Methods Research Report

The Empirical Evidence of Bias in Trials Measuring Treatment Differences

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Research Report

The Empirical Evidence of Bias in Trials Measuring Treatment Differences

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10056-I

Prepared by:
RTI-UNC Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Nancy D. Berkman, Ph.D., M.L.I.R.
P. Lina Santaguida, Ph.D.
Meera Viswanathan, Ph.D.
Sally C. Morton, Ph.D.

AHRQ Publication No. 14-EHC050-EF
September 2014

This report is based on research conducted by the RTI International-University of North Carolina Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10056-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The Empirical Evidence of Bias in Trials Measuring Treatment Differences. Methods Research Report. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I.) AHRQ Publication No. 14-EHC050-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Yen-pin Chiang, Ph.D.
Acting Deputy Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang M.D., M.P.H.
Director, EPC Program and Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors thank our Research Assistant Priyanka Sista for the support she provided to the senior researchers through the peer review portion of this project. We would also like to thank Maureen Rice, Mary Gauld, and Julianna Beckett at the McMaster EPC, who assisted with literature identification, retrieval, and review. Lastly, we acknowledge the expert document preparation provided by Loraine Monroe at RTI International.

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who participated in developing this report follows:

Ethan Balk, M.D., M.P.H.
Institute for Clinical Research and Health Policy Studies Tufts Medical Center
Boston, MA

Julian Higgins, Ph.D.
University of Bristol
Bristol, UK

David Moher, Ph.D.
Ottawa Hospital Research Institute
Ottawa, Ontario, Canada

Jonathan Sterne, Ph.D.
University of Bristol
Bristol, UK

Jonathan Treadwell, Ph.D.
ECRI
Plymouth Meeting, PA

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers follows:

Ethan Balk, M.D., M.P.H.
Tufts Center for Clinical Evidence Synthesis
Institute for Clinical Research and Health Policy Studies
Tufts Medical Center
Boston, MA

Gerald Gartlehner, M.D., M.P.H.
Danube University
Krems, Austria

Lisa Hartling, Ph.D.
University of Alberta
Edmonton, Alberta, Canada

Julian Higgins, Ph.D.
University of Bristol
Bristol, UK

Joseph Lau, M.D.
Brown University
Providence, RI

David Moher, Ph.D.
Ottawa Hospital Research Institute
Ottawa, Ontario, Canada

Susan L. Norris M.D., M.Sc., M.P.H.
Oregon Health and Science University
Portland, OR

Paul Shekelle, M.D., Ph.D.
RAND
Santa Monica, CA

Jonathan Sterne, Ph.D.
University of Bristol
Bristol, UK

Jonathan Treadwell, Ph.D.
ECRI
Plymouth Meeting, PA

The Empirical Evidence of Bias in Trials Measuring Treatment Differences

Structured Abstract

Objectives. To comprehensively and systematically review and compare empirical evaluations of specific types of bias on effect estimates in randomized controlled trials (RCTs) reported in systematic reviews.

Data sources. MEDLINE®, the Cochrane Library, and the Evidence-based Practice Center methods library located at the Scientific Resource Center. Additional studies were identified from reference lists and technical experts. We included meta-epidemiological studies (studies drawing from multiple meta-analyses), meta-analyses, and simulation studies (in relation to reporting bias only) intended primarily to examine the influence of bias on treatment effects in RCTs.

Review methods. Approaches to minimizing potential biases considered in the review included selection bias through randomization (sequence generation and allocation concealment); confounding through design or analysis; performance bias through fidelity to the protocol, avoidance of unintended interventions, patient or caregiver blinding and clinician or provider blinding; detection bias through outcome assessor and data analyst blinding and appropriate statistical methods; detection/performance bias through double blinding; attrition bias through intention-to-treat analysis or other approaches to accounting for dropouts; and reporting bias through complete reporting of all prespecified outcomes. Two people independently selected, extracted data from, and rated the quality of included studies. We did not pool the results quantitatively due to the heterogeneity of included studies.

Results. A total of 38 studies of trials (48 publications) met our inclusion criteria, from our review of 4,844 abstracts. Of these, 35 had usable evidence. Some studies concerned the effect of more than one type of bias on effect estimates. We reviewed 23 studies on allocation concealment, 14 studies on sequence generation, 2 studies on unspecified bias in randomization, 2 studies on confounding, 2 studies on fidelity to protocol and unintended interventions, 4 studies on patient and/or provider blinding, 8 studies on assessor blinding, 2 studies on appropriate statistical methods, 18 studies on double blinding, 15 studies on attrition bias, and 9 studies on selective outcome reporting.

Although a trend toward exaggeration of treatment effects was seen across bodies of evidence for most biases, the magnitude and precision of the effect varied widely across studies. We generally found evidence that was precise and consistent in direction of effect for assessor and double blinding, specifically in relation to subjective outcomes, and for selective outcome reporting. Evidence was generally consistent in direction of effect but with variable precision across studies for allocation concealment, sequence generation, and assessor blinding of objective or mixed outcomes. In contrast, evidence was generally inconsistent and imprecise in relation to confounding, adequate statistical methods, fidelity to the protocol, patient/provider blinding, and attrition bias.

Studies differed markedly on a number of dimensions including measures/scales used to measure biases, the thoroughness of reporting of trial conduct that was required, approaches to statistical modeling and adjustment for potential confounding, types of outcomes and stratification by treatment or condition. Within many epidemiological studies, the included meta-analyses or trials varied along these dimensions as well.

Conclusions. Theory suggests that bias in the conduct of studies would influence treatment effects. Our review found some evidence of this effect in relation to some aspects of RCT study conduct. When the bias was present, commonly the treatment effect was increased, but rarely were the estimates precise in the individual studies. However, because this evidence is limited and uncertain with respect to the magnitude of the impact, this does not necessarily imply that systematic reviewers can eliminate assessment of risk of bias. Due to the complexity of evaluating precision in meta-epidemiological studies developed from potentially heterogeneous meta-analyses or trials, we cannot be sure that studies were sufficiently powered. We suggest that systematic reviewers consider subgroup analyses, with and without studies with flaws in relation to specific biases of importance for review questions. Future studies evaluating the impact of biases on treatment effect should follow the lead of the BRANDO study and use modeling approaches that include careful construction of large datasets of trials (and eventually observational studies) designed to look at the effect of specific aspects of study conduct and the interrelationship between bias concerns.

Contents

Figures

Tables

Appendixes

Background

Epidemiological theory supports that there is a relationship between various types of bias in study conduct and the observed effect.[1] Bias may occur in selection of the sample, detection of treatment effects, performance of the study, and reporting of outcomes, each of which may increase the size of the observed effect. Because of this, systematic reviews evaluate the risk of bias of each included study. However, the empirical evidence supporting the theoretical relationships has been mixed or has not yet been tested. Prior reviews have attempted to summarize and compare the empirical evidence, but they are either outdated or not exhaustive. West et al. (2002),[2] focusing on systems to rate the quality (risk of bias) of studies, presented some empirical evidence concerning the relationship between bias and treatment effect, as did Deeks et al. (2003)[3] comparing randomized and nonrandomized studies and Song et al. (2000)[4] limited to publication and related biases. In reviewing the Cochrane Handbook, we found, that only empirical evaluations supporting concern about each of the potential sources of bias were cited, while studies that showed no relationship were not.[5] The recent BRANDO study (Bias in Randomized and Observational Studies) analyzed combined data from a large number of trials included in seven earlier methodological evaluations with the aim of further quantifying the effect of biases and other flaws in trial conduct and their effect on outcomes.[6] The empirical results from the BRANDO study are included in this review. BRANDO did not include a systematic review of the earlier evaluations that had similar or related goals (personal correspondence, Jonathan Sterne, 2012).

Key Questions

The goal of this review was to comprehensively and systematically review and compare empirical evaluations of the effect of specific types of bias on effect estimates of outcomes among studies reported in systematic reviews. The types of biases and approaches that trials may use to adjust for each bias are presented in Table 1; this classification draws primarily from the Cochrane Handbook for Systematic Reviews,[5] but also draws on material from other sources, including Hernan et al., (2004),[7] Rothman et al., (2003),[1] and Viswanathan and Berkman (2012).[8] They include the potential sources of bias that are commonly of concern among systematic reviewers who are evaluating the risk of bias in individual studies. We note that instruments designed to identify risk of bias in individual studies differ in their approach, including the specific questions used for measurement. We did not restrict this analysis to particular approaches to measurement of biases and we included all measures, specifying the exact question where provided.

Table 1. Potential sources of bias in studies (threats to internal validity) and approaches used to address specific biases

Type of Bias	Description	Approaches to Addressing the Bias
Selection bias at baseline including confounding	Systematic differences between baseline characteristics of the groups that are compared as a result of non-random enrolment in the study. Includes confounding that occurs when patient prognostic characteristics, such as disease severity or comorbidity, influence both treatment source and outcomes. Confounders are the common cause for intervention and exposure; they occur before exposure. Confounding can occur from self-selection of treatments or physician-directed selection of treatments.	1. Adequate sequence generation (randomization) 2. Allocation concealment 3. Confounding in RCTs typically addressed through randomization; in RCTs with randomization issues, confounding accounted for through analysis
Performance bias	Systematic differences in the care provided to participants and protocol deviation. Examples include contamination of the control group with the exposure or intervention, unbalanced provision of additional interventions or co-interventions, difference in co-interventions, and inadequate blinding of providers and participants.	1. Fidelity to the protocol 2. Avoidance of unintended interventions or co-interventions 3. Patient/caregiver blinding 4. Clinician/provider blinding
Detection bias	Systematic differences in outcomes assessment among groups being compared, including systematic misclassification of the exposure or intervention, covariates, or outcomes because of variable definitions and timing, diagnostic thresholds, recall from memory, inadequate assessor blinding, and faulty measurement techniques, including a lack of valid and reliable measures. Erroneous statistical analysis might also affect the validity of effect estimates.	1. Outcome assessor blinding 2. Data analyst blinding 3. Appropriate statistical methods
Attrition bias	Systematic differences in the loss of participants from the study and how they are accounted for in the results (e.g., incomplete followup, differential attrition). Bias resulting from incomplete outcomes assessment. Bias from missing data can result in selection bias during the course of the study.	1. Intention-to-treat analysis 2. Accounting for dropouts 3. Obtaining complete data
Reporting bias	Systematic differences between reported and unreported findings (e.g., differential reporting of outcomes or harms, incomplete reporting of prespecified study findings).	1. Complete reporting of all prespecified outcomes

Abbreviations: RCT = randomized controlled trials.

Our initial key questions were the following:

1. What sources of bias have been evaluated empirically for their impact on study effect estimates? Sources of bias include selection bias and confounding, performance bias, attrition bias, detection bias, and selective outcome reporting bias. For the purposes of this review, we do not include publication bias.
2. What evidence exists that these sources of bias influence effect estimates?

Input from a Technical Expert Panel advised the study team to further refine the analysis in several important ways. These refinements are reflected in revised key questions and include the following: limit the review to randomized controlled trials (RCTs) because we would be unlikely to find meta-analyses or meta-epidemiological studies focusing on examining risk of bias in observational studies; highlight evidence of the effect of one source of bias, adjusting for other biases and interactions between sources of bias; and distinguish between evidence that was obtained from meta-epidemiological studies and other designs.

Our revised key questions are as follows:

1. What sources of risk of bias in the design and conduct of RCTs have been evaluated empirically for their impact on study effect estimates?

We include the sources of bias that are most likely to be of concern in an RCT: selection, confounding, performance, detection, attrition, and outcome reporting.

2. What is the evidence that selection, confounding, performance, detection, attrition, and selective outcome reporting bias influence effect estimates? Findings will include the following:
    a. If results measuring one source of bias adjust for the effect of one or more additional biases or the interaction between sources of bias.
    b. If results differ by whether analyses were conducted using particular analytic approaches or when limited to specific types of diseases, interventions, or outcomes.

For consistency and clarity throughout the report, we refer to our analyses in this report as a review, the meta-epidemiological or other empirical publications that we reviewed as studies, and the individual RCTs on which they were based as trials.

Methods

Literature Search Strategy

We conducted focused searches of MEDLINE (via PubMed), the Cochrane methods group library, and the Evidence-based Practice Center (EPC) methods library located at the Agency for Healthcare Research and Quality AHRQ's Effective Health Care Program's Scientific Resource Center. We also asked members of our Technical Expert Panel and others knowledgeable of this body of work to identify relevant studies. An experienced research librarian used a predefined list of search terms and medical subject headings (MeSH). The librarian completed the MEDLINE and Cochrane methods library search on January 8, 2012, and an update search on September 24, 2012. The EPC methods library search was completed on December 5, 2011. We limited the search to studies published in English, based on limited resources. The complete search strategies for the MEDLINE and Cochrane methods library searches, including specific limitations used for each database, are presented in Appendix A. The EPC methods library search used the term "bias" with "effect," "estimate," "empiric," and/or "evidence".

Inclusion and Exclusion Criteria

The study's inclusion and exclusion criteria are presented in Table 2. We excluded publication bias because it is not a threat to internal validity and it has been studied extensively in earlier reviews.[9-14] We focused on studies that were primarily methodological and intended to evaluate the relationship between biases in study design and/or conduct and effect estimates of outcomes by comparing treatment effects in randomized controlled trials (RCTs) that implemented an approach to minimize bias (i.e., allocation concealment, blinding) with those that did not. We excluded sensitivity analyses occurring within a single systematic review or meta-analysis that was primarily evaluating treatment effects.

We did not impose any inclusion restrictions on the manner in which the evaluation of the impact of a bias was defined or measured by a study author. For example, we included any evaluative definition of appropriate statistical methods or intention-to-treat analysis and noted the definition used in each study. Similarly, we included any approach to measurement (continuous or categorical) or formulation of the comparison group (not adjusted, not reported, or both).

