**Abstract**

**Background:** The validity of research synthesis is threatened if published studies comprise a biased selection of all studies that have been conducted. We conducted a meta-analysis to ascertain the strength and consistency of the association between study results and formal publication.

**Methods:** The Cochrane Methodology Register Database, MEDLINE and other electronic bibliographic databases were searched (to May 2009) to identify empirical studies that tracked a cohort of studies and reported the odds of formal publication by study results. Reference lists of retrieved articles were also examined for relevant studies. Odds ratios were used to measure the association between formal publication and significant or positive results. Included studies were separated into subgroups according to starting time of follow-up, and results from individual cohort studies within the subgroups were quantitatively pooled.

**Results:** We identified 12 cohort studies that followed up research from inception, four that included trials submitted to a regulatory authority, 28 that assessed the fate of studies presented as conference abstracts, and four cohort studies that followed manuscripts submitted to journals. The pooled odds ratio of publication of studies with positive results, compared to those without positive results (publication bias) was 2.78 (95% CI: 2.10 to 3.69) in cohorts that followed from inception, 5.00 (95% CI: 2.01 to 12.45) in trials submitted to regulatory authority, 1.70 (95% CI: 1.44 to 2.02) in abstract cohorts, and 1.06 (95% CI: 0.80 to 1.39) in cohorts of manuscripts.

**Conclusion:** Dissemination of research findings is likely to be a biased process. Publication bias appears to occur early, mainly before the presentation of findings at conferences or submission of manuscripts to journals.
Background

Synthesis of published research is increasingly important in providing relevant and valid research evidence to inform clinical and health policy decision making. However, the validity of research synthesis based on published literature is threatened if published studies comprise a biased selection of the whole set of all conducted studies [1].

The observation that many studies are never published was termed “the file-drawer problem” by Rosenthal in 1979 [2]. The importance of this problem depends on whether or not the published studies are representative of all studies that have been conducted. If the published studies are a random sample of all studies that have been conducted, there will be no bias and the average estimate based on the published studies will be similar to that based on all studies. If the published studies comprise a biased sample of all studies that have been conducted, the results of a literature review will be misleading [3]. For example, the efficacy of a treatment will be exaggerated if studies with positive results are more likely to be published than those with negative results.

Publication bias is defined as “the tendency on the parts of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings” [4]. The existence of publication bias was first suspected by Sterling in 1959, after observing that 97% of studies published in four major psychology journals provided statistically significant results [5]. In 1995, the same author concluded that the practices leading to publication bias had not changed over a period of 30 years [6].

Evidence of publication bias can be classified as direct or indirect [7]. Direct evidence includes the acknowledgement of bias by those involved in the publication process (investigators, referees or editors), comparison of the results of published and unpublished studies, and the follow-up of cohorts of registered studies [8]. Indirect evidence includes the observation of disproportionately high percentage of positive findings in the published literature, and a larger effect size in small studies as compared with large studies. This evidence is indirect because factors other than publication bias may also lead to the observed disparities.

In a Health Technology Assessment (HTA) report published in 2000, we presented a comprehensive review of studies that provided empirical evidence of publication and related biases [8]. The review found that studies with significant or favourable results were more likely to be published, or were likely to be published earlier, than those with non-significant or unimportant results. There was limited and indirect evidence indicating the possibility of full publication bias, outcome reporting bias, duplicate publication bias, and language bias. Considering that the spectrum of the accessibility of research results (dissemination profile) ranges from completely inaccessible to easily accessible, it was suggested that a single term ‘dissemination bias’ could be used to denote all types of publication and related biases [8].

Since then, many new empirical studies on publication and related biases have been published. For example, Egger et al provided further empirical evidence on publication bias, language bias, grey literature bias, and MEDLINE index bias [9]. and Moher et al evaluated language bias in meta-analyses of randomised controlled trials [10]. Recently, more convincing evidence on outcome reporting bias has been published [11-13]. In addition, a large number of studies of publication bias in conference abstracts have been published [14]. As this new empirical evidence may strengthen or contradict the empirical evidence included in the previous review, we have updated our review of dissemination bias, by synthesizing findings from newly and previously identified studies [15]. This paper focuses on findings from a review of cohort studies that provided direct evidence on dissemination bias by investigating the association between publication and study results.

Methods

Criteria for inclusion

This review included any studies that tracked a cohort of studies and reported the rate of publication by study results. The relevant empirical studies may have tracked a cohort of protocols approved by research ethics committees, registered by research funding bodies, submitted to regulatory authorities, presented at conferences, or submitted to journals. Primary studies included in such cohorts could be clinical trials, observational or basic research. We separated the cohorts into four subgroups according to stages in a simplified pathway of research publication from inception to the journal publication (Figure 1). A study that followed up a cohort of research from the beginning (even if retrospectively) was termed an inception cohort study. A regulatory cohort study refers to a study that examined formal publication of clinical trials submitted to regulatory authorities (such as the US Food and Drug Administration or similar). An abstract cohort study investigated the subsequent full publication of abstracts presented at conferences. A manuscript cohort study followed the publication fate of manuscripts submitted to journals. Studies that did not provide sufficient data to compare the rates of publication between studies with different results were excluded.
Literature Search strategy
The identification of cohort studies for this review was conducted as part of a comprehensive search for empirical and methodological studies on research dissemination bias. We searched MEDLINE, the Cochrane Methodology Register Database (CMRD), EMBASE, AMED and CINAHL to August 2008 (see Additional file 1). PubMed, PsycINFO and OpenSIGLE were searched in May 2009 to locate more recently published studies. References (titles with or without abstracts) identified by the MEDLINE and CMRD searches were examined independently by two reviewers, while those from other databases were assessed by one reviewer. The reference lists of retrieved studies and reviews were examined to identify additional studies.

