Few systematic reviews exist documenting the extent of bias: a systematic review

Andrea C. Tricco MSc,1,2 Jennifer Tetzlaff BSc,1,3 Margaret Sampson MLIS,1,4 Dean Fergusson PhD,3 Elise Cogo ND,1,2 Tanya Horsley PhD,3 David Moher PhD.1,3

1 Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada;
2 Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada;
3 Department of Epidemiology & Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada
4 Department of Information Studies, University of Wales, Aberystwyth, Wales.

Please address correspondence and reprint requests to:
David Moher
Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute
401 Smyth Road, Ottawa, Ontario, Canada
Tel: 613-738-3591, email: dmoher@uottawa.ca, fax: 613-738-4800

1 This research was funded, in part, by Chalmers Research Group. Andrea C. Tricco is funded, in part, by a Canadian Institutes of Health Research Doctoral Research Award and the University of Ottawa. Dr. Moher is funded, in part, by a University of Ottawa Research Chair.
What’s New

- To minimize bias during the conduct of systematic reviews (SRs), evidence suggests that authors should include unpublished material, update SRs periodically, search multiple databases, conduct hand searches, use the Cochrane Highly Sensitive Search Strategy to locate randomized controlled trial reports, and assess for publication bias.

- Additional empirical research examining language bias, outcome reporting bias, effects of study risk of bias assessment, and the effects of blinding reviewers during a SR have not been summarized by a SR.

- Future research should also examine bias that can occur during the selection of studies for inclusion and bias that occurs during the synthesis of studies.

- SRs examining many widespread practices for conducting SRs were not identified. Based on the identified gaps in the SR literature, perhaps methodological reviewers have fallen behind.
Abstract

Objective: To summarize the evidence concerning bias and confounding in conducting systematic reviews (SRs).

Study Design and Setting: Literature was identified through searching the Cochrane Library, MEDLINE, PsycINFO until November 2006 and the authors’ files. Studies were included if they were SRs of bias that can occur while conducting a SR. Risk of bias in the SRs was appraised using the Oxman and Guyatt index.

Results: Ten SRs were included. All examined biases related to searching for evidence (e.g., publication bias). One also reported bias associated with obtaining data from included studies (e.g., outcome reporting bias). In order to minimize bias, data suggest including unpublished material, hand searching for additional material, searching multiple databases, assessing for publication bias, and periodically updating SRs. No SRs were found examining bias related to choosing studies for inclusion or combining studies.

Conclusion: There is little evidence from SRs to support commonly practiced methods for conducting SRs. No SRs summarized studies with prospective designs and most had moderate or minimal risk of bias. Future research should examine bias that can occur during the selection of studies for inclusion and the synthesis of studies, as well as systematically review the existing empirical evidence.

Key words: systematic review, bias, research methodology, publication bias

Running title: bias in systematic reviews

Main text: 3454 words; Abstract: 197 words; 1 Glossary; 2 Figures; 5 Tables; 1 Appendix
Introduction

Systematic reviews (SRs) are becoming increasingly popular in evidence-based healthcare [1] and have as their strength, methodological features designed to minimize bias. However, evidence suggests that a large proportion of SRs are poorly reported and susceptible to bias [2,3]. For example, a recent cross-sectional survey found that less than half of published SRs reported using a protocol (46.3%) [2]. The use of protocols minimizes bias because hypotheses and methods are stated \textit{a priori} without prior knowledge of results [4]. Furthermore, the importance of assessing publication bias in SRs has been clearly established [5-7], yet few published SRs consider issues related to publication bias (31.3%) [2]. These finding have serious implications; SRs are often used in the development of clinical practice guidelines [8] and are increasingly viewed as a useful tool for health decision-makers [1,9,10].

Bias that can occur while conducting a SR has been described previously. Fifteen years ago in the journal, Felson put forth a framework of such biases [11]. This framework explained biases in finding all studies (e.g., publication bias, citation bias); biases that can occur while choosing studies to include in the SR (e.g., inclusion criteria bias, selector bias); and obtaining accurate data bias (e.g., bias in scoring study quality, outcome reporting bias).

Evidence-based information regarding the biases outlined in Felson’s framework [11] would provide guidance when conducting a SR. This information can be obtained from SRs that summarize the evidence on biases explained by Felson [11]. We aimed to summarize the evidence for minimizing bias and confounding in conducting SRs and examine gaps in this literature by conducting a SR.
Methods

Eligibility criteria

A SR was defined as any study for which “the authors' stated objective was to summarize evidence from multiple studies and the article described explicit methods, regardless of the details provided” [2,12]. When it was clear that the intent of the authors was a literature review (e.g., authors identified the review as a brief overview with no specific review question), as opposed to a SR, articles were excluded [2,12]. We included SRs of empirical studies examining bias and confounding that can occur during the conduct of a SR. The following types of biases relevant to our review identified a priori included: publication, indexing/citation, language, time-lag, multiple/duplicate publication, outcome reporting, and study quality biases. These were identified from two articles outlining bias in SRs [5,11] and discussions between the investigators.