Meta-epidemiological studies are commonly conducted to answer the methodological questions of interest in this review. A meta-epidemiological study analyzes a collection of meta-analyses, in each of which the component studies have been classified according to some study-level characteristic.[15, 16] To include the largest number of informative studies, we did not require that included meta-epidemiological studies conduct any, or any particular type, of statistical analysis of the included meta-analyses. These studies are sometimes called meta-meta-analyses.

When describing a meta-epidemiological study, we considered the target population of meta-analyses, how the study meta-analyses were chosen, and how the number of studies that were included was justified and described. The target population refers to the population of meta-analyses from which the study meta-analyses were taken. For example, the target population might be "all meta-analyses in Cochrane Library between data X and date Y that had at least one RCT." The target population may be representative of a certain clinical condition only, for example, meta-analyses of back pain. We also noted the list, or to use sampling language, the sampling frame from which the meta-analyses were drawn. We noted the inclusion and exclusion

criteria, if appropriate. For example, only meta-analyses that contained at least one RCT were included. Finally, the target population of meta-analyses was described, either a census, that is, all meta-analyses within the population, or a random or convenience sample of meta-analyses chosen from the population.

Table 2. Study inclusion and exclusion criteria

Category	Inclusion	Exclusion
Population, intervention, comparators, and length of followup, setting	• All populations, length of followup, and setting	• Medical tests (screening, diagnosis, genetic tests) • In vitro studies
Time period	• Published 1980 onward	• Published before 1980
Study design	• Meta-epidemiology study of randomized controlled trials • Simulation studies for outcome reporting bias only	• Meta-epidemiology studies of observational studies
Outcomes	• Differences in treatment effect size and directions • Difference in significance of effect	• Estimates of heterogeneity
Publication language	• English	• Non-English languages
Admissible evidence for on patient-level, provider-level, or systems-level interventions (study design and other criteria)	• Meta-analyses of data from systematic reviews that were designed primarily to examine the effect of biases on results • Meta-epidemiological studies • Simulation studies for outcome reporting bias only	• Primary data • Simulation studies of the effect of biases other than outcome reporting bias • Narrative reviews • Studies not designed primarily to look at the effect of a bias on the size or direction of effect, such as sensitivity analyses included in systematic reviews of treatment effects. • Studies of the potential effect of other types of biases not related to study conduct (for example, the presence of dampness vs. nondampness in asthma studies, adjusted vs. nonadjusted meta-analyses and the impact on the summary estimate)
Biases	• Selection, detection, performance, attrition, and outcome reporting. Table 1 includes the specific approaches to adjusting for the bias that are examined in studies.	• Publication bias (including time-lag bias, multiple citation bias, language bias) • Other possible sources of biases such as single versus multicenter, country of origin

We considered several aspects of the meta-epidemiological studies when assessing the possible influence of bias on the size and direction of the treatment effect. These aspects are the design of the meta-epidemiological study, particularly how the component meta-analyses were gathered and the resulting generalizability of the study; the measurement approach the study used to assess the biases within the trials; and the analytic approach in the study, such as whether the study fit a regression model to evaluate the relationship between the bias and treatment effect and whether the regression model adjusted for other factors (such as clinical condition).

Many meta-epidemiological studies in our systematic review fit a regression model to examine the relationship between an approach to adjusting a bias component of interest listed in Table 1 (e.g., blinded versus unblinded) and treatment effect. The descriptions of these models varied across the studies. Therefore, we now define terms that we used to classify different

models based on how the meta-epidemiological study reported how its analysis was conducted. We note that in some cases, analytic approaches are comparable. If the meta-epidemiological study—

- Does not fit a regression model, but rather pools treatment effects (e.g., odds ratios) separately for different subgroups of individual trials, we call this a **stratified by bias component analysis**. For example, a meta-epidemiological study may pool odds ratios for all trials that are blinded and separately pool odds ratios for all trials that are not blinded. If the analysis is stratified by clinical condition, we call this a **stratified by clinical condition analysis.** For example, a meta-epidemiological study may pool odds ratios for all surgical trials and separately pool odds ratios for all nonsurgical trials. Both of these are stratified rather than regression approaches. We note that this type of stratified approach is comparable to a regression approach in which one interacts the stratum variable, e.g., clinical conduction, with the other terms in a regression model. If the analysis fits a hierarchical regression model in which the first level of hierarchy is trial, and the second is meta-analysis, we call this a **hierarchical model**. Otherwise, if the model is not hierarchical, we call this a **regression model.**
- Fits a regression model that includes the bias component of interest and other bias components, thereby adjusting for the effects of other biases, we say this model **adjusts for other biases**, that is includes multiple sources of bias. Otherwise, we say this model **does not adjust for other biases**. We note that the biases we considered in this assessment are those that we have chosen to study in this systematic review (Table 1); so, for example, publication bias is not included in this set of biases.
- Fits a regression model that includes other factors not among the set of biases we are studying, we say this model **adjusts for other factors**; otherwise, we say this model **does not adjust for other factors**. We note that the choice of other factors can be quite extensive (for example, types of interventions or diseases or biases that are not being studied as part of this systematic review), because it depends on what the various meta-epidemiological studies report. These factors may or may not be plausible sources of bias or modifiers of the effects of the bias components that we are interested in.

As an example of the language described above, if the authors fit a hierarchical regression model that included a single bias component (such as concealment of allocation) as an independent variable, then we describe this model of the effect of concealment of allocation as "hierarchical regression; does not adjust for other biases; does not adjust for other factors."

We note that meta-epidemiological studies which stratify by clinical condition, or which adjust for clinical condition as another factor, produce a summary effect for each clinical topic. Conversely, meta-epidemiological studies which do not stratify by clinical condition, or which do not adjust for clinical condition as another factor, give us an average effect across clinical topics.

A discussion of modeling approaches is available in Sterne et al.[16] This paper describes the theoretical basis for modeling decisions, such as the manner in which the errors are estimated; such distinctions are beyond the scope of our review. We also note that Odgaard-Jensen et al.[17] further categorized meta-epidemiological studies as to whether they adjusted for clinical differences in the participants and in the interventions.

We did not conduct our own meta-analysis of the meta-epidemiological studies. This decision was based on the dissimilarity in modeling approaches across the meta-epidemiological studies and practical constraints.

In relation to our presentation of results from the included studies, we use the term "treatment effect" to mean any statistic that measures the treatment effect, e.g., a mean difference (MD), a standardized mean difference (SMD), or a dichotomous outcome measure such as a ratio of odds ratios (ROR).

In relation to our presentation of continuous outcomes, we use the Cochrane Collaboration Glossary (http://www.cochrane.org/glossary) to define a MD to be equal to the estimated mean outcome for the treatment group minus the estimated mean outcome for the control group. We define a SMD to be equal to the mean difference divided by an estimate of the standard deviation.

In studies, a MD may be called a weighted mean difference, as MDs are combined meta-analytically using weights proportional to the precision of the estimate. Sometimes a SMD is called an effect size, a measure without units. Occasionally a MD is also called an effect size. We examined each study to determine whether a MD or a SMD had been estimated using the Cochrane definitions. For greater clarity, we generally do not use the term "effect size" in this document.

Study Selection

Two trained members of the research team independently reviewed all titles and abstracts identified through searches for eligibility against predefined inclusion and exclusion criteria. Studies marked for possible inclusion by either reviewer underwent a full-text review. Each full-text article was again independently reviewed by two trained members of the team to determine if it met inclusion criteria. If it did not meet inclusion criteria, each reviewer recorded the reason for exclusion and they later resolved the disagreement by consensus discussion. The reviewers consulted a third party if they were unable to reach a consensus. The full-text review form reviewers used is reproduced in Appendix B. Forms were created and accessed using Web-based systematic review software (DistillerSR, Evidence Partners, Ottawa, Canada). If both reviewers agreed that a study did not meet the eligibility criteria, it was excluded. If the reviewers disagreed, they resolved conflicts by discussion and consensus or by consulting a third member of the review team. The project coordinator tracked results of the abstract and full-text reviews in an EndNote database. Appendix C contains a complete list of studies excluded during the full-text review, denoted by their primary reason for exclusion.

Data Extraction

We developed an evidence table template for data synthesis. For studies that met inclusion criteria, we abstracted relevant information into these evidence tables using Microsoft Excel. We abstracted characteristics of the study including sources of bias examined, outcomes, and inclusion/exclusion criteria and results, including how a bias was measured, statistical approach to measuring differences in outcomes (e.g., odds ratios, RORs), numeric findings, and whether the study adjusted for other biases or other factors. One trained reviewer initially abstracted the relevant data from each included study, and a second member of the team reviewed each data abstraction against the original article for completeness and accuracy.

Quality Assessment

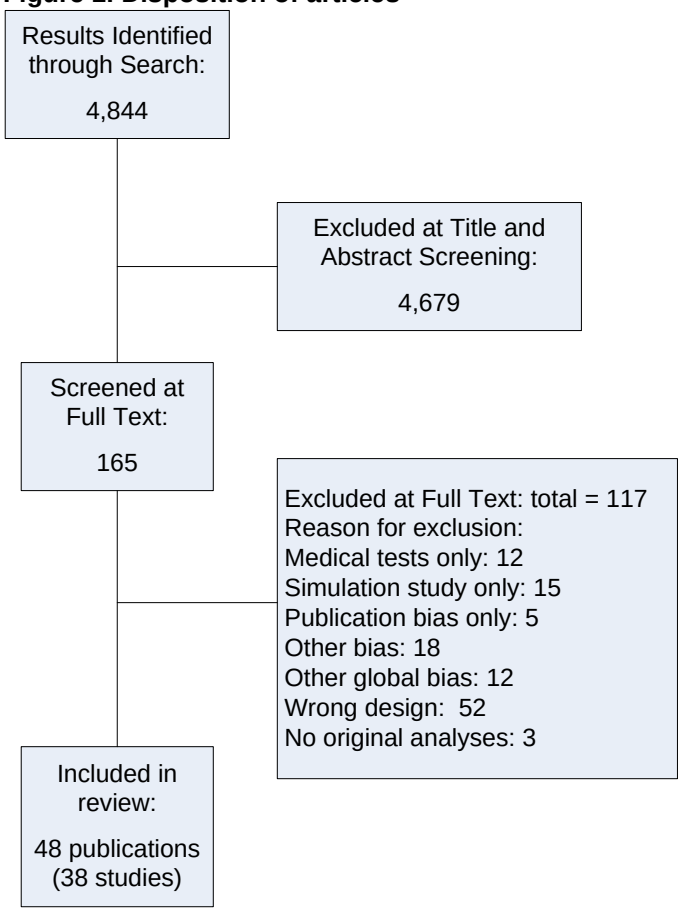
We developed a priori nine criteria by consensus to assess studies for potential design flaws because no standardized checklist exists. In part, our choice of questions for systematic reviews was based on AMSTAR.[18] For each included study, two independent reviewers assessed the potential for bias: selection, performance, attrition, detection, and reporting. Disagreements between the two reviewers were resolved by discussion and consensus or by consulting a third member of the team. Our nine criteria included the following:

1. Were studies selected by a census? If no, how was the sample selected? Was the sample a census or not?
2. Did the study account for interaction among sources of bias?
3. Did the study use dichotomous measures for high versus low quality?
4. If the study used dichotomous measures for high versus low quality, how was the threshold determined?
5. What was the interrater reliability of risk of bias measure?
6. Was the risk of bias measure valid?
7. How was sample size calculated?
8. Do the findings of this meta-epidemiological study apply to multiple clinical areas?
9. Did the study account for duplication of trials (from different meta-analyses)?

Results

In the sections below, we present an overview of our included studies followed by results for each source of bias for which we found empirical evidence. Following a listing and description of the number of eligible studies (Figure 1), we summarize the characteristics of included studies (Table 3), including the approach used to minimize the potential effect of a bias that is evaluated in each study (i.e., blinding, allocation concealment), conditions examined and outcomes measured, the analysis design and data sources, and we note whether the study included stratified analyses or adjustments for other biases or other factors.

Figure 1. Disposition of articles



We provide a more detailed description of each included study in two tables included in Appendix D. Data we present in appendix tables include inclusion and exclusion criteria, unit of analysis (meta-analysis or trial), and sources for locating trials (or meta-analyses) used as data for the analysis, the interventions, comparators, populations, clinical conditions and outcomes of trials included in studies, and adjustments to the analyses.

The results for each source of bias are discussed in the text, with outcomes from each study included in the evidence summarized in a corresponding table (Tables 4 through 14).

Appendix E offers an assessment of the quality for each included study.