Data Extraction and Synthesis
Data extraction included type of cohort, clinical speciality, design of studies, duration of follow-up, definition of study results, and the rate of publication by study results. One reviewer extracted data directly into tables, which were checked by a second reviewer.

The outcome, study publication, was usually defined as full publication in a journal, but study results were often categorised differently between included cohorts. We used the classifications 'statistically significant (p ≤ 0.05)' versus 'non-significant (p > 0.05)' or 'positive' versus 'non-positive'. Positive results included results that were considered to be 'significant', 'positive', 'favourable', 'important', 'striking', 'showed effect', or 'confirmatory', while non-positive results were labelled as being 'negative', 'non-significant', 'less or not important', 'invalidating', 'inconclusive', 'questionable', 'null', or 'neutral'.

The validity of the included cohort studies was not formally assessed in this review, due to a lack of reliable tools for assessment of methodological studies. However, we tried to identify and summarise the main methodological limitations in the included studies.

Data from the included studies were first summarised in narrative form. The odds ratio was used as the outcome statistic to measure the association between study results and subsequent publication. In existing reviews of cohort studies of publication bias, results from different studies have been quantitatively combined [14,16,17], although significant heterogeneity across individual studies has lead to some controversy [18]. We felt it helpful to provide pooled estimates after separating the included cohort studies into appropriate subgroups, in order to improve statistical power and generalisability. Results from individual studies were quantitatively pooled using random-effects meta-analysis. Heterogeneity within each subgroup was measured using the $I^2$ statistic, considering heterogeneity to be moderate or high when $I^2$ is greater than 50% [19]. Rucker et al pointed out that, given the same between-study variance, the value of $I^2$ will increase rapidly as the sample size of individual studies increases in meta-analysis [20]. Therefore, clinical and methodological relevance were the most important issue to consider when deciding whether the results from individual studies could be quantitatively combined in meta-analysis.

Funnel plots were used to assess the association between the point estimates of log odds ratio (a measure of extent of publication bias) and the precision of estimated log odds ratio (inverse of standard errors). The visual assessment of these plots was supported by a formal statistical test using the regression method suggested by Peters et al [21].

Results
Forty-eight cohort studies provided sufficient data to assess rate of publication by study results. These consisted of 12 inception cohort studies [22-33], four regulatory cohort studies [34-37], 28 abstract cohort studies [38-65], and four manuscript cohort studies [66-69]. Eight inception cohort studies were excluded because they did not provide data on the results of unpublished studies or did not examine the association between publication and study results [12,13,70-75].
Main characteristics of included cohort studies

The main characteristics of the included cohort studies are summarised by subgroup in four additional files (Additional file 2, 3, 4 and 5). One inception cohort study [32] and four abstract cohort studies [44,49,63,65] were available only in abstract form. Of the 12 cohort studies in the inception subgroup, eight did not restrict the field of research, and four were limited to AIDS/HIV [29], health effects of passive smoking [30], complementary medicine [33], or eye diseases [32]. Of the four cohort studies in the regulatory subgroup, two did not specify clinical fields [34,36] and two included studies of anti-depressants [35,37]. All of the cohort studies of conference abstracts were restricted to a specific clinical field, such as emergency medicine, anaesthesiology, perinatal studies, cystic fibrosis, or oncology. Of the four cohorts of journal manuscripts, two examined manuscripts submitted to general medical journals (JAMA, BMJ, Lancet, and Annals of Internal Medicine) [66,69] while two included manuscripts submitted to the American version of the Journal of Bone and Joint Surgery [67,68].

Four of the 12 cohort studies in the inception group, all four studies in the regulatory group, and nine of the 28 cohort studies in the abstract group included only clinical trials. The remaining cohort studies included research of different designs, although separate data for clinical trials was available in some of these studies. Follow-up time ranged from 1 to 12 years in the inception cohort studies, and from 2 to 5 years in the regulatory or abstract cohort studies.

Authors of inception cohort studies used postal questionnaires or telephone interviews of investigators or both to obtain information on results of unpublished studies. The response rate to the survey of investigators ranged from 69% to 92% (see Additional file 2). Information on study results was already available in five regulatory cohort studies and in all abstract cohort studies. Bibliographic databases were usually searched to decide publication status.

Study results were categorised as statistically significant (p < 0.05) or non-significant in 19 studies, and a wide range of different methods were used to categorise study results as, for example, positive versus negative, confirmatory versus inconclusive, striking versus unimportant.