Search strategy

SRs were identified through electronic searches in MEDLINE (1966 to November Week 3 2006, Ovid interface), PsycINFO (1806 to November Week 4 2006, Ovid interface), the Cochrane Library (2006, Issue 4, Wiley interface), limited to the English language, and the PubMed ‘related articles’ link for all references from the Quality of Reporting of Meta-analyses (QUOROM) Statement checklist items [13]. Electronic searches were supplemented by using studies from the authors' personal files, contacting experts for grey literature and additional material, and scanning the reference lists of all included SRs. The electronic search strategies were developed and validated by two information
specialists (EC, MS; Appendix A). Electronic searches were conducted on February 1, 2006 and updated on November 20, 2006.

Screening

One reviewer (ACT) screened the titles and abstracts identified by the literature search for study inclusion using a pre-defined study relevance form. This was verified by a second reviewer (JT) who screened a random sample of 1/3 of the records. The full-text of potentially relevant articles was obtained for further evaluation to determine inclusion.

Data abstraction

Data were abstracted by one reviewer (ACT), using a pre-specified standardized 20-item data abstraction form, and verified by a second reviewer (JT) using a 1/5 random sample. Abstracted data included study characteristics (e.g., first author, country or countries where the research originated); the number, study designs, and methodological quality of studies included in the SR; types of bias examined; author’s definition of each type of bias; and the SR results. The estimated effect size of bias (e.g., relative risk) and respective confidence intervals were also abstracted.

The SR biases were categorized as follows: 1) biases in finding all studies (sampling bias), 2) biases in selecting studies for inclusion, 3) biases in obtaining accurate data from selected studies, and 4) biases that occur when studies are combined [11].

Risk of bias assessment

The Oxman and Guyatt Overview Quality Assessment Questionnaire was used to assess the risk of bias in the included SRs [14]. This validated instrument consists of 9 main criteria for assessing the scientific quality of review articles [14]. The final item asks the
assessor to rate the overall scientific quality of the SR using a score ranging from 1 (i.e., extensive flaws) to 7 (i.e., minimal flaws). Two reviewers conducted a training exercise using this instrument (ACT, JT) and independently rated all studies (ACT, JT). Disagreements were resolved through discussion.

**Data synthesis**

The agreement between the reviewers who screened the literature (AT, JT) was assessed using a kappa statistic [15]. We determined *a priori* that an acceptable level of agreement would be greater than 60% [15]. Results were summarized narratively and quantitative results from the relevant SRs were visually presented in a forest plot.

**Results**

A total of 3733 records were identified through the searches and subsequently screened. Of these, 221 full-text articles were obtained for further examination to determine relevance, and 10 SRs met our eligibility criteria [5,16-24] (Figure 1). One of these SRs [23] was identified as an update of a previous SR [16], leaving a total of nine unique relevant SRs. We also identified six Cochrane reviews published as protocols [25-30], which will be included in any subsequent update of this SR (Table 1). Good agreement was observed between the two reviewers (ACT, JT) at the full-text screening level (kappa = 0.67).

On average, the SRs included 35 studies (range: 2-79) with retrospective cohort, cross-sectional and/or case series study designs, and all were published between 2000 and 2005 (Table 2). Half of the reviews reported or conducted quantitative data synthesis (Figure 2).
Of all included SRs, one did not assess the risk of bias [22] and two did not report assessing the risk of bias [5,24]. The remaining SRs assessed the risk of bias using a component approach [19,23], a checklist [17,18], and by examining other methodological issues (e.g., response rate, databases searched) [20,21] (Table 3).

The risk of bias in the SRs themselves varied greatly. The Oxman and Guyatt index [14] indicated that one SR had major flaws [24], four had minor flaws [5,21-23] and the remaining had minimal flaws [17-20]. Studies were consistently assessed as having an increased risk of bias on the study selection (n=6) and validity (n=5) criteria (Table 4).

**Systematic review biases**

1. **Biases in identifying studies (sampling bias)**
2. A) **Publication-related biases**

**Publication bias**

Two reviews examining publication bias were identified [5,24]. Seventeen of the 26 (65%) included studies overlapped between both SRs. Song et al. provided information specific to publication bias and included 11 cross-sectional studies [5] (Table 5). The percentage of statistically significant results ranged from 35-97% across emergency, medical, biological, and psychology journals [31-35]. The existence of publication bias was consistently confirmed by four retrospective cohort studies of research approved by research ethics boards and trial registries [6,36-38]. The results of a quantitative data synthesis for which a comprehensive literature search was not performed were reported in the Song et al. SR [7]. The overall adjusted odds ratio for publication bias was 2.54 (95% confidence interval: 1.44 to 4.47; Figure 2).
Dubben et al. included 26 retrospective cohort and cross-sectional studies examining publication bias [24]. The median effect size was 2.3 (confidence interval not reported), indicating that published studies were more than twice as likely to report positive results.

**Grey literature bias**

Two SRs examined the effects of including grey literature in SRs [5,18]. There was no overlap in the included studies in these reviews. Song et al. included one retrospective cohort and four cross-sectional studies, and found that trials with statistically positive results were more likely to be published [5]. Hopewell et al. included eight cross-sectional studies [18]. Overall, published trials included more participants (median 46 (IQR 4-300) versus 5.5 (IQR 4-88)), were more likely to have statistically significant results (30% versus 19%, p<0.05), and were less susceptible to bias than grey literature trials [18].