Results of Literature Searches

We identified 4,844 citations from searches and reviews of reference lists. Figure 1 documents the disposition of the 165 articles retrieved for full-text review for this report. We excluded 117 full-text articles for various reasons, including wrong design (narrative reviews that did not measure the effect of the bias on outcomes, studies not designed primarily to look at the effect of a bias on the size or direction of effect, such as sensitivity analyses included in systematic reviews of treatment effects; or sensitivity analyses to assess nonmethodological biases), a focus on medical tests, and biases not being considered as part of the review. These included publication bias, other biases that can occur when studies are conceived and performed (e.g., sponsorship, language, single versus multisite studies, studies stopped early for benefit, clinical specialty of the review team, and nonrandomized studies only), and other global biases

that refer to atypical methodological biases and biases that can occur in analyzing the study results (e.g., different methods used to measure the summary estimate and evaluating noninferiority trials). Articles excluded during full-text review are listed in Appendix C with reasons for exclusions. The review includes 38 studies, reported in 48 publications. Three of the included studies, while meeting our inclusion criteria, did not contain any usual data for our analyses and are not discussed further.[19-21]

Table 3. Characteristics of included studies

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Schultz et al., 1995[22]	Selection • Allocation concealment • Randomization Attrition • Exclusions after randomization Detection/Performance • Double-blinding (source not necessarily identified: could be identical/ active placebo)	Pregnancy, child birth, and early neonatal period Binary outcomes varied across studies; specific outcomes: NR	Level of analysis: MA (N=33 [250 RCTs]) Outcome: ROR Model: Logistic regression, adjusting for nesting in MA and other biases Data: Cochrane Pregnancy and Childbirth Database	Stratification: No, but assembled homogenous groups of interventions Adjusted for other 3 biases examined. Also: tx group, trial, variation across MAs
Juni et al.,1999[23]	Selection • Allocation concealment Detection • Assessor blinding Attrition • Handling of drop-outs and withdrawals (ITT)	Condition: General surgery Binary outcomes, thromboembolitic events (bleeding and DVT)	Level of analysis: RCT (N=17 included in 1 MA) Outcome: RRR Model: Fixed effects MA Data: MA by Nurmohamed et al.[24]	Stratification: No, all trials included same intervention and outcome No adjusting for other biases or other factors
Linde 1999[25]	Selection • Randomization (explicitly stated) • Allocation concealment Detection/Performance • Double-blinding (Patient and assessor) Attrition • Complete followup	Conditions: Homeopathic interventions for tx or prevention Binary outcomes varied across studies; specific outcomes: NR	Level of analysis: RCT (N=89 included in 1 MA) Outcome: ROR Model: Multivariate Data: MEDLINE, EMBASE and CAM registries	Stratification: No Adjusted for other 3 biases examined

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Moher et al., 1999,[26] Moher et al., 1998[27]	Selection • Randomization • Allocation concealment Detection/Performance • Double blinding (source not always identified: could be identical/active placebo) Attrition • Adequate followup	Conditions: Digestive and circulatory diseases, mental health, stroke pregnancy and childbirth Binary outcomes varied across studies; specific outcomes: NR	Level of analysis: MA (N=11 selected randomly [127 RCTs]) Outcome: ROR Model: Logistic regression Data: Cochrane Database	Stratification: No No adjusting for other 3 biases Adjusting for: intervention effect, trial indicators, variation across MAs
Kjaergard, Villumsen and Gluud, 2001[28, 29]	Selection • Randomization • Allocation concealment Detection/Performance • Double blinding (identical placebo tablets or similar) Attrition • Adequate followup	Conditions: NR Binary outcomes: mortality, neonatal mortality, cesarean section, DVT, dropouts, endocervical cells, resumed smoking	Level of analysis: MA (N=14 [190 RCTs, 23 large, 167 small]) Outcome: ROR Model: Logistic regression Data: Cochrane Database and MEDLINE	Stratification: By small vs. large RCT ≥ 1,000 participants No adjusting for other 3 biases Adjusting for: intervention effect, trial indicators, variation across MA
Balk et al., 2002[30, 31]	Selection • Randomization • Allocation concealment Detection • Assessor blinding • Valid statistical methods Performance • Patient blinding • Provider blinding • Caregiver blinding Detection/Performance • Double blinding Attrition • Intention-to-treat analysis • Dropouts recorded Detection/Performance • Double blinding (source not identified)	Outcomes: mortality in cardiovascular disease studies; varied in other clinical areas. If multiple outcomes, included those examined by the largest number of studies or those most clearly defined.	Level of analysis: MA (N=26 [256 RCTs]) Of these: Cardiovascular: (N=8 [93 RCTs]) Infectious disease: (N=6 [56 RCTs]) Pediatrics: N=5 [60 RCTs]) Surgery:(N=7 [67 RCTs]) Outcome: ROR Model: Baysian meta-regression with random effects, adjusting for clustering within MAs and variability across MAs Data: MEDLINE and Cochrane databases	Stratification: 4 disease areas Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Clifford et al., 2002[32]	Selection • Allocation concealment	Outcomes: Various conditions NS	Level of analysis: RCT (N=100) Outcome: OR Model: ANOVA Data: 1 year of issues of 5 selected high-impact medical journals	Stratification: direction of RCT outcome: (favored new intervention; favored conventional; neutral; unclear) Adjusting for other biases: Funding source
Sterne et al., 2002[16]	Selection • Allocation concealment Detection/Performance • Double-blinding (source not always identified: could be identical/ active placebo)	Same as Schulz et al., 1995[22]	Level of analysis: MA (N=33 [250 RCTs]) Outcome: ROR Models: comparison of logistic regression, fixed and random-effects MA Data: Same as Schultz et al.[22]	Stratification: No No adjusting for other biases Adjusting for variation across studies and MA differed by model type
Als-Nielsen et al., 2003[33]	Detection/Performance • Double-blinding (described as double blind)	Varied/mix of conditions across studies Binary outcomes varied across studies; specific outcomes: NR	Level of analysis: RCT (N=370 [25 MA]) Outcome: OR Model: logistic regression MA Data: Cochrane Database	Stratification: Disease area, and type of intervention Adjusting for other biases: Funding type
Egger et al., 2003[34, 35]	Selection • Allocation concealment Detection/Performance • Double-blinding (described as double blind or included assessor blinding)	Infectious diseases, neurology, obstetrics and gynecology, other Binary outcomes varied across studies; specific outcomes: NR	Level of analysis: MA Allocation concealment: (39 [304 RCTs]; Blinding: (45 [399 RCTs]) Outcome: ROR Model: Meta-regression Data: Cochrane Database; *Health Technology Assessment* and 8 medical journals that regularly publish SRs	Stratification: Yes, by Intervention: drug vs. nondrug, active vs. nonactive control, complementary vs. conventional medicine Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Chan et al., 2004[36]	Reporting • Selective outcome	Cardiology, obstetrics and gynecology, surgery, and pediatrics. Efficacy and harms (binary or continuous: NR), specific outcomes: NR	Level of analysis: RCT (N=48 RCTs [1,233 efficacy outcomes] and N=26 RCTs [169 harms outcomes]) Outcome: OR Model: Random-effects MA Data: PubMed, EMBASE and Cochrane Trials Register	Stratification: Efficacy, harms; fully or partially reported vs. qualitatively reported or unreported Adjusting for other biases: No
Chan et al., 2004[37]	Reporting • Selective outcome	Various conditions NS Efficacy and harms (binary, continuous or survival data), specific outcomes: NR	Level of analysis: RCT (N=50 [2,175 efficacy outcomes and 605 harms outcomes] Outcome: OR Model: Random-effects MA Data: Clinical studies approved by Scientific-Ethnical committees for Copenhagen and Frederiksberg, Denmark	Stratification: Efficacy, harms; fully or partially reported vs. qualitatively reported or unreported Adjusting for other biases: No
Kyzas et al., 2005[38]	Detection/Performance • Blinding (stated as blinded-level not specified) Reporting • Outcome reporting	Head and neck squamous cell cancer Binary outcomes: all-cause mortality and lymph node status	Level of analysis: RCT (N=42) Outcome: RR Model: random and fixed effects MA Data: PubMed and EMBASE	Stratification: No Adjusting for other biases: No
Tierney et al., 2004[39]	Attrition • Intention-to-treat analysis	Cancers of the bladder, brain, lung, esophagus, ovary, lung, and soft tissue sarcoma Hazard of survival and recurrence	Level of analysis: MA and RCT level (N=14 MA & N=92 RCTs with at least one patient exclusion (21,905 patients) Outcome: HR Model: Fixed effects MA Data: SRs with MAs of patient-level RCT data addressing cancer therapies	Stratification: Disease condition Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Derry et al., 2006[40]	Selection • Randomization Detection/Performance • Double blinding (patient and assessor-practitioner [acupuncturist] could not be blinded)	Stroke, various painful conditions, nausea and vomiting, depression other conditions (insomnia, smoking cessation, weight loss, asthma) Binary measures: varied across studies	Level of analysis: RCT (N=35) Outcome: Relative risk Model: Random effects MA Data: PubMed, AMED, Cochrane Database	Stratification: sample size and control event rate by outcome (nausea, vomiting, antiemetic consumption Adjusting for other biases: Randomized and blinded combined
Furukawa et al., 2007[41]	Reporting • Selective outcome reporting	Conditions: Various conditions NS Measures: 2 binary and 2 continuous of greatest patient importance	Level of analysis: RCT (N=156) Outcome: OR and SMD Model: Linear regression Data: Cochrane database	Stratification: No Adjusting for other biases: No
Pildal et al., 2007[42]	Selection • Allocation concealment Detection/Performance • Double-blinding (source not necessarily identified: could be identical/ active placebo, could be patient and provider)	Varied conditions covered in 6 earlier meta-epi studies (Schulz et al., 1995,[22] Moher et al., 1998,[27] Kjaergard, Villumsen and Gluud, 2001,[28] Egger et al., 2003,[34] Balk et al., 2002,[30] Als-Nielson, 2004[43]) Binary measures varied across studies	Level of analysis: MA Allocation concealment: (N=34 [283 RCTs]), Blinding: (N=20 [182 RCTs]) Outcome: ROR Model: Random effects meta regression Data: Cochrane Database, PubMed	Stratification: No Adjusting for other biases: No
Siersma et al., 2007[13]	Selection • Sequence generation • Allocation concealment Detection/Performance • Double-blinding Attrition • Intention-to-treat	Conditions: Various NS Measures: Primary binary, varied across studies	Level of analysis: RCT (N=523 [48 MA]) Outcome: ROR Model: Logistic regression (with and without random effects), Weighted regression (with and without random effects) Data: Cochrane database	Stratification: No Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Fenwick et al., 2008[44]	Selection • Allocation concealment Detection • Assessor blinding	Periodontology Continuous measures: probing depth, and clinical or probing attachment level	Level of analysis: RCT (N=50, [allocation concealment, N=34; blinding, N=33]) Outcome: MD Model: Random effects meta-regression Data: Cochrane Database of SRs	Stratification: Outcomes of Probing depth, CAL/PAL Adjusting for other biases: No
Wood et al., 2008[45]	Selection • Allocation concealment Detection/Performance • Double-blinding (source not necessarily identified: could be identical/ active placebo, could be patient and provider)	Varied conditions covered in 3 earlier meta-epi studies (Schulz 1995,[22] Kjaergard, Villumsen and Gluud, 2001,[28] Egger et al., 2003[34]) Binary measures varied across studies-compared objectively vs. subjectively assessed outcomes	Level of analysis: MA Allocation concealment (N=102 [804 RCTs]) Blinding (N=76 [746 RCTs]) Outcome: ROR Model: Logistic regression and random effects meta regression Data: Cochrane Database, PubMed	Stratification: mortality vs. other outcomes; objective vs. subjective outcomes; drug vs. other interventions Adjusting for other biases: Yes
Hartling et al., 2009[46], additional information :personal correspon-dence, Lisa Hartling, 2013	Selection • Randomization (sequence generation and allocation concealment) Detection/Performance • Blinding	Conditions related to pediatric health Outcomes: Primary but NS	Level of analysis: RCT (N=163) Outcome: SMD Model: Multivariate meta-regression Data: Manuscripts resulting from Society for Pediatric Research meetings between 1992 and 1995	Stratification: Outcome type (binary vs. continuous, objective vs. subjective) Adjusting for other biases: Study type (efficacy vs. equivalence), study design (crossover vs. factorial or parallel)

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Inaba et al., 2009[47]	Selection • Randomization (sequence generation and allocation concealment) Detection • Assessor blinding Attrition • Disclosure of withdrawals and dropouts	Acute myocardial infarction Binary outcomes: mortality, Continuous outcomes: impaired myocardial blush grade, ST-segment resolution	Level of analysis: RCT (N=25) Outcome: MD Model: DeSimonion and Laird random effects model, Meta-regression analysis (univariate and multivariate) Data: Cochrane Database, PubMed, hand searches	Stratification: study size Adjusting for other biases: single or multicenter design, type of devices used, study regions, presence of conflicts of interest
Nuesch et al., 2009[48, 49] Nuesch et al., 2009[48, 49]	Selection • Allocation concealment Performance • Patient blinding Attrition • Excluding patients from the analysis	Pain from osteoarthritis of the knee or hip Nonbinary subjective outcomes from interventions (Patient reported pain)	Level of analysis: MA Concealment: (N=14 [158 trials]) Blinding: (N=10 [122 trials]) Outcome: SMD Model: Logistic regression Data: Cochrane, PubMed, EMBASE and CINAHL, last update: 11/2007	Stratification: difference in treatment effect, small vs. large treatment benefits, high vs. low between-trial heterogeneity, nonpharmacologic vs. pharmacologic interventions, complementary vs. conventional medicine Adjusting for other biases: allocation concealment, intention-to-treat