Pooled estimates of publication bias

Table 1 summarises the main results of the meta-analyses. The formal publication of statistically significant results (p < 0.05) could be compared with that of non-significant results in four inception cohort studies, one regulatory cohort study, 12 abstract cohort studies and two manuscript cohort studies (Figure 2). The rate of publication of studies in the four inception cohorts ranged from 60% to 93% for significant results and from 20% to 86% for non-significant results. The rate of full publication of meeting abstracts ranged from 37% to 81% for statistically significant results, and from 22% to 70% for non-significant results. Heterogeneity across the four cohort studies from the inception subgroup was statistically significant ($I^2 = 61\%$, p = 0.05). There was no statistically significant heterogeneity across studies within the cohort studies of abstracts and cohort studies of manuscripts. The pooled odds ratio for publication bias by statistical significance of results was 2.40 (95% CI: 1.18 to 4.88) for the four inception cohort studies, 1.62 (95% CI: 1.34 to 1.96) for the 12 abstract cohort studies, and 1.15 (95% CI: 0.64 to 2.10) for the two manuscript cohort studies (Figure 2). Only one regulatory cohort study was included in Figure 2, reported an odds ratio of 11.06 (95% CI: 0.56 to 21.97) for publication of statistically significant vs. non-significant results.

To include data from other cohort studies, a positive result was loosely defined as important, confirmatory or significant, while a ‘non-positive’ result included negative, non-important, inconclusive or non-significant results. This more inclusive definition of positive results allowed the inclusion of all 12 inception cohort studies, four regulatory cohort studies, 27 abstract cohort studies, and four manuscript cohort studies (Figure 3). There was statistically significant heterogeneity across cohort studies within regulatory (p = 0.04) and abstract subgroups (p < 0.001). Pooled estimates of odds ratios consistently indicated that studies with positive results were more likely to be published than studies with non-positive results, but this was not true after submission for publication (Figure 3).

Types of studies included in the cohort studies varied from basic experimental, observational and qualitative research to clinical trials. When the analyses were restricted to clinical trials, the results were similar to that based on all studies, and there was no significant heterogeneity in the extent of publication bias among the included inception cohorts and abstract cohorts of clinical trials (Figure 4 and Figure 5).

Funnel plots constructed separately for the four subgroups of cohort studies were not statistically significantly asymmetric (Figure 6).

Factors associated with publication bias

Some cohort studies have examined the impacts of other factors on biased publication of research. The factors investigated include study design, type of study, sample size, funding source, and investigators’ characteristics. However, only a few of the included cohort studies...
reported findings regarding factors associated with publication bias and findings from different studies were often inconsistent.

Easterbrook et al conducted subgroup analyses to examine susceptibility to publication bias amongst various subgroups of studies. They found that observational, laboratory-based experimental studies and non-randomised trials had a greater risk of publication bias than randomised clinical trials. Factors associated with less bias included a concurrent comparison group, a high investigator rating of study importance and a sample size greater than 20 [28].

Dickersin and colleagues investigated the association between the risk of publication bias and type of study (observational, clinical trial), multi or single centre, sample size, funding source and principle investigators’ characteristics (such as sex, degree, rank). They found no statistically significant association between any factors examined and publication bias [26]. In a different inception cohort study, Dickersin and Min reported that the odds ratio for publication bias was significantly different between multi-centre versus single centre studies, and between studies with a female principle investigator and studies with a male principle investigator [27]. They did not find an association between publication bias and the use of randomisation or blinding, having a comparison group or a larger sample size [27].

Stern and Simes found that the risk of publication bias tended to be greater for clinical trials (odds ratio 3.13, 95% CI: 1.76 to 5.58) than other studies (for all quantitative studies odds ratio 2.32, 95% CI: 1.47 to 3.66). When analyses were restricted to studies with a sample size ≥100, publication bias was still evident (hazard ratio 2.00, 95% CI: 1.09 to 3.66) [31].

### Discussion

This updated analysis yielded results similar to previous reviews: studies with statistically significant or positive results are more likely to be formally published than those with non-significant or non-positive results [14,16-18]. In 1997, Dickersin combined the results from four inception cohort studies [26-28,31] and found that the pooled adjusted odds ratio for publication bias (publication of studies with significant or important results versus those with unimportant results) was 2.54 (95% CI: 1.44 to 4.47) [16]. A recent systematic review of inception cohort studies of clinical trials found the existence of publication bias and outcome reporting bias, although pooled meta-analysis was not conducted due to perceived differences between studies [18]. A Cochrane methodology review of publication bias by Hopewell et al[17] included five inception cohort studies of trials registered before the main results being known [22,26,27,29,31], in which the pooled odds ratio for publication bias was 3.90 (95% CI: 2.68 to 5.68). In a Cochrane methodology review by Scherer et al, the association between the subsequent full publication and study results was examined in 16 of 79 abstract cohort studies [14], finding that the subsequent full publication of conference abstracts was associated with positive results (pooled OR = 1.28, 95% CI: 1.15 to 1.42).

Our review is the first to enable an explicit comparison of results from cohort studies of publication bias with fundamentally different sampling frames. Biased selection for publication may affect research dissemination over the whole process from before study completion, to presentation of findings at conferences, manuscript submission to journals, and formal publication in journals (Figure 1). It seems that publication bias occurs mainly before the presentation of findings at conferences and before the submission of manuscript to journals (Figure 2 and Figure 3).