**Funding bias**

Two SRs provided insight into bias pertaining to source of funding [20,22]. Lexchin et al. included 30 primary studies, eight (27%) of which were also included in Bekelman et al. Based on 18 comparisons from 15 cohort, cross-sectional, and case studies, pharmaceutical sponsorship was associated with positive outcomes (odds ratio: 4.05, 95% confidence interval: 2.98 to 5.51; Figure 2, Table 5) [20]. All 13 studies examining an association between study quality and funding source found that industry-funded studies were of comparable quality to studies with other funding sources. Bekelman et al. reported similar results in their SR (Figure 2, Table 5).

**Time-lag bias**
Two SRs examined time-lag bias [5,17]. One SR included two retrospective cohort studies, which followed trials from the date of follow-up completion or date of ethics approval to the date of publication [17]. These studies found that 55-58% of all trials were published in full and trials with statistically significant results in favour of the experimental intervention were published quicker (range: 4-5 years) than those with non-significant results (range: 6-8 years, p<0.05) [17]. Song et al. also included these two studies, as well as another retrospective cohort study and a cross-sectional study [5]. In the retrospective cohort study included by Song et al., the hazard ratio for time to publication from trial completion for positive versus negative trials was 3.7 (95% confidence interval: 1.8 to 7.7) [39]. In a cross-sectional study of 26 meta-analyses included by Song et al., the treatment effect was exaggerated in the majority of trials published early in the drug development cycle compared with later cycle trials (average difference in relative odds: 35%, 95% confidence interval: 15% to 55%) [40].

Abstract to full publication bias

Three SRs examined abstract to full publication bias [5,21,23] and findings were consistent across all reviews. Scherer et al. included 79 retrospective cohort studies, 10 (13%) of which were reviewed in the von Elm et al. SR (n=21 included studies). Scherer et al. determined that 44.5% of abstracts presented at meetings were subsequently published in full-text (95% confidence interval: 43.9-45.1%) [23]. After nine years, 52.6% of all abstracts, 63.1% of randomized controlled trial (RCT) abstracts, and 49.3% of abstracts with other study designs were published in full. Positive results were associated with full publication; as was oral presentation, acceptance for meeting
presentation, RCT design, higher quality RCT abstracts, and clinical research (versus basic science research; Figure 2, Table 5).

von Elm et al. also examined the full publication of results that were presented at meetings [21]. They found similar survival analysis results and noted that abstracts with positive versus negative outcomes and oral presentations versus poster presentations were more likely to be published. However, they found that abstracts about basic science were more likely to be published than those on clinical science. Rejected abstracts had the same long-term publication rate as accepted abstracts in this SR (Figure 2, Table 5).

Song et al. described 22 studies that examined full publication bias [5]. The three studies that were not included by Scherer et al. and von Elm et al. found similar results to those reported above.

**Place of publication bias**

One review (Song et al.) summarized three cross-sectional studies that examined place of publication bias. These cross-sectional studies found that the *British Medical Journal* was more likely to publish articles on the ‘early life hypothesis’ (e.g., relationships between indicators of fetal development and later disease patterns) than *The Lancet* [41]. Furthermore, journals considered to be ‘prominent’, such as *Cancer* and the *New England Journal of Medicine* published a higher proportion of positive trials while less well known journals only published trials with statistically negative results [42].

**Country of conduct bias**

Song et al. summarized two cross-sectional and one case study that examined country of conduct bias [5]. In one of these studies, the estimated efficacy of a complementary and alternative therapy was greater in studies published outside the United States when
compared to those published in the United States [43]. In another study, the proportion of positive results in trials from China, Taiwan, Japan, and Hong Kong was 100% compared to 56.7% for similar trials published in other countries including Canada, the United States, and Germany [44].

Language bias

Song et al. summarized five cross-sectional studies that examined language bias and found conflicting evidence for the impact of including or excluding non-English language reports in SRs [5].

B) Locating studies using electronic databases

Indexing bias

Two SRs examined indexing bias [5,19]. Song et al. reviewed eight cross-sectional studies that consistently reported an increased likelihood of missing relevant reports when only one electronic database is searched. Hopewell et al. examined the effects of hand searching versus electronic searching for identifying reports of RCTs and included 34 cross-sectional studies [19], none of which overlapped with the Song et al. SR. Hand searching identified 92-100% of RCTs compared to 42-80% of RCTs when searched electronically [19]. It was concluded that hand searching is valuable in conducting SRs of RCTs, especially for identifying RCTs reported as grey literature, published in languages other than English, and published in journals not indexed in electronic databases.

Search bias

Hopewell et al. summarized evidence for search bias [19]. The electronic Cochrane Highly Sensitive Search Strategy (HSSS) identified 80% of RCTs compared to 65% for electronic complex searches, and 42% for electronic simple searches.
C) Locating studies using reference lists

Citation bias

One SR examined citation bias [5]. Song et al. summarized nine cross-sectional studies, which found that authors will cite studies that confirm their results and that statistically positive trials are cited more frequently compared to non-positive trials. For example, one study examining 76 trials found that positive references were cited more frequently (58%) than negative (29%) or neutral (13%) citations [45].

Multiple/Duplicate publication bias

Song et al. reviewed four cross-sectional studies and one retrospective cohort study that examined multiple publication bias [5]. In one study that examined 44 multiple publications of 31 trials, it was determined that the conclusions of multiple publications became more positive over time [46]. In another study, reports with statistically significant results were more likely to generate duplicate publications than those with statistically non-significant results [36]. Including duplicate data overestimated the treatment effect in a study that examined a sample of trials [47].