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
van Tulder et al. 2009[50]	Selection • Allocation concealment • Randomization • Similarity of groups at baseline Detection • Assessor blinding • Timing of outcome assessment Performance • Patient blinding • Care provider blinding • Effect of co-intervention differential compliance Attrition • Drop-out rate • ITT	Nonspecific low back pain Continuous and binary measures: Pain, function, or similar improvement	Level of analysis: RCT (N=216) Outcome: SMD Model: Random effects meta-regression Data: Cochrane Library 2005: all Back Review Group reviews of nonsurgical tx for nonspecific low back pain.	Stratification: No Adjusting for other biases: No
Dwan et al., 2010[51]	Reporting • Selective outcome reporting	Acute asthma Outcome type: NS, Pulmonary function tests, including peak expiratory flow rate (PEFR) and forced expiratory volume in one second (FEV1), and hospital admission	Level of analysis: RCT (N=24) Outcome: RR Model: sensitivity analysis Sources: All RCTs included in the SR 'Intravenous and nebulized magnesium sulfate for acute asthma'	Stratification: Intervention and outcome Adjusting for other biases: No
Hamm et al, 2010,[52] additional information: personal correspon-dence, Lisa Hartling, 2013	Selection: • Allocation concealment • Randomization Detection/Performance • Double-blinding (source not identified) Attrition • Drop-out rate Reporting • Selective outcome reporting	Unspecified Continuous and binary: Unspecified (primary trial outcome)	Level of analysis: RCT (N=236) Outcome: SMD when continuous & converted OR when binary Data: Pediatric RCTs published in Cochrane Registry in 2007	Stratification: No Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Kirkham et al., 2010[53]	Reporting • Selective outcome reporting	Hepato-biliary, pregnancy and childbirth, neonatal, oral health, menstrual disorders and subfertility Outcomes not reported	Level of analysis: SR (N=283 [2486 RCTs]) Outcome: NR Model: Sensitivity analysis Data: Issue of the Cochrane library	Stratification: No Adjusting for other biases: No
Hartling et al., 2011[54]	Selection • Randomization (sequence generation and allocation concealment) Detection/Performance • Blinding (source not identified) Attrition • Incomplete data Reporting • Selective outcome reporting	Persistent asthma Continuous: Forced expiratory volume in 1 second (FEV_1)	Level of analysis: RCT (N = 107) Outcome: MD Model: Random effects MA Data: RCTs included in a SR	Stratification: No Adjusting for other biases: No
Hempel et al., 2012 [35], 2011[55]	Selection • Randomization (sequence generation and allocation concealment) Detection • Assessor blinding Performance • Patient blinding • Care provider blinding • Similar co-interventions/ compliance/timing Attrition • Drop-out rate • ITT	Variety of conditions including back pain, digestive diseases, circulatory diseases, mental health, stroke, and pregnancy and childbirth. Variety of continuous and categorical outcomes measured as absolute treatment effect sizes, standardized treatment effect sizes and odds ratios	Level of analysis: RCT (N=600) Outcome: SMD, ROR Models: Random and fixed effects MA Data: Four large datasets that included RCTs that had been used in a variety of earlier MAs	Stratification: size of the treatment effect, condition, type of outcome Adjusting for other biases: No
Herbison 2011[56, 57] Herbison 2006[56-58]	Selection • Allocation concealment	Condition: Not limited by condition Outcome: Any binary outcome	Level of analysis: MA (N =65 [389 RCTs]) Outcome: ROR Model: Random effects meta-regression Data: Cochrane library	Stratification: No Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Liu, LaValley and Latham, 2011[59]	Detection • Assessor blinding Attrition • ITT	Not limited by condition Continuous measure of lower limb muscle strength	Level of analysis: RCT (N=73) Outcome: SMD Model: Random effects MA Data: RCTs included in a recently published SR	Stratification: No Adjusting for other biases: ITT
Hartling et al., 2012 [24]	Selection • Randomization • Allocation concealment Detection/Performance • Double blinding Reporting • Selective outcome reporting Attrition • Incomplete data	Conditions: Varied across studies and included circulatory and respiratory health, nutrition, metabolism, and diabetes, musculoskeletal health and arthritis. Outcomes: Objective and subjective	Level of analysis: RCT (N=154) Outcomes: SMD Model: Logistic meta-regression Data: random sample from previous study by Hopewell et al (2010)	Stratification: Type of outcome (objective or subjective) Adjusting for other biases: No
Hrobjartsson et al., 2012 [60]	Detection • Assessor blinding	Wound/ulcer, fractured bone, angina pectoris, facial folds, other Binary, varied by study, mostly subjective, such as patient function	Level of analysis: RCT (N=21) Outcome: ROR Model: Random effects meta-analysis Data: PubMed, EMBASE, PsychINFO, Cochrane library, High Wire Press, Google Scholar	Stratification: clinical problem, whether study arms were same type of procedure Adjusting for other biases: Funding, whether blinding procedure considered effective, whether patients seen by one or two assessors
Mhaskar et al., 2012 [61]	Reporting • Methodological quality	Conditions: Cancer Outcomes: Survival	Level of analysis: RCT (N = 429) Outcome: Ratio of HR Model: Univariate Data: All NCI Cooperative Group trials between 1968 and 2006 with protocols and publications available with unique RCTs	Stratification: Publication vs. protocol plus publication Adjusting for other biases: No

Table 3. Characteristics of included studies (continued)

Study	Bias Examined in Study	Conditions Evaluated Outcomes Reported/How measured	Analysis Design Data Source	Analysis Stratification by Disease, Condition, Trial Size Adjust for Other Biases or Other Factors
Savovic et al., 2012 [6]; Savovic et al., [62]; Savovic 2012 [62]	Selection • Allocation concealment • Randomization Detection/Performance • Double-blinding (source not necessarily identified)	Pregnancy and childbirth, mental and behavioral, circulatory system, digestive system, other factors, respiratory system, Other ICD-10, unclassified All-cause mortality, other objective, objectively measured but potentially influenced by clinician judgment, subjective, mixture of subjective and objective	Level of analysis: MA (N=234 [1973 RCTs]) Allocation concealment: (N=88 [811 RCTs]) Randomization: (N=104 [911 RCTs]) Blinding: (N=60 [592 RCTs]) Outcomes: ROR Model: Bayesian hierarchical bias model (allows for random intervention effects (between trial heterogeneity) within MA. Data: 7 earlier published meta-epidemiological studies	Stratification: outcome grouping (mortality, other objective, subjective or mixed) Adjusting for other biases: other two biases being evaluated

Abbreviations: AMED = Allied and Complementary Medicine Database; CAL/PAL = clinical or probing attachment level; CAM = complementary and alternative medicine; DVT = deep vein thrombosis; FEV1 = forced expiratory volume in 1 second; HR = hazard ratio; ITT = intention-to-treat; MA = meta-analysis; MD = mean difference; N = number; NCT = National Cancer Institute; NR = not reported; NS = ; OR = odds ratio; PEFR = peak expiratory flow rate; RCT = randomized controlled trial; ROR = ratio of odds ratios; RR = risk ratio; RRR = ratio of risk ratios; SMD = standardized mean difference; SR = systematic review; tx = treatment.

Overview of Included Studies

Among the 35 included studies that we include as evidence, the unit of analysis for 11 of the studies was meta-analyses,[6, 16, 22, 26-28, 30, 34, 42, 45, 48, 49, 56, 57, 62, 63] for 22 it was trials,[23, 25, 32, 33, 35-38, 40, 44, 47, 50-52, 54, 55, 59-61, 64-67] for one it was both meta-analyses and trials,[39] and for one it was systematic reviews.[53] Some of the studies included data from earlier studies. Sterne et al.[16] reanalyzed data included in the Schulz et al. study[22] based on additional model specifications. Wood et al.[45] combined data from three earlier studies included in this review—Schulz et al.,[22] Kjaergard et al.,[28] and Egger et al.[34] Pildal et al.[42] combined data from six earlier studies—Schulz et al.,[22] Moher et al.,[27] Kjaergard et al.,[28] Egger et al.,[34] Balk et al.,[30] and Als-Nielson et al.[43] Hempel et al. compared results among four datasets, two of which were based on studies included in other studies reported in this review—Moher et al.[27] and Balk et al.[30] Results of one of the datasets analyzed in Hempel et al.[35, 55] were also presented in Hamm et al.[52] The largest dataset was constructed by Savovic et al.,[6, 12, 62] combining data from seven earlier studies, including six included in this review: Schulz et al.,[22] Siersma et al.,[64] Balk et al.,[30] Egger et al.,[34] Kjaergard et al.,[28] and Pildal et al.[42]

While more than half of the studies were not limited to a particular condition or treatment approach or did not report on these characteristics in their included trials, 16 studies were limited

to trials in particular treatment areas or approaches to treatment. Specifically, Schulz et al.[22] and Sterne et al.[16] were limited to obstetric and neonatal trials; Juni et al.[23]was limited to use of heparin in surgery trials; Linde et al.[25] concerned homeopathy; Kyzas et al.,[38] Tierney,[39] and Mhaskar et al.[61] focused on cancer outcomes; Hartling et al.[66] was limited to pediatric trials; Nuesch et al.[48, 49] concerned pain outcomes after treatment for osteoarthritis; Hartling et al.[54] and Dwan et al.[51]were limited to asthma; van Tulder et al.[50] focused on treatment for low back pain; Clifford et al.[32] was limited to pharmaceutical interventions; Derry et al.[40] treatment with acupuncture; Fenwick et al.[44] periodontology; and Balk et al.[30]subgroup analyses limited to trials concerning cardiovascular disease, infectious disease, pediatrics, and surgery. Seven studies adjusted for the effect of other biases: Schulz et al.,[22] Linde et al.,[25] Wood et al.,[45] Nuesch et al.[48, 49], Liu et al.,[59] Hrobjartsson et al.,[60] and Savovic et al.[6, 62] Four additional studies adjusted for other study-level characteristics: Clifford et al.[32] funding source, Als-Nielsen et al.[33] funding source and treatment effect, and lastly, Inaba et al.[47] and Hartling et al.[66] adjusted for a number of factors other than the biases being examined in this review (Tables 3 and D-2). Each specific results section below provides additional detail about the characteristics of the studies included in the evidence, including statistical approach and adjustment for other variables.

Selection Bias: Allocation Concealment

Description of Included Studies

Study Characteristics

A total of 25 studies (31 publications) met our inclusion criteria for the evaluation of selection bias, operationalized as allocation concealment, random sequence generation, or generally described as randomization. Twenty three[5, 12, 13, 16, 22-28, 30, 32, 34, 42, 44-46, 48-50, 55-57, 61, 62, 68, 69] of 25 studies reported on allocation concealment. Two studies[40, 47] did not distinguish between sequence generation and allocation concealment but specified the bias as randomization; however, we include the data for these studies in the allocation concealment results (Table 4).

From these 25 studies evaluating the impact of inadequate allocation concealment, three studies (five publications) reevaluated trials within previously published empirical evaluations that included data from studies also eligible for inclusion in this systematic review. From these three meta-epidemiological publications, one study[55] contained bias ratings from two previously published datasets,[26, 27, 50] another study[45] from three datasets,[22, 28, 34] and the third study[12, 68] from five datasets[22, 28, 30, 34, 42] eligible in our systematic review. One study[16] reanalyzed the Schultz et al.[22] data to assess the effect of improvements in the statistical computational methods used to estimate the impact of allocation concealment bias.

The meta-analyses and trials included in the majority of the eligible meta-epidemiological studies were derived from specialized registries for systematic reviews (predominately Cochrane) or from traditional bibliographic databases (predominately MEDLINE and EMBASE). For all meta-epidemiological studies, including those that did not base their analyses on previous datasets, there was the potential for overlap of included meta-analyses or trials across the different studies (Table 3); overlap in the years of meta/analyses/trial publication, the bibliographic sources searched, the types of disease/population areas, and the datasets that were reanalyzed account for this potential inclusion of duplicate trials across studies.

Table 4. Summary of allocation concealment risk of bias results by study

Study Identification	Results of Allocation Concealment Bias	Impact on Treatment Effect + Increase - Decrease = No Difference
Studies with Odds Ratios (OR) or Ratio of Odds Ratios (ROR) or Ratio of Relative Risks = (RRR)		
Schulz et al., 1995[22]	ROR = 0.70 (0.62 to 0.79)[a/b] (p < 0.001)	+ (U)[a/b]
Juni et al.,1999[23]	RRR = 1.12 (0.76 to 1.65) [^] (p = 0.58)	= (I)[^]
Moher et al., 1998[26] companion Moher et al., 1999[27]	ROR = 1.11 (0.76 to 1.63)[c] (p = Not reported)	= (I)[c]
Linde et al., 1999[25]	ROR = 0.84 (0.60 to 1.18)[d] (p = Not reported)	= (I)[d]
Kjaergard, et al. 2001[28, 29]	Original ROR = 0.60 (0.31 to 1.15)[e] (p = 0.12) ROR = 0.48 (0.25 to 0.92)[f] (p = 0.027) Revised ROR = 0.82 (0.31 to 0.995)	Original = (I)[e] + (I)[f] Revised = (I)
Balk et al., 2002[30]	Overall ROR = 1.05 (0.91 to 1.21) CV ROR = 1.14 (0.96 to 1.42) ID ROR = 0.97 (0.68 to 1.42) PA ROR = 0.90 (0.58 to 1.28) SX ROR = 0.73, (0.36 to 1.24)	Overall = CV = (I) ID = (I) PA = (I) SX = (I)
Sterne et al., 2002[16]	ROR = 0.66 (0.55 to 0.78) p<0:001[J] ROR = 0.67 (0.57 to 0.78) p<0:001[k] ROR = 0.67 (0.61 to 0.82) p<0:001[l]	+ (I)[j/k/l]
Egger et al.,2003[34]	ROR 0.79 (0.70 to 0.89) p <0.001 Active control yes ROR = 0.71 (0.56 to 0.90) Active Control no ROR = 0.84 (0.74 to 0.97) p > 0.05 Drug intervention ROR = 0.79 (0.69 to 0.91) Non-drug intervention ROR = 0.79 (0.62 to 1.00) p > 0.05	+ (I +U) No statistical differences for active control or drug intervention.
Pildal et al., 2007[42]	ROR = 0.90 (95% CI: 0.81 to 1.01) p=0.08	= (I + U)
Siersma et al., 2007 [13]	ROR = 1.04 (0.90 to1.19)[f] (p = Not reported) ROR = 1.03 (0.90 to1.17)[g] (p = Not reported)	= (I)[f, g]
Wood et al., 2008[45]	Any outcome: ROR = 0.83 (0.74 to 0.93) (p^{Tol} N/A) Other outcomes: ROR = 0.76 (0.66 to 0.87) (p^{Tol} = 0.002) All-cause mortality: ROR = 1.01 (0.90 to 1.14) Other outcomes: ROR = 0.80 (0.73 to 0.87)* Subjective outcomes: ROR = 0.69 (0.59 to 0.82) (p^{Tol} = 0.009) Objective outcomes: ROR = 0.91 (0.80 to 1.03) Subjective outcomes: ROR = 0.75 (0.63 to 0.88)*	Any outcome: + (I + U) Other outcomes: + (I + U) Subjective outcomes: +(I+U)
Herbison 2011[57] Herbison 2006[56]	All group comparison: G1: Reference (central randomization) G2: ROR = 1.02 (0.85 to 1.22) G3: ROR = 0.87 (0.76 to 1.00) G4: ROR = 0.86 (0.78 to 0.96) G5: ROR = 0.76 (0.66 to 1.15) G6: ROR = 0.89 (0.70 to 1.15) Two group comparison: ROR = 0.91 (0.83 to 0.99)	+ (G4) = (G2/G3/G5/G6) + (I + U)