### Table 1: Results of meta-analyses of cohort studies of publication bias

| Cohort category          | No. of cohort studies | Pooled odds ratio (95% CI) | Heterogeneity test: $I^2$ (p value) |
|--------------------------|-----------------------|---------------------------|-----------------------------------|
| **Statistically significant versus non-significant results** | | | |
| Inception cohorts        | 4                     | 2.40 (1.18 to 4.88)       | 61% (0.05)                        |
| Regulatory cohorts       | 1                     | 11.06 (0.56 to 219.68)    |                                   |
| Abstract cohorts         | 12                    | 1.62 (1.34 to 1.96)       | 40% (0.08)                        |
| Manuscript cohorts       | 2                     | 1.15 (0.64 to 2.10)       | 48% (0.17)                        |
| **Positive versus non-positive results** | | | |
| Inception cohorts        | 12                    | 2.78 (2.10 to 3.69)       | 37% (0.09)                        |
| Regulatory cohorts       | 4                     | 5.00 (2.01 to 12.45)      | 64% (0.04)                        |
| Abstract cohorts         | 27                    | 1.70 (1.44 to 2.02)       | 61% (<0.001)                      |
| Manuscript cohorts       | 4                     | 1.06 (0.80 to 1.39)       | 22% (0.28)                        |
Rate of publication of statistically significant versus non-significant results - all studies

| Study or Subgroup                  | Significant Events | Non-significant Events | Odds Ratio M-H, Random, 95% CI | Odds Ratio M-H, Random, 95% CI |
|------------------------------------|--------------------|------------------------|--------------------------------|--------------------------------|
| **1.1 Inception cohorts**          |                    |                        |                               |                                |
| Easterbrook 1991                   | 93                 | 154                    | 39.2%                          | 2.91 [1.80, 4.73]              |
| Misakian 1998                      | 29                 | 40                     | 16.6%                          | 0.44 [0.11, 1.79]              |
| Stern 1997a                        | 99                 | 146                    | 36.0%                          | 3.51 [1.95, 6.33]              |
| Wormald 1997                       | 14                 | 15                     | 8.3%                           | 5.60 [0.60, 52.54]             |
| Subtotal (95% CI)                  | 355                | 245                    | 100.0%                         | 2.40 [1.18, 4.88]              |
| Total events                       | 235                | 105                    |                                |                                |
| Heterogeneity: Tau² = 0.27; Chi² = 7.65, df = 3 (P = 0.05); I² = 61% |          |            |                                |                                |
| Test for overall effect: Z = 2.42 (P = 0.02) |            |            |                                |                                |
| **1.1.2 Regulatory cohorts**       |                    |                        |                               |                                |
| Melander 2003                      | 21                 | 21                     | 100.0%                         | 11.06 [0.56, 219.68]           |
| Subtotal (95% CI)                  | 21                 | 21                     | 100.0%                         | 11.06 [0.56, 219.68]           |
| Total events                       | 21                 | 17                     |                                |                                |
| Heterogeneity: Not applicable       |                    |                        |                               |                                |
| Test for overall effect: Z = 1.58 (P = 0.12) |            |            |                                |                                |
| **1.1.4 Abstract cohorts**         |                    |                        |                               |                                |
| Akbari 2008                         | 23                 | 45                     | 4.3%                           | 1.05 [0.46, 2.40]              |
| Castillo 2002                      | 160                | 361                    | 7.5%                           | 1.14 [0.64, 2.02]              |
| Eloubeidi 2001                     | 36                 | 98                     | 38.0%                          | 2.08 [1.29, 3.37]              |
| Evers 2000                         | 41                 | 69                     | 38.9%                          | 1.00 [0.89, 1.24]              |
| Glick 2006                         | 208                | 397                    | 13.7%                          | 1.61 [1.16, 2.23]              |
| Harris 2006                        | 12                 | 24                     | 76.0%                          | 2.59 [1.09, 6.16]              |
| Harris 2007                        | 69                 | 101                    | 177.0%                         | 2.26 [1.37, 3.71]              |
| Krzyzanowska 2003                  | 181                | 223                    | 287.0%                         | 2.03 [1.34, 3.09]              |
| Scherer 1994                       | 33                 | 46                     | 47.0%                          | 1.72 [0.72, 4.10]              |
| Smith 2007                         | 521                | 1120                   | 202.0%                         | 1.17 [0.87, 1.59]              |
| Timmer 2002                        | 177                | 354                    | 174.0%                         | 1.13 [0.77, 1.66]              |
| Zamakhashary 2006                  | 105                | 151                    | 13.0%                          | 3.34 [1.52, 7.32]              |
| Subtotal (95% CI)                  | 2989               | 1877                   | 100.0%                         | 1.62 [1.34, 1.96]              |
| Total events                       | 1566               | 801                    |                                |                                |
| Heterogeneity: Tau² = 0.04; Chi² = 18.24, df = 11 (P = 0.08); I² = 40% |            |            |                                |                                |
| Test for overall effect: Z = 4.97 (P < 0.00001) |            |            |                                |                                |
| **1.1.5 Manuscripts to journals**  |                    |                        |                               |                                |
| Lee 2006                           | 35                 | 718                    | 32.7%                          | 0.75 [0.32, 1.73]              |
| Olson 2002                         | 78                 | 383                    | 55.3%                          | 1.43 [0.98, 2.09]              |
| Subtotal (95% CI)                  | 1101               | 471                    | 100.0%                         | 1.15 [0.64, 2.10]              |
| Total events                       | 113                | 62                     |                                |                                |
| Heterogeneity: Tau² = 0.10; Chi² = 1.91, df = 1 (P = 0.17); I² = 48% |          |            |                                |                                |
| Test for overall effect: Z = 0.47 (P = 0.64) |            |            |                                |                                |