2. Choosing studies biases

This includes inclusion criteria bias and selector bias [11]. We did not identify a SR that specifically assessed these biases.

3. Obtaining accurate data biases

A) Bias by the systematic reviewer

This includes bias in scoring study quality and extractor bias [11]. We did not identify a SR that specifically assessed these biases.

B) Bias due to inaccurate reporting of study results
This includes *study quality bias* and *recording error bias* [11]. We did not identify a SR that specifically assessed these biases.

**Outcome reporting bias**

One review examined the phenomenon of outcome reporting bias [5]. The review described two cross-sectional and one case study; all found that published trials were more likely to report statistically significant outcomes.

**4. Combining studies biases**

Bias can occur when statistically combining studies in a SR (e.g., *indirect comparison bias*). We did not identify a SR that specifically assessed these biases.

**Discussion**

We identified few SRs documenting the extent of bias that can occur while conducting a SR. Our extensive literature search only identified 10 SRs, one of which was an updated SR. Although few SRs were identified, our results have implications for systematic reviewers. Empirical evidence for publication bias, time-lag bias, abstract to full publication of bias, funding bias, and grey literature bias was identified. These will be described further below.

SRs based only on published material may have exaggerated effect sizes, thus grey literature should be included in SRs. Grey literature, such as conference abstracts, should be sought and included, as evidence suggests that ‘positive’ trials presented as abstracts, oral versus poster presentations, and RCT designs have a greater likelihood of being published in full. Statistically significant studies tend to be published earlier, over-estimating effect sizes of SRs. Therefore, SRs should be routinely updated. Issues
surrounding how and when to update SRs was recently addressed in a recent SR conducted by some of us and is forthcoming in the *journal* [48].

Furthermore, at least one database should be searched and although labour-intensive, hand searching should be considered whenever feasible. The Cochrane HSSS has the potential to locate a large proportion of RCTs in major English electronic databases and should be used, whenever possible. Although funding bias is omnipresent in the published literature, industry-sponsored publications were of comparable risk of bias to those sponsored by other sources.

We identified SRs examining several biases, yet further investigation into the following is warranted: 1) place of publication bias, 2) country bias, 3) search bias, 4) citation bias, 5) multiple publication bias, and 6) outcome reporting bias (a SR is planned; Dr. P. Williamson, personal communication). The SRs themselves should be updated, as new evidence may have emerged. Although not a mandate of this review, we believe it is important to explore whether common SR practices do in fact decrease bias, such as having two people independently screen potentially relevant material, as well as scanning the reference lists of the included studies in a SR.

Although many types of bias were covered in the included SRs, gaps in the SR methodological literature were apparent. Our literature search identified additional studies that have yet to be included in a SR of bias. Six cross-sectional studies examining language bias [49-54], six retrospective cohort studies examining outcome reporting bias [4,55-59]; 11 cross-sectional studies examining the effects of study risk of bias [52,60-69], although we are aware of a very recent publication examining one component of this [70]; and six trials examining whether systematic reviewers should be masked while
scoring study quality have not been systematically reviewed [65,71-75]. There may be
other gaps in the SR methodological literature that have not yet been fully realized. For
example, ‘incomplete reporting bias’; a bias that occurs when studies are omitted from
meta-analysis because of incompletely reported information (e.g., measure of dispersion)
[76,77]. Even though the literature search identified six SR protocols (Table 1), gaps in
the literature are still apparent.

Forty years ago, Archie Cochrane challenged health researchers to systematically
review research across all specialties [78]. Based on the identified gaps in the SR
literature, perhaps methodological reviewers have fallen behind. For example, in the
Cochrane Library Issue 2 2007 there were 4801 Cochrane reviews, only 19 of which were
methodological reviews.

We challenge systematic reviewers to conduct additional high quality
methodological reviews. This will not only inform systematic reviewers in general, but
will also impact and inform granting agencies that fund SRs; groups that conduct SRs,
such as the Cochrane Collaboration and the Agency for Healthcare Research and Quality
(AHRQ) Evidence-based Practice Center program; those interested in developing
reporting guides for SRs, such as the Preferred Reporting Items for Systematic Reviews
and Meta-Analyses (PRISMA) Group and the Enhancing the Quality and Transparency
of health Research (EQUATOR) Network; and texts of how best to conduct SRs.

One of the included SRs had major flaws, four had minor flaws, and four had
minimal flaws when assessed for risk of bias. Although 5/9 of the included SRs
conducted or reported a meta-analysis, none addressed the potential confounding or effect
modification of their results. The effects of a variety of biases on the results should be
examined in methodological reviews that conduct a quantitative synthesis, perhaps through stratification, sensitivity analyses, or specifically controlling for confounders and effect modifiers in their analyses. Furthermore, none of the SRs located or included prospective studies as their primary studies. It may be the case that not many methodological prospective studies (e.g., RCTs, prospective cohort studies) examining bias at the SR level have been conducted, which should be a priority research area for methodologists.

Our SR has limitations. Studies published in languages other than English were excluded. Only one independent reviewer screened the titles and abstracts for inclusion. We relied on the study-level risk of bias assessment reported in the SRs. Furthermore, we may have missed pooled analyses that were not based on evidence from a SR (e.g., [79]).