Table 4. Summary of allocation concealment risk of bias results by study (continued)

Study Identification	Results of Allocation Concealment Bias	Impact on Treatment Effect + Increase - Decrease = No Difference
Hempel et al., 2011[43, 70]	Dataset 3: Jadad and Schultz Schulz: concealment: ROR = 0.72 (0.46, 1.13)[m] Schulz: concealment: ROR = 0.77 (0.61, 0.97)[n] (p<0.05) Dataset 4: Heterogeneity set CBRG: ROR = 0.89 (0.73, 1.09) CBRG: ROR = 0.89 (0.66, 1.18) corrected for clustering	= (I + (U)[m] + (I + (U)[n] = (I + (U) = (I + (U)
Savovic et al., 2012[12, 62, 68]	All outcomes = ROR 0.89 (0.81 to 0.99)[o] -> Increase in BW trial SD = 0.06 and MA SD = 0.05 Mortality outcomes = ROR 1.03 (0.82 to 1.31[o] -> Increase in BW trial SD = 0.07 and MA SD = 0.07 Other objective outcomes = 0.92 (0.76 to 1.12)[o] -> Increase in BW trial SD = 0.06 and MA SD = 0.06 Subjective or mixed = 0.82 (0.70 to 0.94)[o] -> Increase in BW trial SD = 0.08 and MA SD = 0.08	All outcomes + (I)[l] Mortality = (I)[l] Objective = (I)[l] Subjective +(I)[l]
Mhaskar et al., 2012[61]	Publication: RHR = 0.94 (0.89, 0.99) Protocol plus publication: RHR = 0.95 (0.87, 1.04)	Publication: + (I) Protocol +(I)
Studies With Mean Differences in Treatment Effect (MD), Standardized Mean Difference in Treatment Effect (SMD), Standardized Mean Effect Size (SMES) or Relative Risk (RR)		
Fenwick et al., 2008[44]	Outcome of probing depth: adequate vs. MD = 0.22 (-0.58 to 1.03) p=0.59 (unclear) MD = 0.60 (-1.70 to 1.89) p=0.37 (inadequate) MD = 0.25 (-0.53 to 1.03) p=0.53 (unclear or inadequate) Outcome of clinical or probing attachment level (CAL/PAL): adequate vs. MD = 0.05 (-0.95 to 1.06) p=0.92 (unclear) MD = -0.09 (-2.0 to 1.08) p=0.93 (inadequate) MD = 0.10 (-1.01 to 1.04) p=0.98 (unclear or inadquate) (Positive MD in treatment effect indicates a tendency for a poorer quality study to obtain a larger treatment effect while a negative MD indicates a good quality studies obtain a greater treatment effect.)	Probing Depth: Both: =(I + U) CAL/PAL: Both: = (I + U)
Nuesch et al., 2009[48] Nuesch et al., 2009[49]	Adequate vs. not: (negative SMD indicates that trials with adequate concealment show a less beneficial effect) Overall SMD = -0.15 (95% CI, -0.31 to 0.02) Large vs. small treatment benefit: SMD = -0.79 (95% CI, -1.02 to -0.50) (p^{TOI} = <0.001) Drug vs. nondrug: SMD = -0.24 (95% CI, -0.53 to 0.04) (p^{TOI} = 0.26) Complementary vs. conventional: SMD = -0.52 (95% CI, -0.93 to -0.10) (p^{TOI} = 0.019)	Overall: = (I + U) Large vs. small treatment benefit + (I+U) Complimentary vs. conventional + (I + U)
Hempel et al., 2011[70]	Dataset 1 (Back Pain): Allocation concealment: SMD= -0.08 (-0.23 to 0.07) Dataset 2 (EPC set): Allocation concealment: SMD = -0.05 (-0.22 to 0.11)	= (I + U)
van Tulder et al., 2009[50]	SMD = -0.08 (-0.23, 0.07) (- value indicates that adequate concealment (higher quality) shows smaller effect)	= (I)

Table 4. Summary of allocation concealment risk of bias results by study (continued)

Study Identification	Results of Allocation Concealment Bias	Impact on Treatment Effect + Increase - Decrease = No Difference
Hamm et al, 2010,[52]	High risk of bias: SMD = 0.25 (-0.04 to 0.53) Unclear risk of bias: SMD = 0.39 (0.28 to -0.50) Low risk of bias: SMD = 0.38 (0.20 to -0.57) (CIs overlap with each other. High risk of bias had smaller SMD than low risk of bias.)	= = =
Hartling et al., 2009[46]	In comparison to adequate concealment, inadequate ($p < 0.678$) and unclear concealment ($p < 0.480$) results were not significantly different	=
Hartling, 2011[54]	Forced expiratory volume: High/unclear RoB vs. low RoB: MD = 0.05 (-0.03 to 0.12) Symptom-free days: High/unclear RoB vs. low RoB: MD = 0.55 (-4.48 to 5.59) High RoB vs. unclear and low RoB: MD Not estimable High RoB vs. unclear and low RoB: MD Not estimable	= (H + I) = (H + I)
Hartling et al., 2012[24]	High risk of bias: SMES = - 0.38 (-0.29 to 1.05) Unclear risk of bias: SMES = 0.77 (0.65 to 0.88) Low risk of bias: SMES = 0.49 (0.30 to 0.68) Overall meta-regression p value = 0.10 Stratified analysis: For allocation concealment, objective outcomes had higher RoB than subjective outcomes ($p = 0.007$).	= + +
Inaba et al., 2009[47]	Mortality Unclear randomization RR = 1.1 (0.62 to 1.99) Yes randomization RR = 0.66 (0.46 to 0.95) Incomplete ST resolution Unclear randomization RR = 0.67 (0.52 to 0.86); Yes randomization RR = 0.83 (0.73 to 0.93) Impaired Myocardial Blush Grade Unclear randomization RR = 1.1 (0.62 to1.99); Yes randomization RR = 0.66 (0.46 to 0.95)	Mortality: = (U) Incomplete ST: = (U) Myocardial Blush: = (U)
Studies with Other Metrics for Showing Impact of Allocation Concealment: Odds Ratio (OR) and Relative Risk (RR)		
Clifford et al., 2002[32]	Funding not for profit: OR = 0.55 (0.21 to 1.42) Funding mixed sources: OR = 1.35 (0.46 to 3.98) OR suggest that not for profit funding is protective against unclear rating of allocation concealment.	Not for profit protective
Derry et al., 2006[40]	Nausea: RR = 0.73 (0.57 to 0.93) Vomiting: RR = 0.71 (0.56 to 0.91) Antiemetic consumption: RR = 0.79 (0.61 to 1.02)	N/A

Notes: (+) = exaggerated impact from bias; (=) = no significant effect from bias, no significant difference between groups; (-) = reduced impact from bias, magnitude is decreased; (U) = relative to unclear allocation concealment; (I) = relative to inadequate allocation concealment

(G1) Trials that used some form of central randomization that clearly should hide the allocation (e.g., remote telephone service or randomization by a pharmacy).
(G2) Trials that used sealed envelopes with security enhancement (e.g., opaque and numbered); "Inadequate or unclear concealment" included.
(G3) Trials that used sealed envelopes without further details.
(G4) Trials that were reported as randomized without details, and also "double-blind."
(G5) Trials that simply said they were randomized with no further details.

(G6) Trials where the allocation was clearly not hidden (e.g., based on an open list, odd/even days of the week, participant's birth date, team member on duty at enrollment).
[a] Adjusting for sequence generation, blinding, and post allocation exclusions
[b] Excluded trials with inadequate concealment and compared with trials with unclear concealment
^ = Univariate analysis
c = Allowing for summary Odds Ratio (OR) to vary simultaneously according to the components (i.e. component by treatment interactions)
d = Adjusting for allocation concealment, blinding, post allocation exclusions and intent to treat.
e = Large vs small trials with adequate versus inadequate trials.
f = small trials with adequate versus inadequate trials
g = Logistic regression with random effect
h = Weighted regression with random effect
j = Logistic regression model
k = meta-analytic approach, within meta-analysis differences estimated using logistic regression
l = meta-analytic approach, within meta-analysis differences estimated using random effects meta- regression
m = random effects model
n = fixed effects model
o = adjusted for other biases
ToI = p value of test of interaction

Abbreviations: CBRG = Cochrane Back Review Group; CI = confidence interval; EPC = Evidence-based Practice Center; ID = infectious disease; MA = meta-analysis; N/A = not available; OR = odds ratio; PA = pediatrics; RHR = ratio of hazard ratio; RoB = risk of bias; ROR = ratio of odds ratio; RR = risk ratio; RRR = ratio of risk ratio; SMES = standardized mean effect size; SD = standard deviation; SX = surgery.

The year of publication of the included meta-analyses/trials in the meta-epidemiological studies may affect the ratings for presence or absence of allocation concealment bias, because reporting standards across time may differ. Approximately half of the meta-epidemiological studies captured and included trials published up to the year 2000,[30, 32, 57] 2002,[12, 42, 68] 2005,[40] 2006,[50, 55, 61] 2007,[5, 44, 48, 49] and 2009.[47] For the remaining meta-epidemiological studies, the included trials were published up to the late 1990s.

Definition of Allocation Concealment

Definitions of adequate concealment varied across studies (Appendix Table D-3), with the most comprehensive categorization for the presence, absence, or "unclear" category being specified by Schultz et al.;[22] this definition was used by five other studies.[16, 26, 27, 32, 42, 55] Seven studies based the definition from the Cochrane Handbook for the 2006 version of this tool[44] and the 2008 version.[5, 24, 46, 47, 54, 61] Three studies[28, 30, 48] used more complex definitions (relative to Schultz). One of these studies evaluated concealment based on any of 25 different quality assessment instruments.[23] A second meta-epidemiological study[45] that pooled trials from three previous studies[22, 28, 34] used the definitions provided within the original studies. Similarly, the third study[12, 62, 68] considered any definition of allocation concealment in the original datasets; however, in this study agreement between studies for trials overlapping across different meta-epidemiological datasets were compared (individual ratings of the included trials were dichotomized into adequate and inadequate allocation concealment). One study[56, 57] categorized various components of allocation concealment into six distinct categories; the first three partition out various aspects of concealing allocation; the latter three concern lack of adequate description or clearly inadequate concealment.

Four studies[22, 24, 44, 54] evaluated allocation concealment in three categories (adequate, inadequate, and unclear); the remaining studies typically grouped inadequate and unclear concealment into one category. However, one of these studies[22] removed trials with inadequate concealment ratings and presented results comparing adequate and unclear trials. Four studies[25-

[27, 55] used the Jadad scale to assess allocation concealment; however, the Jadad scale does not specifically query allocation concealment. As such, one study added additional items from the Internal Validity Scale,[25] which is not a validated scale, and the others added items from the Schultz definition.[26, 55, 71]

Two studies did not specify the criteria for establishing randomization (sequence generation and allocation concealment).[40, 47]

Population, Interventions, Comparators, and Outcomes of Included Studies

For many of the included meta-epidemiological studies evaluating allocation concealment, the types of population, interventions, comparators, or outcomes were not sufficiently specified; however, the titles of included meta-analyses or systematic reviews or trials included in the meta-epidemiological study were reviewed and, where possible, these were extracted (see Appendix D and Table 3).

Eight of the 25 studies evaluating allocation concealment included trials with specific patient or disease groups and these included trials restricted to periodontal disease,[44] osteoarthritis,[48, 49] acute myocardial infarction with adjunctive devices,[47] back pain,[50] cancer,[61] and pediatric populations.[5, 46] The remaining studies included a wide variety of patient and disease populations. Another study[30] compared four patient/disease categories that included cardiovascular, infectious diseases, pediatrics, surgical interventions, and an overall category combining all groups. The remaining meta-epidemiological studies included meta-analyses/trials with very heterogeneous populations and diseases.

Two studies limited the included trials to acupuncture[40] and pharmaceutical interventions.[32] Three studies[34, 45, 49] presented stratified results comparing different interventions that included drug versus nondrug interventions and estimated the impact of inadequate allocation concealment. One study[49] presented stratified analyses based on complementary versus conventional medicine interventions. Overall, there was great variability in the types of interventions in the trials included across the meta-epidemiological studies.

Two meta-epidemiological studies included only trials with placebo[25] or placebo, sham, and noninterventions[49] comparators. Three studies had a single intervention as the comparator and this included standard percutaneous coronary intervention,[47] acupuncture,[40]and standard weight heparin.[23] A single study[34] presented stratified analyses comparing active versus nonactive comparators within the included trials. The remaining studies did not specify the type of comparators of the included trials.

The majority of meta-epidemiological studies did not specify the types of outcomes, and some of these were extracted from reviewing titles of included trials in the reference list (see Table 3 and Appendix D). Fifteen studies[16, 22, 23, 25-28, 30, 34, 42, 45, 57, 61, 68, 70, 72] evaluated the impact of allocation concealment using binary outcomes (Table 4). The remaining studies evaluated mean MD and SMD. Two studies compared the impact of allocation concealment on subjective versus objective outcomes,[12, 45, 68] and one study compared mortality versus morbidity outcomes.[12, 68]

Two meta-epidemiological studies considered trial characteristics such as study sample size,[48, 49] funding source,[32] and method of allocation concealment,[56, 57] when evaluating the impact of inadequate allocation concealment on the treatment effect.