Figure 2
Rate of publication of statistically significant versus non-significant results - all studies.
| Study or Subgroup | Positive Events | Non-positive Events | Odds Ratio M-H, Random, 95% CI | Odds Ratio M-H, Random, 95% CI |
|------------------|----------------|--------------------|--------------------------------|--------------------------------|
| **2.1.1 Inception cohorts** | | | | |
| Bardy 1998       | 52             | 111                | 3.36 [1.73, 6.53]              |                                |
| Cronin 2004      | 26             | 54                 | 1.84 [0.65, 5.22]              |                                |
| Decullier 2005   | 129            | 188                | 5.03 [2.91, 10.49]             |                                |
| Decullier 2006   | 26             | 37                 | 1.58 [0.37, 6.71]              |                                |
| Dickens 1992     | 259            | 314                | 2.48 [1.64, 3.74]              |                                |
| Dickersin 1993   | 121            | 124                | 7.04 [1.90, 26.16]             |                                |
| Easterbrook 1991 | 93             | 154                | 2.91 [1.80, 4.73]              |                                |
| Ioannidis 1998   | 20             | 27                 | 4.11 [1.41, 11.99]             |                                |
| Misakian 1998    | 29             | 40                 | 0.44 [0.11, 1.79]              |                                |
| Stern 1997       | 153            | 232                | 2.85 [1.72, 4.71]              |                                |
| Wormald 1997     | 14             | 15                 | 5.60 [1.60, 19.94]             |                                |
| Zimpel 2000      | 43             | 107                | 1.70 [0.84, 3.47]              |                                |
| **Subtotal (95% CI)** | 1383           | 411                | 2.78 [2.10, 3.89]              |                                |
| **Total events** | 965            | 401                |                                |                                |

Heterogeneity: Tau² = 0.08; Chi² = 17.53, df = 11 (P = 0.09); I² = 37%
Test for overall effect: Z = 7.10 (P < 0.00001)

| **2.1.3 Regulatory cohorts** | | | | |
| Lee 2008          | 285            | 432                | 3.43 [2.31, 5.09]              |                                |
| Melander 2003     | 21             | 21                 | 11.06 [5.56, 21.68]            |                                |
| Rising 2008       | 102            | 124                | 2.60 [1.17, 5.78]              |                                |
| Turner 2008       | 37             | 38                 | 58.14 [7.15, 473.03]           |                                |
| **Subtotal (95% CI)** | 615            | 240                | 5.00 [2.01, 12.45]             |                                |
| **Total events**  | 445            | 108                |                                |                                |

Heterogeneity: Tau² = 0.45; Chi² = 8.35, df = 3 (P = 0.04); I² = 64%
Test for overall effect: Z = 3.46 (P < 0.0005)

| **2.1.4 Abstract cohorts** | | | | |
| Alkari 2008        | 59             | 137                | 1.26 [0.29, 5.49]              |                                |
| Brazzelli 2009     | 107            | 141                | 1.15 [0.51, 2.12]              |                                |
| Callaham 1998      | 77             | 153                | 1.07 [0.61, 1.86]              |                                |
| Cassell 2002       | 160            | 361                | 1.14 [0.64, 2.02]              |                                |
| Cheng 1998         | 43             | 113                | 1.23 [0.58, 2.59]              |                                |
| De Bellafuite 1992 | 48             | 65                 | 2.74 [1.43, 5.25]              |                                |
| Delamere 2005      | 15             | 52                 | 15.00 [1.54, 146.54]           |                                |
| Eloubeidi 2001     | 36             | 98                 | 2.08 [1.29, 3.37]              |                                |
| Evers 2000         | 41             | 69                 | 1.70 [0.49, 6.24]              |                                |
| Glick 1995         | 208            | 397                | 0.97 [0.76, 1.24]              |                                |
| Ha 2008            | 288            | 982                | 3.26 [1.80, 5.98]              |                                |
| Halpern 2001       | 29             | 83                 | 2.27 [0.96, 5.33]              |                                |
| Harris 2006        | 45             | 132                | 1.55 [0.80, 2.99]              |                                |
| Harris 2007        | 123            | 203                | 1.80 [1.13, 2.86]              |                                |
| Hasselblad 1993    | 54             | 112                | 2.62 [1.55, 4.43]              |                                |
| Kronf 2001         | 98             | 139                | 2.61 [1.62, 4.22]              |                                |
| Klasson 2002       | 162            | 235                | 2.24 [1.51, 3.34]              |                                |
| Krzyzanowska 2003  | 148            | 229                | 1.81 [1.17, 2.80]              |                                |
| Landry 1996        | 24             | 58                 | 3.18 [1.56, 6.48]              |                                |
| Ping 2006          | 189            | 337                | 1.28 [0.57, 2.84]              |                                |
| Petticrew 1999     | 18             | 36                 | 0.95 [0.39, 2.33]              |                                |
| Sannosian 2006     | 136            | 220                | 0.98 [0.63, 1.52]              |                                |
| Scharl 1994        | 33             | 46                 | 1.72 [0.72, 4.10]              |                                |
| Smith 2007         | 521            | 1120               | 1.17 [0.87, 1.59]              |                                |
| Sinnar 2002        | 177            | 354                | 1.26 [0.96, 1.66]              |                                |
| Vecchi 2006        | 120            | 161                | 2.22 [1.48, 3.32]              |                                |
| Zamakkahy 2006     | 105            | 151                | 3.34 [1.52, 7.32]              |                                |
| **Subtotal (95% CI)** | 6108           | 4180               | 1.70 [1.44, 2.02]              |                                |
| **Total events**  | 3064           | 1915               |                                |                                |