Our findings recommend including unpublished material in SRs, updating SRs periodically, searching more than one database, hand searching for additional material, using the Cochrane HSSS to locate RCT reports, and assessing for publication bias. Further examination of the other types of bias identified in our SR is warranted and the existing empirical evidence should be systematically reviewed.

Acknowledgements

We would like to thank Nick Barrowman for his statistical consultation. The first author would like to thank Drs Moher and Fergusson for supporting this work, as an earlier version was conducted as an assignment for their ‘Systematic reviews and meta-analysis’ course at the Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa.
Potential conflict of interest

None declared.
Glossary of biases:

**Publication bias:** Occurs when investigators, reviewers, and editors submit or accept manuscripts for publication based on the direction or strength of the study findings [7].

**Grey literature bias:** Occurs when the results reported in journal articles are systematically different from those presented in other reports, such as working papers, dissertations, or conference abstracts [5].

**Funding bias:** Biases in the design, outcome, and reporting of industry sponsored research in order to show that a drug shows a favourable outcome [22].

**Time-lag bias:** Occurs when the speed of publication depends on the direction and strength of the trial results [80].

**Abstract to full publication bias:** Occurs when the full publication of studies that have been initially presented at conferences or in other informal formats is dependent on the direction and/or strength of their findings [5].

**Place of publication bias:** Occurs when a journal is more enthusiastic towards publishing articles about a given hypothesis than other journals because of editorial policy or readers' preference [41].

**Country of conduct bias:** Occurs when the country of publication is associated with the strength or direction of research findings [5].

**Language bias:** Occurs when languages of publication depend on the direction and strength of the study results [81].

**Indexing bias:** Occurs when there is biased indexing of published studies in literature databases [11].

**Search bias:** Occurs when there is a bias in captured studies resulting from an inadequate or incomplete search [11].

**Citation bias:** Occurs when the chance of a study being cited by others is associated with its results [5]. Therefore, retrieving literature from scanning reference lists may produce a biased sample of articles, rendering the conclusions of an article less reliable [45].

**Multiple/duplicate publication bias:** Occurs when studies with significant or supportive results are more likely to generate multiple publications than studies with non-significant or unsupportive results [5]. It can be classified as overt or covert, the latter being more difficult to deal with in systematic reviews [47].

**Inclusion criteria bias:** Occurs when the inclusion criteria of a review purposely excludes important studies that the reviewer knows of [11].
Selector bias: Occurs when the inclusion criteria are not specific enough, leaving the reviewer free to choose studies, which may be susceptible to bias [11].

Bias in scoring study quality: Occurs when the systematic reviewer systematically scores studies published by their peers or in high-impact journals as being more methodologically rigorous [11].

Extractor bias: Occurs when the data is not extracted accurately from the study [11].

Study quality bias: Occurs when studies of lower or higher quality are associated with positive or favourable results [11].

Recording error bias: Occurs when the actual study results and the recorded results in the published paper differ [11].

Outcome reporting bias: Occurs when a study in which multiple outcomes were measured reports only those that are significant [5].

Indirect comparison bias: Occurs when indirect comparisons rather than direct comparisons are used to combine results in a systematic review.
Figure 1: Study flow chart

Level 1 (Titles & Abstracts)

N=3733 potentially relevant records

N=3512 excluded records:
Main reasons:
1) study did not examine bias
2) editorial/commentary

Level 2 (Full text)

N=221 potentially relevant records

N=211 excluded records:
1) study did not examine bias (n=42)
2) study did not examine bias that can occur during systematic review conduct (n=9)
3) systematic review protocol (n=6)
4) study was not a systematic review (n=154)

Level 3 (Abstraction)

N=10 included systematic reviews, of which 9 were unique
Figure 2: Forest plot of pooled estimates from random-effects meta-analyses examining bias

Publication bias
Positive versus negative results
Song 2000

Funding bias
Positive versus negative results
Lexchin 2003
Bekelman 2003

Abstract to full publication bias
Positive versus negative results
von EIm 2003
Scherer 2005
Basic versus clinical research
von EIm 2003
Scherer 2005
Results favour treatment
Scherer 2005

RCT positive versus negative results
Scherer 2005

Accepted for meeting versus not
Scherer 2005

RCT versus non-RCT
Scherer 2005

High quality RCT versus low
Scherer 2005

Favours Bias
Abbreviation: RCT randomized controlled trial. Legend: Odds ratio ☐; relative risk  ●
<table>
<thead>
<tr>
<th>Author/Year, Reference</th>
<th>Topic</th>
<th>Search methods</th>
<th>Search limits</th>
<th>Bias category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarke/2007 [25]</td>
<td>Individual patient MA versus aggregate MA</td>
<td>CMR, MEDLINE, EMBASE, hand search, Internet search, contact experts, PubMed related articles*</td>
<td>NR</td>
<td>Combining studies bias</td>
</tr>
<tr>
<td>Olsen/2007 [28]</td>
<td>Publication bias in clinical trials</td>
<td>CMR, MEDLINE, contact experts, search reference lists*</td>
<td>NR</td>
<td>Sampling bias</td>
</tr>
<tr>
<td>Song/2007 [29]</td>
<td>Indirect comparisons</td>
<td>CDSR, DARE, MEDLINE, hand search, search reference lists*</td>
<td>NR</td>
<td>Combining bias</td>
</tr>
</tbody>
</table>

Notes: *Years of search not reported.