Findings From Studies Concerning Allocation Concealment

Table 4 shows the main findings from the studies that evaluated the impact of inadequate allocation concealment on the treatment effect. Fifteen meta-epidemiological studies estimated the ROR or ratio of hazard ratios (RHR). Across studies, a value of less than one indicates that the presence of allocation concealment bias exaggerates the treatment effect. The ROR estimates for inadequate concealment varied, and showed a general trend that the presence of this bias could exaggerate treatment effect from 52 percent[28] to 8 percent;[12, 68] however, not all of these estimates were statistically significant. From these 15 studies that evaluated the impact using ROR/HR, nine meta-epidemiological studies were statistically significant for some comparisons; generally all studies showed consistently that the presence of inadequate or unclear allocation concealment exaggerated the treatment effect. However, in some of the studies demonstrating statistical significance, the effects were not consistent across all conditions evaluated within the same meta-epidemiological study (see Table 4). One study[29, 73] reanalyzed their results with more advanced statistical tests and showed a much smaller exaggerated effect that was only marginally significant. One study[35, 55] that evaluated different datasets showed different results with different methods of assessing the risk of bias. The populations, interventions, and outcomes in the meta-epidemiological studies that were statistically significant were notably heterogeneous (see Table 3).

Nine studies measured mean difference MD or SMD to evaluate the impact of this bias. Generally, the point estimates of the MD or SMD were negative, suggesting that studies with adequate concealment had smaller treatment effects. However, among these nine studies, only three differences in estimates were statistically significant (Table 4). One study[48, 49] found significant differences in the impact of inadequate concealment in trials that showed a large versus small benefit; similarly, the impact of bias in complementary versus conventional interventions was also statistically significant. However, the difference in the overall estimate (including all trials) was not significant in this meta-epidemiological study. The second study[24, 69] showed that the magnitude of the treatment effect was smaller for trials at high risk of bias relative to those with low risk of bias. The third study[47] reported the relative risk (RR) for three different outcomes associated with myocardial infarction and adjunctive devices. Although this study did not specify which aspects of randomization were evaluated, the findings suggest for two of these outcomes, the RR was greater in magnitude for unclear randomization compared with adequate randomization; however, the confidence intervals for all outcomes overlapped substantially.

Among the three studies evaluating the impact of allocation bias based on other considerations, the results were varied (Table 4). The findings from one study[32] suggest that trials that differ by source of funding (industry only, not for profit sources, or mixed sources) or the direction of the trial findings (favoring new intervention, conventional interventions, neutral findings or unclear) did not differ with respect to whether they had unclear or clear concealment (concealment adequate or inadequate). A second study[40] did not specify which aspect of selection bias associated with randomization was evaluated; the impact of the bias was assessed as a RR for selected adverse events associated with acupuncture.[40] The third study[47] compared each of three cardiac outcomes and showed that unclear randomization was associated with a larger pooled RR (relative to those that had adequate randomization) for all but two of the outcomes; however, the confidence intervals overlapped significantly for all outcomes.

Impact of Other Factors on the Effects of Allocation Concealment Bias

Only one[30] of 25 meta-epidemiological studies compared the impact of inadequate allocation concealment (relative to adequate concealment) on treatment effects based on different populations (four patient disease categories—cardiovascular, infectious diseases, pediatrics, surgical interventions—and an overall category combining all groups) and showed no significant differences for any of the four disease groups or overall. However, the trials within the disease/intervention areas may still be heterogeneous with respect to patient characteristics and interventions.

The effect of inadequate concealment by the type of intervention was considered in four studies. Three of these studies presented stratified results comparing drug versus nondrug interventions with respect to inadequate allocation concealment and calculated the ROR,[34, 45] or SMD.[49] Only one of these studies[49] showed that differences in treatment effects were larger in drug trials relative to nondrug trials (suggesting drug trials with inadequate concealment show larger treatment benefits). This same study[49] presented stratified analyses based on complementary versus conventional medicine and showed no significant differences in treatment effects between these intervention groups with respect to inadequate concealment.

A single study[34] presented stratified analyses comparing active versus nonactive comparators within the included trials and showed no statistical difference between those that did and did not adequately conceal allocation.

Four studies evaluated the impact of inadequate concealment on different outcomes. Two studies compared the impact on subjective (patient reported, physician assessed), objective, and mortality-related outcomes. Both studies showed that inadequate concealment exaggerated the treatment effect significantly; from 31 percent [ROR, 0.69 (0.59 to 0.82, p=0.009)][45] to 18 percent [ROR, 0.82 (0.70 to 0.94,].[12, 62, 68] For objective outcomes, one study was statistically significant and the other was not.[12, 62, 68] Another study[44] that presented stratified analyses for three outcomes in periodontal disease found no difference in the trials with or without adequate concealment. A fourth study[47] compared different outcomes associated with myocardial infarction and adjunctive devices. Although this study did not specify which aspects of randomization were evaluated, the findings suggest that the magnitude of the RR varies by the type of outcome and that for two of three outcomes the magnitude was greater when randomization was unclear; however, the confidence intervals overlapped.

Some of the studies evaluated the potential for other factors to impact the magnitude or direction of bias associated with allocation concealment. One study[28] showed that among small trials only, this bias was associated with significantly exaggerated treatment effects [ROR, 0.48 (0.25 to 0.92, p=0.027)]; it was also significant in small trials when compared with large (greater than 1,000 subjects) trials [ROR, 0.49 (0.27 to 0.86, p=0.014)]. One study[61] compared the impact of the protocol and the publication of a study and showed a significant difference (exaggerated effect 5 percent to 6 percent) when concealment bias was present.

Selection Bias: Sequence Generation (Randomization)

Description of Included Studies

Study Characteristics

Fourteen meta-epidemiological studies (18 publications)[5, 12, 13, 22, 24-28, 30, 46, 54, 55, 61, 62, 68, 69] reported on bias associated with sequence generation during randomization. From these meta-epidemiological studies, two (four publications) reevaluated trial data from previously published empirical evaluations that were also included in this systematic review. From these two meta-epidemiological publications, one study [70]contained bias information from two previously published datasets,[26, 27, 50] and the second study[12, 62, 68] contained bias information from five datasets[22, 28, 30, 34, 42]included in our systematic review.

The meta-analyses and trials included in the majority of the 14 meta-epidemiological studies evaluating the impact of poor sequence generation were derived from specialized registries for systematic reviews (predominately Cochrane) or from traditional bibliographic databases (predominately MEDLINE and EMBASE). Given the years of publication, the bibliographic databases, and the disease/intervention areas included, there was the potential for overlap of meta-analyses and trials across the meta-epidemiological studies. Eight meta-epidemiological studies captured more recently published trials, to the years 2000,[30] 2002,[6, 62, 63] 2006,[24, 50, 54, 55, 61] and 2007.[5] The remaining five meta-epidemiological studies included trials published up to the late 1990s.

Definition of Sequence Generation

Definitions of what was adequate and inadequate sequence generation varied across studies (Appendix Table D-4). From the 14 studies evaluating the impact of bias associated with sequence generation, two[25-27] employed the Jadad scale to define adequate sequence generation (Appendix Table D-4). One study[12, 62, 68] included any definition of sequence generation in the original dataset; it compared agreement between the individual trial ratings for bias from different studies and dichotomized ratings to indicate those that were deemed adequate or inadequate. Five studies used the criteria from the Cochrane Handbook version 2008.[5, 24, 46, 54, 61] The remaining five studies[13, 22, 28, 30, 50] had similar criteria for defining the presence or absence of this bias as the other meta-epidemiological studies in our systematic review.

Population, Interventions, Comparators, and Outcomes of Included Studies

Table 3 shows the different populations (diseases or age groups) included within the studies evaluating the impact of inadequate sequence generation. For many of the studies, the specific population, intervention, comparator, and outcome information was not fully reported; where possible, we reviewed the titles of included meta-analyses or systematic reviews and extracted information related to these characteristics (Appendix D). Overall, great variability for all these characteristics was noted across the trials included within these 14 meta-epidemiological studies.

Four of the 14 studies evaluating sequence generation included trials with specific patient or disease groups and these included trials restricted to back pain,[50] cancer,[61] and pediatric populations.[5, 46] Another study[30] compared four patient/disease categories that included cardiovascular, infectious diseases, pediatrics, surgical interventions, and an overall category

combining all groups The remaining meta-epidemiological studies included a wide variety of patient and disease populations.

One study[25] included only trials with placebo arms. One study had a single intervention as the comparator that included medications for persistent asthma.[54] The remaining studies did not specify the type of comparators within the included trials.

The majority of studies did not specify the types of outcomes, and we derived some of these based on the titles of included meta-analyses/trials where possible. Nine studies[12, 13, 22, 25-28, 30, 61, 62, 68, 70] evaluated the impact of sequence generation using the ROR. Six studies evaluated SMD,[5, 50, 70] and MD.[24, 46, 54, 69]

Findings From Studies Concerning Sequence Generation

Table 5 shows the findings from the meta-epidemiological studies that evaluated the potential effect of inadequate sequence generation. Nine meta-epidemiological studies estimated the ROR or RHR. A value of less than one indicates that the presence of allocation concealment bias exaggerates the treatment effect. The ROR estimates for inadequate sequence generation varied, and estimates showed the presence of this bias could exaggerate treatment effect from 51 percent[28] to 5 percent;[29, 70] however, not all of these estimates were statistically significant. From these nine studies that evaluated the impact using ROR/RHR, three meta-epidemiological studies were statistically significant for some comparisons. One study[28, 29] used more advanced statistical computations and showed that sequence generation bias was no longer statistically significant following the revised analysis. One study[12, 62, 68] found a significant difference when all outcomes were considered together but not for more specific outcomes (Table 5). Two other studies showed a significant exaggeration among studies with inadequate sequence generation; one[25] showed a 36 percent amplification of the treatment effect, and the other showed the largest impact (48 percent to 51 percent) and was consistent in small and large trials.[28]

Table 5. Summary of sequence generation risk of bias results by study

Study Identification	Results of Sequence Generation Bias	Impact on Treatment Effect* + Increase - Decrease = No Difference
Studies with results measured through Odds Ratios (OR), Ratio of Odds Ratios (ROR), Ratio of Hazard Ratios (RHR) or Ratio of Relative Risks (RRR)		
Schulz et al., 1995[22]	ROR = 0.95 (0.81 to 1.12)[a] (p = 0.58) ROR = 0.75 (0.55, 1.02)[b] (p <0.07)	= (U)[a] = (A vs. ID AC)[b]
Moher et al., 1998/Moher et al., 1999[26, 27]	ROR = 0.89 (0.67 to 1.20)[c] (p = Not reported)	= (I)[c]
Linde et al., 1999[25]	ROR = 0.64 (0.43 to 0.94)[d] (p = Not reported)	+ (I)[d]
Kjaergard, et al. 2001[28, 29]	Original ROR = 0.49 (0.30 to 0.81)[e] p < 0.001) ROR = 0.52 (0.28 to 0.93)[f] (p = 0.029) Revised ROR = 0.95 (0.86 to 1.04)	Original + (I)[e] + (I)[f] Revised = (I)
Balk et al., 2002[30]	Overall ROR = 1.01 (0.85 to 1.18) (p not reported) CV ROR = 1.14 (0.91 to 1.49) (p not reported) ID ROR = 0.93 (0.58 to 1.64) (p not reported) PA ROR = 0.88 (0.49 to 1.51) (p not reported) SX ROR = N/A	Overall =(I) CV = (I) ID = (I) PA = (I) SX N/A

Table 5. Summary of sequence generation risk of bias results by study (continued)

Study Identification	Results of Sequence Generation Bias	Impact on Treatment Effect* + Increase - Decrease = No Difference
Siersma et al., 2007[13]	ROR = 0.87 (0.74 to 1.01)[g] ROR = 0.85 (0.72 to 1.01)[h]	= (I)[g,h]
Hempel et al., 2011[35, 55]	Dataset 3: Jadad and Schultz: Jadad: randomization: ROR = 0.95 (0.62, 1.44)[j] Jadad: randomization: ROR = 0.88 (0.70, 1.09)[k] Schulz: sequence: ROR = 1.01 (0.66, 1.56)[j] Schulz: sequence: ROR = 0.92 (0.74, 1.15)[k] Dataset 4: Heterogeneity set CBRG: ROR = 1.18 (0.94, 1.49) CBRG: ROR = 1.18 (0.83, 1.67) corrected for clustering	 Jadad = (I)[j,k] Schultz = (I)[k] = (I) = (I)
Savovic et al., 2012[12, 62, 68]	All outcomes: ROR=0.90 (0.82 to 0.99)[l] -> Increase in between trial SD = 0.06/MA SD =0.05 Mortality outcomes ROR=0.86 (0.69 to 1.06)[l] -> Increase in between trial SD=0.08/ MA SD =0.06 Other objective outcomes: ROR= 1.00 (0.84 to1.20)[l] -> Increase in between trial SD = 0.07/MA SD = 0.07 Subjective/ mixed outcome: ROR=0.88 (0.76 to 1.00)[l] -> Increase in between trial SD = 0.05 and MA SD = 0.06	All outcomes + (I)[l] Mortality =(I)[l] Objective =(I)[l] Subjective =(I)[l]
Mhaskar et al., 2012[61]	Publication: RHR =1.01 (0.96, 1.05) Protocol plus publication: RHR = 1.00 (0.96, 1.05)	Publication: = Protocol: =
Studies with Results Measured through: Mean Difference (MD) or Standardized Mean Difference (SMD) in treatment effect, or Standardized Mean Effect Size (SMES)		
Hempel et al., 2011[55]	Dataset 1 (Back Pain): Randomization adequate vs. not: SMD = 0.02 (-0.12, 0.16)	= (I) = (I)
	Dataset 2 (EPC reports): Randomization adequate vs not: SMD = 0.01 (-0.15, 0.17)	= (I) = (I)
van Tulder et al., 2009[50]	SMD = 0.02 (-0.12, 0.16)	= (I)
Hamm et al, 2010,[52]	SMES = 0.23 (0.07 to 0.39) High RoB SMES = 0.45 (0.30 to 0.60) Unclear RoB SMES = 0.34 (0.21to 0.46) Low RoB	-(high vs. low)
Hartling et al., 2009[46]	SMD: Inadequate vs. adequate (p < 0.560) SMD:Unclear vs. adequate (p < 0.262)	= (I) = (U)
Hartling et al., 2011[54]	Forced expiratory volume: High or unclear vs. low: MD = -0.01 (-0.04 to 0.03) Symptom free days: High or unclear vs. low: MD = 1.71 (-4.11 to 7.54) Forced expiratory volume: High vs. Unclear or low: not estimable	= (I + U) = (I + U)
Hartling et al., 2012[24]	High bias: SMES = 0.86 (-0.35, 2.06) Unclear bias: SMES = 0.86 (0.68, 1.04) Low bias: SMES = 0.57 (0.47, 0.68) Overall meta-regression: p value = 0.10 Stratified analysis: For sequence generation, objective outcomes had greater bias than subjective outcomes (p=0.01).	 =