Heterogeneity: Tau² = 0.10; Chi² = 66.82, df = 26 (P < 0.00001); I² = 61%
Test for overall effect: Z = 6.19 (P < 0.00001)

| **2.1.5 Manuscripts to journals** | | | | |
| Lee 2006           | 35             | 718                | 0.75 [0.32, 1.73]              |                                |
| Lynch 2007         | 45             | 148                | 0.97 [0.51, 1.84]              |                                |
| Okiwa 2008         | 132            | 620                | 0.91 [0.63, 1.30]              |                                |
| Olson 2002         | 78             | 383                | 1.43 [0.98, 2.09]              |                                |
| **Subtotal (95% CI)** | 1869           | 767                | 1.06 [0.80, 1.39]              |                                |
| **Total events**  | 290            | 135                |                                |                                |

Heterogeneity: Tau² = 0.02; Chi² = 3.83, df = 3 (P = 0.28); I² = 22%
Test for overall effect: Z = 0.39 (P = 0.69)

Figure 3
Rate of publication of positive versus non-positive results - all studies.
Figure 4
Rate of publication of statistically significant versus non-significant results - clinical trials only.
### Rate of publication of positive versus non-positive results - clinical trials only

**Figure 5**

Rate of publication of positive versus non-positive results - clinical trials only.

| Study or Subgroup | Positive Events | Non-positive Events | Odds Ratio M-H, Random, 95% CI | Odds Ratio M-H, Random, 95% CI |
|-------------------|----------------|--------------------|--------------------------------|--------------------------------|
| **4.1.1 Inception cohorts** | | | | |
| Bardy 1998 | 52 | 111 | 16 | 77 | 28.7% | 3.36 [1.73, 6.53] |
| Dickersin 1992 | 84 | 96 | 52 | 72 | 20.1% | 2.69 [1.22, 5.96] |
| Dickersin 1993 | 121 | 124 | 63 | 74 | 7.4% | 7.04 [1.90, 28.16] |
| Ioannidis 1998 | 20 | 27 | 16 | 39 | 11.0% | 4.11 [1.41, 11.99] |
| Stern 1997 | 69 | 100 | 27 | 67 | 30.3% | 3.30 [1.73, 6.29] |
| Wormald 1997 | 14 | 15 | 15 | 21 | 2.5% | 5.60 [0.60, 52.54] |
| Subtotal (95% CI) | 473 | 350 | 100.0% | 3.50 [2.45, 4.99] |
| **Total events** | 360 | 189 | | |
| Heterogeneity: Tau² = 0.00; Chi² = 1.81, df = 5 (P = 0.87); I² = 0% |
| Test for overall effect: Z = 6.89 (P < 0.00001) |

| **4.1.3 Regulatory cohorts** | | | | |
| Lee 2008 | 285 | 432 | 52 | 144 | 43.8% | 3.43 [2.31, 5.09] |
| Melander 2003 | 21 | 21 | 17 | 21 | 7.8% | 11.06 [0.56, 219.68] |
| Rising 2008 | 102 | 124 | 25 | 39 | 34.9% | 2.60 [1.17, 5.78] |
| Turner 2008 | 37 | 38 | 14 | 36 | 13.5% | 58.14 [7.15, 473.03] |
| Subtotal (95% CI) | 615 | 240 | 100.0% | 5.00 [2.01, 12.45] |
| **Total events** | 445 | 108 | | |
| Heterogeneity: Tau² = 0.45; Chi² = 8.35, df = 3 (P = 0.04); I² = 64% |
| Test for overall effect: Z = 3.46 (P = 0.0005) |

| **4.1.4 Abstract cohorts** | | | | |
| Akbari 2008 | 59 | 137 | 3 | 8 | 1.6% | 1.26 [0.29, 5.49] |
| Cheng 1998 | 43 | 113 | 14 | 42 | 6.2% | 1.23 [0.58, 2.59] |
| Delamere 2005 | 15 | 22 | 1 | 8 | 0.7% | 15.00 [1.54, 146.54] |
| Evers 2000 | 41 | 69 | 38 | 82 | 8.3% | 1.70 [0.89, 3.24] |
| Klassen 2002 | 162 | 235 | 93 | 187 | 21.8% | 2.24 [1.51, 3.34] |
| Krzyzanowska 2003 | 148 | 163 | 229 | 327 | 18.1% | 1.81 [1.17, 2.72] |
| Scherer 1994 | 33 | 46 | 28 | 47 | 4.6% | 1.72 [0.72, 4.10] |
| Timmer 2002 | 84 | 140 | 86 | 186 | 17.5% | 1.74 [1.12, 2.72] |
| Vecchi 2006 | 120 | 161 | 239 | 420 | 21.2% | 2.22 [1.48, 3.32] |
| Subtotal (95% CI) | 1106 | 1307 | 100.0% | 1.92 [1.60, 2.31] |
| **Total events** | 705 | 731 | | |
| Heterogeneity: Tau² = 0.00; Chi² = 6.35, df = 8 (P = 0.61); I² = 0% |
| Test for overall effect: Z = 6.88 (P < 0.00001) |

| **4.1.5 Manuscripts to journals** | | | | |
| Olson 2002 | 78 | 383 | 55 | 362 | 100.0% | 1.43 [0.98, 2.09] |
| Subtotal (95% CI) | 383 | 362 | 100.0% | 1.43 [0.98, 2.09] |
| **Total events** | 78 | 55 | | |
| Heterogeneity: Not applicable |
| Test for overall effect: Z = 1.84 (P = 0.07) |

