Gøtzsche 1987

Gøtzsche PC. Reference bias in reports of drug trials. *British Medical Journal (Clinical Research Edition)* 1987; 295: 654-656.

Gøtzsche 1989

Gøtzsche PC. Multiple publication of reports of drug trials. *European Journal of Clinical Pharmacology* 1989; 36: 429-432.

Grégoire 1995

Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *Journal of Clinical Epidemiology* 1995; 48: 159-163.

Harbord 2006

Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006; 25: 3443-3457.

Hartling 2004

Hartling L, Craig WR, Russell K, Stevens K, Klassen TP. Factors influencing the publication of randomized controlled trials in child health research. *Archives of Pediatrics and Adolescent Medicine* 2004; 158: 983-987.

Hedges 1992

Hedges LV. Modeling publication selection effects in meta-analysis. *Statistical Science* 1992; 7: 246-255.

Hemminki 1980

Hemminki E. Study of information submitted by drug companies to licensing authorities. *British Medical Journal* 1980; 280: 833-836.

Heres 2006

Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry* 2006; 163: 185-194.

Hetherington 1989

Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989; 84: 374-380.

Hopewell 2004

Hopewell S. *Impact of grey literature on systematic reviews of randomized trials* (PhD thesis). University of Oxford, 2004.

Hopewell 2007a

Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000011.

Hopewell 2007b

Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000010.

Hopewell 2008

Hopewell S, Loudon K, Clarke M, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews* (to appear).

Huston 1996

Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *The Lancet* 1996; 347: 1024-1026.

Hutchison 1995

Hutchison BG, Oxman AD, Lloyd S. Comprehensiveness and bias in reporting clinical trials. *Canadian Family Physician* 1995; 41: 1356-1360.

Ioannidis 1998

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.

Ioannidis 2001

Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001; 285: 437-443.

Ioannidis 2007a

Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clinical Trials* 2007; 4: 245-253.

Ioannidis 2007b

Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal* 2007; 176: 1091-1096.

Irwig 1998

Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. *BMJ* 1998; 316: 470-471.

Iyengar 1988

Iyengar S, Greenhouse JB. Selection problems and the file drawer problem. *Statistical Science* 1988: 109-135.

Johansen 1999

Johansen HK, Gøtzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis [see comments]. *JAMA* 1999; 282: 1752-1759.

Jüni 2002

Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology* 2002; 31: 115-123.

Kjaergard 2002

Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. *Journal of Clinical Epidemiology* 2002; 55: 407-410.

Lexchin 2003

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326: 1167-1170.

Liebeskind 2006

Liebeskind DS, Kidwell CS, Sayre JW, Saver JL. Evidence of publication bias in reporting acute stroke clinical trials. *Neurology* 2006; 67: 973-979.

Light 1984

Light RJ, Pillemer DB. *Summing up. The science of reviewing research* (1). Cambridge (MA): Harvard University Press, 1984.

Macaskill 2001

Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 2001; 20: 641-654.

Mahoney 1977

Mahoney MJ. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1977; 1: 161-175.

Mandel 1987

Mandel EH, Rockette HE, Bluestone CD, Paradise JL, Nozza RJ. Efficacy of amoxicillin with and

without decongestant-antihistamine for otitis media with effusion in children. *New England Journal of Medicine* 1987; 316: 432-437.

McAuley 2000

McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet* 2000; 356: 1228-1231.

Melander 2003

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: 1171-1173.

Moher 1996

Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *The Lancet* 1996; 347: 363-366.

Moher 2000

Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology* 2000; 53: 964-972.

Moher 2003

Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment* 2003; 7: 1-90.

Moher 2007

Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine* 2007; 4: e78.

Moscatti 1994

Moscatti R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Academic Emergency Medicine* 1994; 1: 267-271.

Olson 2002

Olson CM, Rennie D, Cook D, Dickersin K, Flanagan A, Hogan JW, Zhu Q, Reiling J, Pace B. Publication bias in editorial decision making. *JAMA* 2002; 287: 2825-2828.