Notes: (+) = exaggerated odds ratio from bias; (=) = no significant effect from bias; (-) = reduced odds ratio from bias; (U) = relative to unclear rating; (I) = relative to inadequate rating

[a] Adjusting for allocation concealment, blinding, and post allocation exclusions.
[b] Estimate based on trials stratified by adequate versus inadequate allocation concealment
[c] Allowing for summary odds ratio (OR) to vary simultaneously according to the components (i.e., component by treatment interactions)
[d] Adjusting for allocation concealment, blinding, post allocation exclusions and intention-to-treat.
[e] Large versus small trials with adequate versus inadequate trials
[f] Small trials with adequate versus inadequate trials
[g] Logistic regression with random effect
[h] Weighted regression with random effect
[i] Adjusting for blinding
[j] Random effects model
[k] Fixed effects model
[l] Adjusted for other biases

Abbreviations: CBRG = Cochrane Back Review Group; CV = cardiovascular disease; RoB = risk of bias; EPC = Evidence-based Practice Center; ID = infectious disease; MA = meta-analysis; N/A = not applicable; OR = odds ratio; PA = pediatrics; RHR = ratio of hazard ratios; ROR = ratio of odds ratio; RRR = ratio of relative risks; SD = standard deviation; SMES = standardized mean effect size; SX = surgery.

Six studies evaluated the impact of inadequate sequence generation using MD and SMD, and only two studies showed statistically significant differences. One of these studies[5]showed that the treatment effect for trials with high risk of bias was smaller than those with low risk of bias. In contrast, the second study[24, 69] showed that the treatment effect in the low risk of bias trials was of a smaller magnitude than the unclear and high risk of bias groups.

Impact of Other Factors on the Effects of Sequence Generation

Only 1[30] of 14 meta-epidemiological studies compared the impact of inadequate sequence generation based on different populations (four patient/disease categories—cardiovascular, infectious diseases, pediatrics, surgical interventions—and an overall category combining all groups) and showed no significant differences for any of the four groups or the overall category.

The effect of inadequate sequence generation by the type of intervention was not evaluated in any study. A single study[34] presented stratified analyses comparing active versus nonactive comparators within the included trials and showed no statistical difference between those that did and did not adequately undertake sequence generation.

A single study[12, 62, 68] evaluated the impact of different outcomes and showed that the presence of inadequate sequence generation did not influence the treatment effect (Table 5).

Two studies evaluated the potential for other factors and the impact on the magnitude or direction of bias associated with sequence generation. One study[28] compared large versus small trials and showed that with small trials only the impact of the presence of this bias was significant [ROR, 0.49 (0.30 to 0.81, p=0.001)]; it was also significant when comparing large (greater than 1000 subjects) to small trials [ROR, 0.52 (0.28 to 0.93, p=0.029)]. One study[61] compared the impact of the protocol and the publication of a study and showed no significant difference when sequence generation bias was present.

Confounding

Description of Included Studies

Study Characteristics

Two studies examined the relationship between confounding and treatment effect (Table 6).[30, 50] One study calculated SMDs in a meta-analysis of back pain trials identified from a review of 15 Cochrane reviews in the 2005 issue that focused on conservative treatments for low back pain.[50] The second study was a meta-epidemiological assessment of the effect of risk of bias on treatment effects, stratified by clinical condition.[30]

Table 6. Summary of confounding results by study

Study Identification	Allocation Concealment: Results of Bias	Effect[a]
Balk et al., 2002[30]	Overall ROR: 0.96 (0.79 – 1.23)	=
	Infectious disease ROR: 0.94 (0.60 – 1.49)	=
	Pediatrics ROR: 0.96 (0.50-1.65)	=
	Surgery ROR: 1.20 (0.76 – 1.72)	=
van Tulder et al., 2009[50]	SMD: -0.10 (-0.24, 0.05)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias

Abbreviations: ROR = ratio of odds ratio; SMD = standardized mean difference.

Definition of Confounding

One study evaluated whether baseline differences in groups that could be confounders were examined using a dichotomous response.[30] The study did not specify how unclear or no information was handled. The other study considered similarity of prognostic factors at baseline and recorded responses as "yes," "no," or "don't know." In order to be scored "yes," reviewers had to agree that demographic factors, duration and severity of complaints, percentage of patients with neurologic symptoms, and value of main outcome measure(s) were similar. In the analysis, the authors combined "no" and "don't know" to generate a dichotomous judgment.[50]

Population, Interventions, Comparators, and Outcomes of Included Studies

As noted earlier, one study was limited to low back pain.[50] The second study included and stratified results for three clinical conditions: infectious diseases, pediatrics, and surgery. Both included a variety of interventions within each clinical area. Neither study specified types nor limitations on comparators. As noted previously, the meta-epidemiological study assessed ratios of odds ratios,[30] and the meta-analysis evaluated treatment effect differences.[50]

Findings Related to Confounding

Although the newer study offered much greater specificity in identifying confounding,[50] neither study found evidence that confounding influences the treatment effect of studies.[30, 50] Notably, both studies included RCTs only. Even in well-conducted RCTs, a dissimilarity of prognostic variables between intervention and control arms may be expected as a matter of chance: across all possible randomizations, interventions and control arms should be similar. Such dissimilarity between arms, because it occurs at random, may not influence treatment effect

consistently. As a result of their exclusive focus on RCTs, these studies may not be able to address the issue of the effect of confounding on effect size.

Performance Bias: Fidelity to Protocol, Unintended Interventions, or Cointerventions

Description of Included Studies

Study Characteristics

Two studies examined the relationship between performance bias and treatment effect (Table 7).[50, 74] One study calculated treatment effect differences in a meta-analysis of back pain trials identified from a review of 15 Cochrane reviews in 2005 (216 trials) that focused on conservative treatments for low back pain.[50] The second study reevaluated results from four datasets: the van Tulder analysis of back pain trials,[50] an analysis of Agency for Healthcare Research and Quality (AHRQ) systematic reviews (165 trials),[55] a dataset compiled by Moher et al. that demonstrated the effect of quality criteria on treatment effect[27] ("pro-bias," 100 trials), and a dataset compiled by Balk et al. that found no effect of quality criteria on most treatment effects ("heterogeneity," 149 trials).[30] No trials appeared in more than one dataset.

Definition of Performance Bias Measures

Both publications evaluated two measures of performance bias: fidelity to protocol and unintended interventions. They relied on the Cochrane Back Pain Review Group criteria. Specifically, for compliance, reviewers marked "yes" if there were no co-interventions or they were similar between the index and control groups, "no" for dissimilar intervention, and "don't know" for unclear information. For compliance, reviewers based their judgment on the reported intensity, duration, number, and frequency of sessions for both the index intervention and control intervention(s) and selected between "yes," "no," and "don't know" for responses. For single-session interventions such as surgery, reviewers marked the item as irrelevant. In all cases, the analysis compared "yes" with "not yes" responses.

Population, Interventions, Comparators, and Outcomes of Included Studies

The datasets cover an extremely wide range of populations, interventions, comparators, and outcomes. Of the four databases covered by this body of evidence, the van Tulder back pain group is the only one limited to a single clinical condition (nonspecific low back pain), but even this review includes numerous nonsurgical treatments.

Findings Related to Performance Bias

Overall, the results (Table 7 and Table 8) are inconsistent in magnitude, direction, and statistical significance. The Moher et al. database stands out as the only dataset supporting a statistically significant difference in treatment effects as a result of compliance or cointerventions. Analysis of studies from the Moher et al. dataset suggest that poor compliance inflates treatment effect and dissimilar cointerventions reduce treatment effect, but neither of these conclusions is supported by the analyses of studies from the other datasets.

Table 7. Summary of compliance results by study

Study Identification	Acceptable Compliance: Results of Bias	Effect[a]
van Tulder et al., 2009[50]	ROR: -0.01 (-0.15, 0.14)	=
Hempel et al., 2011[74b]	SMD for dataset 2 (EPC reports): 0.02 (-0.12, 0.17)	=
	ROR for dataset 3 ("pro-bias"): 0.72 (0.59, 0.88)	+
	ROR for dataset 4 ("heterogeneity"): 1.12 (0.89, 1.41)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias.
[b] This row does not repeat results from dataset 1 because they appear in the row for van Tulder.

Abbreviations: EPC = Evidence-based Practice Center; ROR = ratio of odds ratio; SMD = standardized mean difference.

Table 8. Summary of cointerventions results by study

Study Identification	Similar Cointerventions: Results of Bias	Effect[a]
Van Tulder et al., 2009[50]	SMD: -0.09 (-0.23, 0.05)	=
Hempel et al., 2011[74b]	SMD for dataset 2 (EPC reports): 0.05 (-0.15, 0.28)	=
	ROR for dataset 3 ("pro-bias"): 1.50 (1.22, 1.85)	-
	ROR for dataset 4 ("heterogeneity"): 0.85 (0.69, 1.03)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias
[b] This row does not repeat results from dataset 1 because they appear in the row for van Tulder.

Abbreviations: EPC = Evidence-based Practice Center; ROR = ratio of odds ratio; SMD = standardized mean difference.

Performance Bias: Patient/Caregiver or Provider Blinding

Description of Included Studies

Study Characteristics

Four studies separately analyzed the effect of lack of patient, caregiver, or provider blinding.[30, 35, 50] Only Balk et al.[30] conducted a meta-epidemiological hierarchical analysis, nesting trials within meta-analyses. Nuesch et al. evaluated the effect of blinding, adjusting for allocation concealment in one analysis and intention-to-treat in a second and also conducted separate sensitivity analyses based on the size of the treatment benefits, trial heterogeneity, whether the treatment was pharmacologic, and whether it would be considered complementary medicine.[48] The Hempel et al. study of four databases[35] (also reported in van Tulder et al.[50]) did not adjust for the possible effect of other biases on outcomes.

Population, Interventions, Comparators, and Outcomes of Included Studies

Bulk et al.[30] presented results overall and separately for three disease or treatment areas, while Nuesch et al.[48] was limited to osteoarthritis trials. Studies of nonsurgical treatment for low back pain were presented in van Tulder et al.,[50] and Hempel et al.[74] also compared outcomes in three additional datasets; one concerning a diverse set of topics, one described as "pro-bias" because Moher et al.[27] had found more attenuated results in biased analyses, and the last based on a subgroup of the studies included in Bulk et al.[30] that exhibited heterogeneity across studies.

Findings Related to Patient, Caregiver, and Provider Blinding

We found few significant differences in results based on either patient or provider blinding. In Bulk et al.'s eight analyses concerning patient or caregiver blinding, only one was statistically significant[30] (Table 9). In addition to large confidence intervals, point estimates differed in

direction across analyses; in three of the eight, the point estimate showed a larger effect in the blinded group. Nuesch et al. also found no significant differences in treatment effects, after adjusting for allocation concealment or intent to treat.[48] In a sensitivity analysis, the authors found some evidence that lack of patient blinding would be more likely to increase effect estimates in nonpharmaceutical trials. The Hempel[74] comparison of four datasets found a significant difference between groups in one of the four in relation to both patient and provider blinding. However, in both analyses, the treatment effect was larger in the blinded arms (opposite of the direction one would hypothesize).

Table 9. Summary of patient and/or provider blinding risk of performance bias results by study

Study Identification	Patient and/or Provider Blinding: Results of Bias	Effect[a]
Balk et al., 2002[30]	Patients blinded vs. not:	
	• Overall: ROR = 0.95 (0.70 to 1.13)	=
	• Cardiovascular disease: ROR = 1.08 (0.86 to 1.38)	=
	• Infectious disease: ROR = 0.70 (0.46 to 1.11)	=
	• Pediatrics: ROR = 0.79 (0.39 to 1.19)	=
	Caregivers blinded vs. not:	
	• Overall: ROR = 0.98 (0.75 to 1.20)	=
	• Cardiovascular disease: ROR = 1.09 (0.91 to 1.29)	=
	• Infectious disease: ROR = 0.62 (0.43 to 0.91)	+
	• Pediatrics: ROR = 1.13 (0.73 to 1.84)	=
Nuesch et al., 2009[48]	Patients blinded vs. not:	
	• SMD = -0.15 (-0.39 to 0.09)	=
	• SMD =0.01 (-0.18 to 0.18)[b]	=
	• SMD =-0.06 (-0.20 to 0.09)[c]	=
	Interaction of patient blinded (vs. not) and:	
	• Small vs. large treatment benefits: p=0.75	=
	• High vs. low between-trial heterogeneity: p=0.19	=
	• Nonpharmacologic vs. pharmacologic trial: p<0.001 (Differences more pronounced in MAs of nonpharmacologic interventions)	+
	• Complementary vs. conventional medicine: p=0.07	=
van Tulder et al., 2009,[50] Hempel et al., 2011[55] Dataset 1	• Patients blinded vs. not: SMD = -0.03 (-0.18 to 0.11)	=
	• Provider blinded vs not: SMD = -0.10 (-0.26 to 0.06)	=
Hempel et al., 2012,[74] 2011[55]	Dataset 2	
	• Patient blinded vs. not: SMD =0.21 (0.04 to 0.39)[d]	-
	• Provider blinded vs. not: SMD =0.19 (0.03 to 0.35)[d]	-
	Dataset 3	
	• Patient blinded vs. not: ROR=0.97 (0.78 to 1.21)	=
	• Provider blinded vs not: ROR=0.83 (0.67 to 1.02)	=
	Dataset 4	
	• Patient blinded vs. not: ROR=0.88 (0.71 to 1.08)	
	• Provider blinded vs. not: ROR=0.94 (0.77 to 1.15)	=
		=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias
[b] Adjusting for allocation concealment
[c] Adjusting for intention-to-treat
[d] Point estimate is larger in blinded studies

Abbreviations: MA = meta-analyses; ROR = ratio of odds ratios; vs. = versus; SMD = standardized mean difference.