Events

| Study or Subgroup | Positive | Non-positive |
|-------------------|----------|-------------|
| **4.1.1 Inception cohorts** | 52 | 111 |
| **4.1.3 Regulatory cohorts** | 285 | 432 |
| **4.1.4 Abstract cohorts** | 59 | 137 |
| **4.1.5 Manuscripts to journals** | 78 | 383 |
The subsequent publication of conference abstracts was still biased but the extent of publication bias tended to be smaller as compared to all studies conducted. After the submission of a manuscript for publication, editorial decisions were not clearly associated with study results. However, publication bias may still be an issue for rejected manuscripts, if the possibility of their re-submission to a different journal is associated with the study results. One excluded cohort study (in which data on publication bias was not available) found that psychological research with statistically significant results was more likely to be submitted for publication than research with non-significant results (74% versus 4%) [72].

Since the acceptance of manuscripts for publication by journal editors was not determined by the direction or strength of study results, the existence of publication bias may be largely due to biased selection by investigators of submitted studies. This is supported by data suggesting that a large proportion of submitted papers show significant or positive results (72%) in four cohort studies (Table 2). Since authors inevitably consider the possibility of their manuscripts being rejected before the submission, submitted studies with negative results may be a biased selection of all studies with negative results. In addition, although no conflict of interest was declared in the four cohort studies of submitted manuscripts, this kind of study will always need support or collaboration from journal editors. In prospective studies, editors’ decision on manuscript acceptance may be influenced by their awareness of the ongoing study [69]. Therefore, biased selection for publication by journals cannot be completely ruled out. In Olson et al’s cohort study of manuscripts submitted to JAMA, there was a tendency that studies with significant results had a higher rate of acceptance than studies with non-significant or unclear results (20.4% vs 15.2%, p = 0.07) [69]. In the cohort study by Okike et al, a subgroup analysis of 156 manuscripts with a high level of evidence (level I or II) found that the acceptance rate was significantly higher for studies with positive or neutral results than studies with negative results (37%, 36% and 5% respectively; p = 0.02) [68].

Figure 6
Funnel plots -- publication of positive and non-positive results. (Funnel plot asymmetry test: p = 0.178 for inception cohort studies, p = 0.262 for regulatory cohort studies, p = 0.142 for abstract cohort studies, and p = 0.942 for manuscript cohort studies.)
Important and convincing evidence on the existence of publication bias comes from the inception cohort studies. Study results were defined differently among the empirical studies assessing publication bias. The most objective method would be to classify quantitative results as statistically significant ($p < 0.05$) or not. However, this was not always possible or appropriate. When other methods were used to classify study results as important or not, bias may be introduced due to inevitable subjectivity.

Large cohort studies on publication bias were often highly diverse in terms of research questions, designs, and other study characteristics. Many factors (e.g., sample size, design, research question, and investigators' characteristics) may confound the association as they are associated with both study results and the possibility of publication. However, due to insufficient data, it was impossible to exclude the impact of confounding factors on the observed association between study results and formal publication. To improve our understanding of factors associated with publication bias findings from qualitative research on the process of research dissemination may be helpful [76,77].

There was a statistically significant heterogeneity within subgroups of inception, regulatory and abstract cohort studies (Table 1), although restricting analyses to clinical trials reduced the heterogeneity (Figure 4 and 5). The observed heterogeneity may be a result of differences in study designs, research questions, how the cohorts were assembled, definitions of study results, and so on. For example, the significant heterogeneity across inception cohort studies was due to one study by Misakian and Bero (Figure 2 and Figure 3) [30]. After excluding this cohort study, there was no longer significant heterogeneity across inception cohort studies (see Notes to Figure 2 and Figure 3). The cohort study by Misakian and Bero included research on health effects of passive smoking, and the impact of statistical significance of results on publication may be different from studies of other research topics [30].

The four cohorts of trials submitted to regulatory authorities showed a greater extent of publication bias than other subgroups of cohort studies (Figure 3) [34-37]. Only 855 primary studies were included in the regulatory cohort studies, and two of the four regulatory cohort studies focused on trials of antidepressants [35,37]. Therefore, these regulatory cohort studies may be a biased selection of all possible cases.

Studies of publication bias themselves may be as vulnerable as other studies to the selective publication of significant or striking findings [1,78-80]. In this review, the funnel plot asymmetry was non-significant for each of the four research cohorts (Figure 6). However, we identified a large number of reports of full publication of meeting abstracts, and the association between study results and full publication had not been reported in most of these reports. It is often unclear whether this association had not been examined, or was not reported because the association proved to be non-significant. As an example, Zaretsky and Imrie reported no significant difference ($p = 0.53$) in the rate of subsequent publication of 57 meeting abstracts between statistically significant and non-significant results; but this study could not be included in the analysis due to insufficient data [65].