Abbreviations: MA (meta-analysis), CMR (Cochrane Methodology Register), NR (not reported), SCI (Science Citation Index), CINAHL (Cumulative Index to Nursing and Allied Health Literature), LISA (Library and Information Science Abstracts), SIGLE (System for Information on Grey Literature in Europe), CDSR (Cochrane Database of Systematic Reviews), DARE (Database of Abstracts of Review of Effects).
<table>
<thead>
<tr>
<th>Author/Year, Reference</th>
<th># of studies included and study designs</th>
<th>Search methods</th>
<th>Search limits</th>
<th>SR bias category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopewell/2001 [17]</td>
<td>2 RC</td>
<td>CMR, MEDLINE, EMBASE, SCI, PubMed related articles, search reference lists, hand search, contact experts**</td>
<td>NR</td>
<td>Sampling bias</td>
</tr>
<tr>
<td>von Elm/2003 [21]</td>
<td>64 RC</td>
<td>MEDLINE, EMBASE, Cochrane Library, CINAHL, BIOSIS, SCI, search reference lists, handsearch, author contact, Internet searches**</td>
<td>All languages, all formats</td>
<td>Sampling bias</td>
</tr>
<tr>
<td>*Scherer/2005 [23]</td>
<td>79 RC</td>
<td>MEDLINE (up to 2003), EMBASE (up to 2003), CMR (2003), search reference lists, SCI (up to 2003), author's files, word of mouth</td>
<td>NR</td>
<td>Sampling bias</td>
</tr>
<tr>
<td>Scherer/1994 [16]</td>
<td>11 RC</td>
<td>MEDLINE, author's files, word of mouth**</td>
<td>NR</td>
<td>Sampling bias</td>
</tr>
</tbody>
</table>

**Notes:** *Represents the major publication of the systematic reviews, **Years of search not reported.

**Abbreviations:** RC (retrospective cohort), CS (cross-sectional), CMR (Cochrane Methodology Register), LISA (Library and Information Science Abstracts), DA (Dissertation abstracts), BEI (British Education Index), SIGLE (System for Information on Grey Literature in Europe), NR (not reported), SCI (Science Citation Index), AMED (Allied and Complementary Medicine Database), BIOSIS (Biological abstracts), CINAHL (Cumulative Index to Nursing and Allied Health Literature).
Table 3: Risk of bias in studies included in the systematic reviews

<table>
<thead>
<tr>
<th>Author/Year, Reference</th>
<th>Type of risk of bias assessment tool</th>
<th>Risk of bias assessment results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song/2000 [5]</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Hopewell/2001 [17]</td>
<td>Checklist</td>
<td>Explicit inclusion criteria (2/2), control for clinical differences (1/2) &amp; unclear control (1/2), complete data (1/2), no important flaws (1/2), possible important flaws (1/2)</td>
</tr>
<tr>
<td>Hopewell/2002 [19]</td>
<td>Component</td>
<td>Appropriate hand search (17/34) &amp; unclear (17/34), appropriate electronic search (29/34) &amp; unclear (5/34), eligibility hand search agreement (11/34) &amp; unclear (23/34), eligibility electronic search agreement (8/34) &amp; unclear (24/34), comparable search (28/34) &amp; unclear (6/34)</td>
</tr>
<tr>
<td>Lexchin/2003 [22]</td>
<td>Not conducted**</td>
<td>Not applicable</td>
</tr>
<tr>
<td>von Elm/2003 [21]</td>
<td>Formal risk of bias not conducted; instead indicators of methodological quality (e.g., details on follow-up period, databases searched) were assessed</td>
<td>Not reported</td>
</tr>
<tr>
<td>Bekelman/2003 [20]</td>
<td>Formal risk of bias not conducted; instead components of the study design of included studies were assessed. For example, the sample size and response rate were assessed for cross-sectional studies.</td>
<td>Not reported</td>
</tr>
<tr>
<td>*Scherer/2005 [23]</td>
<td>Component</td>
<td>Most studies used an unbiased sample of abstracts, most had at least 2 years of follow-up, and there was fair ascertainment of subsequent publication</td>
</tr>
<tr>
<td>Scherer/1994 [16]</td>
<td>Not conducted**</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Dubben/2005 [24]</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

Notes: *Represents the major publication of the systematic reviews, **Not conducted means that they reported that they did not assess risk of bias.
### Table 4: Risk of bias in included systematic reviews (Oxman & Guyatt tool)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Search methods</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>Search comprehensiveness</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>Inclusion criteria</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>Bias in study selection</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>Criteria for validity</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>Appropriate validity items</td>
<td>C</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>C</td>
<td>Y</td>
<td>Y</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>Combining methods</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>Appropriate combining</td>
<td>C</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>Appropriate conclusions</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td><strong>Score</strong></td>
<td></td>
<td><strong>4</strong></td>
<td><strong>7</strong></td>
<td><strong>7</strong></td>
<td><strong>7</strong></td>
<td><strong>4</strong></td>
<td><strong>5</strong></td>
<td><strong>6</strong></td>
<td><strong>5</strong></td>
<td><strong>3</strong></td>
<td><strong>2</strong></td>
</tr>
</tbody>
</table>

**Notes:** *Represents the major publication of the systematic reviews, **Scoring: Total score is out of 7. A score of 1 means the review has extensive flaws, 2-3 major flaws, 4-5 minor flaws, and 6-7 minimal flaws