Orwin 1983

Orwin RG. A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics* 1983; 8: 157-159.

Peters 1982

Peters DP, Ceci SJ. Peer review practices of psychology journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences* 1982; 5: 187-255.

Peters 2006

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; 295: 676-680.

Peters 2007

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine* 2007; 26: 4544-4562.

Peters 2008

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. The contour enhanced funnel plot: an aid to interpreting funnel asymmetry. *Journal of Clinical Epidemiology* 2008; 61: 991-996.

Pham 2005

Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic

reviews gave different results depending on whether the intervention was conventional or complementary. *Journal of Clinical Epidemiology* 2005; 58: 769-776.

Pittler 2000

Pittler MH, Abbot NC, Harkness EF, Ernst E. Location bias in controlled clinical trials of complementary/alternative therapies. *Journal of Clinical Epidemiology* 2000; 53: 485-489.

Pocock 1987

Pocock S, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *New England Journal of Medicine* 1987; 317: 426-432.

Poole 1999

Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999; 150: 469-475.

Ravnskov 1992

Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992; 305: 15-19.

Rennie 1991

Rennie D. The Cantekin affair. *JAMA* 1991; 266: 3333-3337.

Rennie 1997

Rennie D. Thyroid Storms. *JAMA* 1997; 277: 1238-1243.

Rosenthal 1979

Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychological Bulletin* 1979; 86: 638-641.

Rücker 2008

Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* 2008; 27: 746-763.

Sampson 2003

Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St John PD, Viola R, Raina P. Should meta-analysts search Embase in addition to Medline? *Journal of Clinical Epidemiology* 2003; 56: 943-955.

Scherer 2007

Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000005.

Schulz 1995

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.

Schwarzer 2007

Schwarzer G, Antes G, Schumacher M. A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine* 2007; 26: 721-733.

Simes 1987

Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Statistics in Medicine* 1987; 6: 11-29.

Smith 1999

Smith R. What is publication? A continuum. *BMJ* 1999; 318: 142.

Sterling 1959

Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association* 1959; 54: 30-34.

Sterling 1995

Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician* 1995; 49: 108-112.

Stern 1997

Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; 315: 640-645.

Sterne 2000

Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 2000; 53: 1119-1129.

Sterne 2001

Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 2001; 54: 1046-1055.

Stuck 1998

Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity. Letter. *BMJ* 1998; 316: 469-471.

Tang 2000

Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology* 2000; 53: 477-484.

Tannock 1996

Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *Journal of the National Cancer Institute* 1996; 88: 206-207.

Taylor 1998

Taylor SJ, Tweedie RL. Practical estimates of the effect of publication bias in meta-analysis. *Australian Epidemiologist* 1998; 5: 14-17.

Teo 1993

Teo KK, Yusuf S, Furberg CD. Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction. An overview of results from randomized controlled trials [see comments]. *JAMA* 1993; 270: 1589-1595.

Terrin 2003

Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine* 2003; 22: 2113-2126.

Terrin 2005

Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology* 2005; 58: 894-901.

Tetzlaff 2006

Tetzlaff J, Moher D, Pham B, Altman D. Survey of views on including grey literature in systematic reviews. *14th Cochrane Colloquium*, Dublin (Ireland), 2006.

Tramèr 1997

Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; 315: 635-640.

Vevea 1995

Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; 60: 419-435.

Vevea 2005

Vevea JL, Woods CM. Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods* 2005; 10: 428-443.

Vickers 1998

Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials* 1998; 19: 159-166.

Villar 1997

Villar J, Piaggio G, Carroli G, Donner A. Factors affecting the comparability of meta-analyses and largest trials results in perinatology. *Journal of Clinical Epidemiology* 1997; 50: 997-1002.

Weber 1998

Weber EJ, Callahan ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA* 1998; 280: 257-259.

Zarin 2005

Zarin DA, Tse T, Ide NC. Trial Registration at ClinicalTrials.gov between May and October 2005. *New England Journal of Medicine* 2005; 353: 2779-2787.