**Implications**

Despite many caveats regarding the available empirical evidence on publication bias, there is little doubt that dissemination of research findings is likely to be a biased process. All funded or approved studies should be prospectively registered and these registers should be publicly accessible. Regulatory authorities should also provide publicly accessible databases of all studies received from pharmaceutical industry. Investigators should be encouraged and supported to present their studies at conferences. A thorough literature search should be conducted in systematic reviews to identify all relevant studies, including searches of registers of clinical trials and available databases of unpublished studies.

**Conclusion**

There is consistent empirical evidence that the publication of a study that exhibits statistically significant or 'important' results is more likely to occur than the publication of a study that does not show such results. Indirect evidence indicates that publication bias occurs mainly before the presentation of findings at conferences and the submission of manuscripts to journals.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

FS, YKL, LH, JR, AJS and IH developed the review protocol. SP, JR and FS conducted literature search. SP, FS, LH, YKL, JR and CH extracted and/or checked data from included

---

Table 2: Proportion of studies with significant or positive results in studies included in cohort studies of publication bias

| Cohort category | % | Total |
|-----------------|---|-------|
| Inception       | 63.0% | 2,194 |
| Regulatory      | 71.9% | 855   |
| Abstract        | 59.4% | 10,288 |
| Manuscript      | 70.9% | 2,636 |

---

(page number not for citation purposes)
Acknowledgements

This methodological review was funded by the UK Department of Health HTA Research Methodology Programme (code 06/92/02). The views expressed in this report are those of the authors and do not necessarily reflect those of the funders. The authors take the responsibility for any errors. We would like to thank Julie Reynolds for providing secretarial support to this review.

References

1. Begg CB, Berlin JA: Publication bias; a problem in interpreting medical data. J R Statist Soc A 1988, 151:445-463.
2. Rosenthal R: The "file drawer problem" and tolerance for null results. Psychol Bull 1979, 86(3):638-641.
3. Begg CB, Berlin JA: Publication bias and dissemination of clinical research. J Natl Cancer Inst 1989, 81:107-115.
4. Dickersin K: The existence of publication bias and risk factors for its occurrence. JAMA 1990, 263(10):1385-1389.
5. Sterling T: Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. Am Stat Assoc J 1959, 54:30-34.
6. 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 Statistican 1995, 49(1):108-112.
7. John D: Publications bias and the evaluation of psychotherapy efficacy In reviews of the research literature. Clin Psychol Rev 1996, 16(2):147-156.
8. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ: Publication and related biases. Health Technol Assess. 2000, 4(10):1-115.
9. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J: How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003, 7(1):1-76.
10. 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 Technol Assess 2003, 7(41):1-90.
11. Chan AW, Altman DG: Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ 2005, 330(7494):753.
12. Chan AW, Hrobjartsson A, Ashar MT, Gotzsche PC, Altman DG: Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004, 291(20):2457-2465.
13. Chan AW, Krizea-Jeri K, Schmid I, Altman DG: Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research [see comment]. CMAJ 2004, 171(7):735-740.
14. Scherer KW, Langenberg P, von Elm E: Full publication of results initially presented in abstracts. Cochrane Database Syst Rev 2007:MR000005.
15. Song F, Parekh S, Hooper L, Loke Y, Ryder J, Sutton AJ, Hing C, Shing C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
16. Dickersin K: How important is publication bias? A synthesis of available data. AIDS Educ Prev 1997, 9(1 SA):15-21.
17. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009:MR000006.
18. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
19. Dickersin K: How important is publication bias? A synthesis of available data. Aids Educ Prev 1997, 9(1 SA):15-21.
20. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009:MR000006.
21. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
22. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009:MR000006.
23. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
24. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
25. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
26. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
27. Dwan H, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Pang C, Harvey I: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010 in press.
71. Druss BG, Marcus SC: Tracking publication outcomes of National Institutes of Health grants. Am J Med 2005, 118(6):658-663.
72. Cooper H, Charlton K: Finding the missing science: the fate of studies submitted for review by a human subjects committee. Psychological Methods 1997, 2(4):447-452.
73. Hahn S, Williamson PR, Hutton JL: Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. J Eval Clin Pract 2002, 8(3):353-359.
74. Pich J, Carne X, Arnaiz J-A, Gomez B, Trilla A, Rodes J: Role of a research ethics committee in follow-up and publication of results. Lancet 2003, 361(9362):1015-1016.
75. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M: Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. Swiss Med Wkly 2008, 138(13-14):197-203.
76. Dickersin K, Ssemanga E, Mansell C, Rennie D: What do the JAMA editors say when they discuss manuscripts that they are considering for publication? Developing a schema for classifying the content of editorial discussion. BMC Med Res Methodol 2007, 7:44.
77. Calnan M, Smith GD, Sterne JAC: The publication process itself was the major cause of publication bias in genetic epidemiology. J Clin Epidemiol 2006, 59(12):1312-1318.
78. Laupacis A: Methodological studies of systematic reviews: is there publication bias? Arch Intern Med 1997, 157:357.
79. Dubben H-H, Beck-Bornholdt H-P: Systematic review of publication bias in studies on publication bias. BMJ 2005, 331(7514):433-434.
80. Song F: Review of publication bias in studies on publication bias: studies on publication bias are probably susceptible to the bias they study. BMJ 2005, 331(7517):637-638.
81. Hopewell S, Clarke M, Stewart L, Tierney J: Time to publication for results of clinical trials. Cochrane Database Syst Rev 2007:MR0000011.

Pre-publication history
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/9/79/prepub