**Abbreviations:** Y (yes), N (no), P (partially), C (can't tell)
Table 5: Biases examined in the included systematic reviews

<table>
<thead>
<tr>
<th>Author/Year, Reference</th>
<th>Details of biases examined</th>
<th>Pooled results of SR: #comparisons, ES: (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Sampling bias: biases in identifying studies for the systematic review</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A) Publication bias</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i) Publication bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song/2000 [5]</td>
<td>Determining the proportion of statistically significant results in the published literature over time</td>
<td>Not conducted</td>
</tr>
<tr>
<td>Song/2000 [5]</td>
<td>Examining the association between type of result (e.g., statistically significant favourable) and publication status (e.g., published)</td>
<td>4 comparisons, adjusted OR for publication bias 2.54 (1.44, 4.47)</td>
</tr>
<tr>
<td>Dubben/2005 [24]</td>
<td>Examining the association between type of result and publication status</td>
<td>Not conducted, reported a median reported OR of 2.3</td>
</tr>
<tr>
<td>ii) Grey literature bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song/2000 [5]</td>
<td>Examining the association between grey literature and type of result</td>
<td>Not conducted</td>
</tr>
<tr>
<td>Hopewell/2002 [18]</td>
<td>Examining the association between grey literature and type of result</td>
<td>Not conducted</td>
</tr>
<tr>
<td>iii) Funding bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexchin/2003 [22]</td>
<td>Examining the association between funding source and publication status</td>
<td>Not conducted</td>
</tr>
<tr>
<td></td>
<td>Examining the association between funding source and type of result</td>
<td>18 comparisons, pharmaceutical sponsorship associated with positive outcomes OR: 4.05 (2.98, 5.51), homogeneity (p=0.17)</td>
</tr>
<tr>
<td></td>
<td>Examining the association between funding source and study quality</td>
<td>Not conducted</td>
</tr>
<tr>
<td>Bekelman/2003 [20]</td>
<td>Examining the association between industry sponsorship and type of result</td>
<td>8 comparisons, industry sponsored studies associated with positive results OR:3.60 (2.63, 4.91), homogeneity (p=0.75)</td>
</tr>
<tr>
<td></td>
<td>Examining the association between industry sponsorship and study quality</td>
<td>Not conducted</td>
</tr>
<tr>
<td>iv) Time-lag bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song/2000 [5]</td>
<td>Examining the association between time to publication and type of result</td>
<td>Not conducted</td>
</tr>
<tr>
<td>Hopewell/2001</td>
<td>Examining the association between time to</td>
<td>Not conducted</td>
</tr>
</tbody>
</table>
### v) Abstract to full publication bias

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song/2000</td>
<td>Examining the association between abstract characteristics (e.g., basic science, favourable result) and being published in full</td>
<td>Not conducted</td>
</tr>
<tr>
<td>von Elm/2003</td>
<td>Examining the association between abstract characteristics and being accepted for a conference presentation</td>
<td>46% abstracts submitted to meetings were accepted, acceptance when topic was basic vs clinical OR:3.49 (2.50, 4.86) and the outcome was statistically significant favourable vs statistically significant unfavourable OR:1.67 (1.16, 2.39), heterogeneity NR</td>
</tr>
<tr>
<td></td>
<td>Examining the association between abstract characteristics and being published in full</td>
<td>Abstracts were more likely to be published when topic was basic vs clinical OR:2.29 (1.75, 2.98), the outcome was positive vs negative OR:2.07 (1.58, 2.71), and it was an oral presentation vs poster OR:1.53 (1.15, 2.03), heterogeneity NR</td>
</tr>
<tr>
<td></td>
<td>Examining the association between being rejected for a conference presentation and being published in full</td>
<td>27% abstracts published despite meeting rejection</td>
</tr>
<tr>
<td><em>Scherer/2005</em></td>
<td>Examining the association between abstract characteristics and being published in full</td>
<td>44.5% abstracts subsequently published, more likely to be published when there are statistically significant results RR:1.30 (1.14, 1.47, $\chi^2$: p=0.0006), results favour treatment RR:1.17 (1.02, 1.35, $\chi^2$: p=0.01), positive results from RCTs RR:1.18 (1.07, 1.30, $\chi^2$: p=0.14), oral presentation RR:1.28 (1.09, 1.49, $\chi^2$: p&lt;0.0001), accepted for meeting presentation RR:1.78 (1.50, 2.12, $\chi^2$: p&lt;0.0001), RCT design OR:1.24 (1.14, 1.36, $\chi^2$: p=0.56), basic research RR:0.79 (0.70, 0.89, $\chi^2$: p=0.0009), and higher quality RR:1.30 (1.00, 1.71, $\chi^2$: p=0.68)</td>
</tr>
<tr>
<td>Scherer/1994</td>
<td>Examining the association between abstract characteristics and being published in full</td>
<td>11 comparisons, 51% abstracts subsequently published, more likely to be published when there are significant results RR:1.17 (0.99, 1.39) and a sample size above the median RR:1.48 (1.14, 1.94), homogeneity (p=0.01)</td>
</tr>
</tbody>
</table>

### vi) Place of publication bias

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song/2000</td>
<td>Examining the association between study characteristics (e.g., topic examined, statistically significant favourable result) and being published in different journals</td>
<td>Not conducted</td>
</tr>
</tbody>
</table>

### vii) Country of conduct bias

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song/2000</td>
<td>Examining the association between study characteristics</td>
<td>Not conducted</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>viii) Language bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Song/2000</strong> [5]</td>
<td>Examining the association between language of publication and study characteristics (e.g., publication status, type of result)</td>
<td>Not conducted</td>
</tr>
</tbody>
</table>

*B) Locating studies using electronic databases*

**Indexing bias**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Song/2000</strong> [5]</td>
<td>Examining the association between type of search (e.g. hand searching versus electronic) and identifying all relevant material</td>
<td>Hand search identified 92-100% of RCTs whereas electronic searches identified 42-80% of trials.</td>
</tr>
<tr>
<td><strong>Hopewell/2002</strong> [19]</td>
<td>Examining the association between type of search (e.g. hand searching versus electronic) and identification of all relevant material</td>
<td>Hand search identified 92-100% of RCTs whereas electronic searches identified 42-80% of trials.</td>
</tr>
</tbody>
</table>

**ii) Search bias:**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hopewell/2002</strong> [19]</td>
<td>Examining the association between type of electronic search (e.g., simple versus complex) and identification of all relevant material</td>
<td>Electronic Cochrane Highly Sensitive Search identified 80% of RCTs, electronic complex searches identified 65% RCTs, and electronic simple searches 42% RCTs, results were judged to be homogeneous but no formal test of heterogeneity conducted.</td>
</tr>
</tbody>
</table>

*C) Finding studies using reference lists*

**i) Citation bias:**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Song/2000</strong> [5]</td>
<td>Examining the association between searching reference lists and identifying all relevant material</td>
<td>Not conducted</td>
</tr>
</tbody>
</table>

**ii) Multiple/duplicate publication bias:**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Song/2000</strong> [5]</td>
<td>Examining the association between effect sizes and including duplicate data</td>
<td>Not conducted</td>
</tr>
</tbody>
</table>

2. **Choosing studies bias: biases in study selection in the systematic review**

**A) Inclusion criteria bias†**

**i) Study inclusion bias**

**B) Selector bias†**

3. **Obtaining accurate data bias: biases in obtaining accurate data from included studies in the systematic review**

**A) Bias by the systematic reviewer†**

**i) Bias in scoring study quality**

---

*Note: The table and text are from a publication on bias in systematic reviews.*
ii) Extractor bias

**B) Bias due to inaccurate reporting of the study results**

i) Outcome reporting bias

| Song/2000 [5] | Examining the association between outcome characteristic (e.g., statistically significant favourable) and being reported in the trial. | Not conducted |

ii) Study quality bias†

iii) Recording error bias†

### 4. Combining studies bias: biases that occur when results are combined

i) Indirect comparison bias: Occurs when indirect comparisons rather than direct comparisons are used to combine results in MAs.

**Notes:** *Represents the major publication of the systematic reviews. † We did not identify a systematic review that specifically examined this type of bias.

**Abbreviations:** SR (systematic review), MA (meta-analysis), ES (effect size), CI (confidence intervals), RCTs (randomized controlled trials), OR (odds ratio), RR (relative risk).
Appendix A: Search strategies

MEDLINE (Ovid interface) search strategy:

1. bias$.ti,ab.
2. exp "bias(epidemiology)"/
3. Publication bias/
4. Location bias.mp.
5. Citation bias.mp.
7. Reference bias.mp.
8. Multiple publication bias.mp.
9. RANDOMIZED CONTROLLED TRIAL.pt.
10. CONTROLLED CLINICAL TRIAL.pt.
11. RANDOMIZED CONTROLLED TRIALS.sh.
12. RANDOM ALLOCATION.sh.
13. DOUBLE BLIND METHOD.sh.
14. SINGLE BLIND METHOD.sh.
15. or/9-14
17. 15 not 16
18. CLINICAL TRIAL.pt.
19. exp CLINICAL TRIALS/
21. ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.
22. PLACEBO.sh.
23. placebo$.ti,ab.
24. random$.ti,ab.
25. versus.tw.
26. RESEARCH DESIGN.sh.
27. or/18-26
28. 27 not 16
29. 28 not 17
30. 17 or 29
31. (or/1-8) and 30
32. meta-analys$.mp.
33. systematic$ review$.mp.
34. 32 or 33
35. 31 and 34
36. limit 35 to English language

PsycInfo (Ovid interface) search strategy:

1. EXPERIMENTER BIAS/ or TEST BIAS/ or CULTURAL TEST BIAS/ or RESPONSE BIAS/
2. bias$.ti,ab.
3. bias$.mp.
4. 2 or 3
5. 4 not 1
6. (meta-analys$ or systematic$ review$).mp.
7. 5 and 6
8. remove duplicates from 7
9. limit 8 to English language

Cochrane Database of Methodological Reviews & Cochrane Methodological Registry (Wiley interface) search strategy:
1. Bias* in all fields
2. Meta-analys* or systematic* review* in all fields
3. Meta-analysis or review in publications type
4. MeSH descriptor Meta-analysis explode all tree in MeSH products
5. #1 and (#2 or #3 or #4)
Reference List


13. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: Improving the quality of reports of meta-analyses of randomised controlled trials: the


54. 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. *J Clin Epidemiol* 2005, **58**: 769-776.