Detection Bias: Assessor Blinding

Description of Included Studies

Study Characteristics and Definition of Assessor Blinding

Eight studies evaluated whether assessor blinding is related to differences in outcomes based on the results of meta-analyses.[23, 30, 35, 47, 50, 59, 60] Other studies examining patient, provider, and double-blinding are described separately below. Some of the studies of double blinding include assessor blinding; however, because of combined reporting with patient and/or provider blinding, these studies cannot isolate the effect of lack of assessor blinding.

Of the studies examining assessor blinding, only Balk et al.[30] conducted a meta-epidemiological hierarchical analysis, nesting trials within meta-analyses. All other studies were meta-analyses of trials. Two studies adjusted for potential confounding from other biases in their analyses. Liu et al.[59] adjusted for the effect of intention to treat on outcome results; Inaba et al.[47] adjusted for funding source and other issues. Hrobjartsson et al.[60] focused on trials that included both blinded and nonblinded assessment within each trial.

All of the studies specifically identified the assessor as blinded, or alternatively, as not blinded or blinding was unknown. We did not find any studies evaluating the effect of lack of data analyst blinding.

Population, Interventions, Comparators, and Outcomes of Included Studies

Most studies were limited to one disease condition or treatment area or stratified analyses based on condition; specifically, deep vein thrombosis,[23] cardiovascular disease,[30] infectious disease,[30] pediatrics,[30] periodontology,[44] acute myocardial infarction,[47] low back pain,[55, 75] and progressive resistance muscle strength.[59] Studies commonly included a variety of interventions and comparators; those with more limited comparisons included low molecular weight heparin compared with regular heparin,[23] adjunctive mechanical devices in heart procedures,[47]and progressive resistance muscle strength training.[59]

Three of the studies included categorical outcomes,[23, 30, 60] two included continuous outcomes,[44, 59] and three included both types of outcomes.[35, 47, 50]

Findings Related to Assessor Blinding

Results were mixed in relation to whether lack of assessor blinding was related to larger outcome effects (Table 10). Although the point estimate in virtually all analyses was larger in studies without assessor blinding, the difference was statistically significant in all or some of the analyses reported in half of the eight studies.

Two studies, each limited to a particular clinical area, found significant results using regression analysis to adjust for other biases. Juni et al.[23] found larger relative risk ratios in deep vein thrombosis trials measuring deep vein thrombosis and bleeding, a 35 percent exaggeration of the effect. Results continued to be significant after adjusting for concealed allocation and withdrawals (results not reported). Liu et al.[59] found a significantly larger effect in unblinded progressive resistance muscle strength trials measuring muscle strength, including even more exaggerated differences after adjusting for whether study outcomes were based on an intention-to-treat analysis. Standardized mean differences were 20 percent larger without adjusting for

intention to treat, increasing to 35 percent, after taking this other bias adjustment into account. Hrobjartsson et al.,[60] focusing on a cross-section of studies that included both blinded and unblinded assessment, found outcomes based on unblinded assessment to be significantly exaggerated overall (ROR, 0.64) and even more so when study outcomes were subjective (ROR = 0.55).

Table 10. Summary of assessor blinding risk of detection bias results by study

Study Identification	Assessor Blinding: Results of Bias	Effect[a]
Juni et al.,1999[23]	• RRR = 0.65 (0.43 to 0.99)	+
Balk et al., 2002[30]	• Overall: ROR = 1.02 (0.82 to 1.22)	=
	• Cardiovascular disease: ROR = 1.11 (0.87 to 1.39)	=
	• Infectious disease: ROR = 0.84 (0.55 to 1.27)	=
	• Pediatrics: ROR = 1.02 (0.57 to 1.61)	=
	• Surgery: ROR = 0.87 (0.56 to 1.36)	=
Fenwick et al., 2008[44]	• Probing depth (adequate vs. inadequate: MD=-0.20 (-0.76 to 0.36)	=
	• CAL/PAL: (adequate vs. inadequate) MD =-0.19 (-1.05 to 0.68)	=
Inaba et al., 2009[47]	Mortality	
	Unclear vs. no blinding: MES=0.81 (0.50 to 1.30)[a]	
	Yes vs. no blinding: MES =0.90 (0.51 to 0.95)[a]	
	• CIs overlap and so difference: (p > 0.05)	=
	Incomplete ST resolution	
	Unclear blinding: MES=0.60 (0.42 to 0.87)[b]	
	Yes blinding: MES=0.82 (0.73 to 0.93)[b]	
	• CIs overlap and so difference: (p > 0.05)	=
	Impaired myocardial blush grade	
	Unclear blinding: MES=0.81 (0.50-1.30)[b]	
	Yes blinding: MES=0.90 (0.51 to 1.59)[b]	
	• CIs overlap and so difference: (p > 0.05)	=
van Tulder et al., 2009,[50] Hempel et al., 2011[55] Dataset 1	• SMD: -0.10 (-0.25, 0.04)	=
Liu, LaValley, and Latham, 2011[59]	• SMD=-0.80 (-1.35 to -0.25)	+
	• SMD=-0.65 (-1.26 to -0.04)[c]	+
Hempel et al., 2012,[74] 2011[55]	Dataset 1	
	• SMD =-0.10 (-0.25 to 0.04)	=
	Dataset 2	
	• SMD =0.06 (-0.28 to 0.41)	=
	Dataset 3	
	• ROR=1.35 (1.05 to 1.73)	+
	Dataset 4	
	• ROR=0.99 (0.80 to 1.22)	=
Hrobjartsson et al., 2012[60]	Overall	
	• ROR=0.64 (0.43 to 0.96)	+
	Subgroup of studies with subjective outcomes:	
	• ROR=0.55 (0.32 to 0.95)	+
	Subgroup of studies with moderately subjective outcomes:	
	• ROR=0.93 (0.56 to 1.54)	+

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias
[b] Adjusting for study size, single or multicenter design, type of devices used, study regions, presence of conflicts of interest
c Adjusting for intention-to-treat adjustment

Abbreviations: CAL/PAL = clinical or probing attachment level; MD = mean difference; MES = mean effect size; ROR = ratio or odds ratios; RRR = ratio of relative risk; SMD =standardized mean difference.

In contrast, the Balk et al. study, based on a hierarchical model, adjusting for nesting within meta-analyses, found a small and nonsignificant difference combining all RCTs (OR, 1.02);

within disease-specific subgroups, the direction of the effect of point estimates differed by subgroup.[30] Comparing results across four datasets, Hempel et al. found significant differences in only one of the four, the so-called "pro-bias" dataset that had shown evidence of bias in an earlier analysis.[55]

Detection Bias: Valid Statistical Methods

Description of Included Studies and Findings

Two studies considered the effect of using valid statistical methods (Table 11). In the Balk et al. study,[30] the validity of statistical methods was determined from quality rating evaluations that asked whether the methods were considered valid and appropriate by the reviewers, based on study design and outcomes of interest. Results were only significantly different in the subgroup of surgery trials, one of three clinical subgroups, ROR = 1.63 (95% CI, 1.03 to 2.83). Other point estimates were smaller; the relationship in the pediatrics studies was close to no difference but the estimate was opposite of the hypothesized direction. Mhaskar et al.[61] considered whether prespecification of the α and β error would affect the treatment effect. For both, if the effect was not prespecified in the protocol and publication, the effect was larger, though relatively small, and not statistically significant.

Table 11. Summary results of studies evaluating valid statistical measures

Study Identification	Valid Statistical Measures: Results of Bias	Effect[a]
Balk et al., 2002[30]	Overall: ROR=1.11 (0.95 to 1.31)	=
	Cardiovascular disease: ROR=1.03 (0.81 to 1.33)	=
	Infectious disease: ROR=1.17 (0.78 to 1.77)	=
	Pediatrics: ROR=0.97 (0.48 to 1.73)	=
	Surgery: ROR=1.63 (1.03 to 2.83)	-
Mhaskar et al., 2012[61]	Prespecification of α in protocol and publication	
	RHR = 1.06 (0.96 to 1.18)	=
	Prespecification of β in protocol and publication	
	RHR = 1.07 (0.94 to 1.22)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias

Abbreviations: RHR = ratio of hazard ratios; ROR = ratio of odds ratios.

Detection/Performance Bias: Double Blinding

Description of Included Studies

Study Characteristics

Eighteen studies evaluated whether double blinding is related to differences in outcomes based on the results of meta-analyses.[5, 12, 13, 16, 22, 24, 25, 27, 28, 30, 33, 34, 38, 40, 42, 45, 46, 55, 68] Some of the studies included data from earlier studies. The Sterne et al. study[16] reanalyzed data included in the Schulz et al. study[22] based on additional model specifications. The Wood et al. study[45] combined data from three earlier studies included in this review: Schulz et al.,[22] Kjaergard et al.,[28] and Egger et al.[34] The Pildal et al. study combined data from six earlier studies: Schulz et al.,[22] Moher et al.,[27] Kjaergard et al.,[28] Egger et al.,[34] Balk et al.,[30] and Als-Nielson et al.[43] Hempel et al.[55] reanalyzed data used in the Moher et al.[27]and Balk et al.[30] studies. Savovic et al. combined data from seven earlier studies, six of which are included in this review: Schulz et

al.,[22] Kjaergard et al.,[28] Egger et al.,[34] Balk et al.,[30] Als-Nielson et al.,[43] and Pildal et al.[42] Eight of the studies conducted meta-epidemiological studies, nesting trials within meta-analyses.[16, 22, 27, 28, 30, 34, 42, 68] The modeling in six of the studies adjusted for the potential effect from other sources of bias.[22, 25, 33, 45, 46, 68]

Definition of Blinding

Blinding was generally dichotomized as double-blind versus a comparison; the description of the comparison differed across studies (i.e., not, inappropriate, inadequate, other, absent, unclear, not reported). Studies differed in the categories and specificity that was used in defining double blinding: six described themselves as double blind without specifying the two blinded groups;[5, 24, 27, 33, 42, 68] one used the Jadad scale[76] of 0, 1, or 2 points based on whether the study included an explicit statement that patients and evaluators were blinded and that treatments were indistinguishable;[25] one only required indistinguishable treatments;[28] three required that patients and either caregiver or outcome assessors were blinded;[13, 30, 40] one required that either two unspecified groups were blinded or the outcome assessor;[34] two using the same data required triple blinding—patients, caregivers, and outcome assessors;[16, 22] one included both single and double-blinded trials;[38, 46] two did not specify the level of blinding;[38, 59] and two included more than one approach, incorporating more than one earlier analysis.[45, 55]

Population, Interventions, Comparators, and Outcomes of Included Studies

Seven of the eighteen studies were limited to one disease condition or treatment area, specifically, perinatal trials,[16, 22] homeopathy,[25] tumor suppression in head and neck cancer,[38] acupuncture,[40] progressive resistance muscle strength,[59] and pediatrics[46] (Table 3). Studies commonly included a variety of interventions and comparators; those with more limited comparisons included therapeutic or preventive interventions,[34] homeopathic,[25] tumor suppressor protein TP53,[38] and acupuncture. [40]Two studies evaluated differences in treatment effects based on continuous outcomes.[5, 24] All other studies evaluated differences in binary outcomes.

Findings Related to Double Blinding

We found evidence of an attenuated effect from lack of double blinding in studies with subjective outcomes. Three studies found statistically significant differences, with results in the anticipated direction[34, 45, 68] (Table 12). In Egger et al. (2003) results were limited to psychiatry RCTs (ROR = 0.47); in Wood et al. (2008) and Savovic et al. (2012), results were described by the authors as subjective (more specifically, in Savovic as subjective and mixed) and point estimates were similar (ROR = 0.77).[34, 45, 68] Although we know that there is some overlap in the RCTs included in these studies, we do not know the extent of the overlap, but this is likely the reason for the similarity in the findings. The Wood et al. analysis adjusted for allocation concealment, while the Savovic et al. study adjusted for allocation concealment and sequence generation.

Table 12. Summary of double blinding risk of detection/performance bias results by study

Study Identification	Double Blinding Specification: Results of Bias	Effect[a]
Schulz et al., 1995[22]	Participants, caregivers and outcome assessors all described as blinded: ROR = 0.83 (0.71 to 0.96) (p = 0.01)[b]	 +
Moher et al., 1998[26] companion Moher et al., 1999[27]	Composition of blinded groups not specified: ROR = 1.11 (0.76 to 1.63)	 =
Linde et al., 1999[25]	Jadad scale: patient and assessor blinded and treatment indistinguishable: ROR = 0.26 (0.14 to 0.51)[b]	 +
Kjaergard, Villumsen and Gluud, 2001[28]	Identical placebo tablets or similar vs. not performed or tablets versus injections or similar: ROR = 0.56 (p = 0.041) Studies with inadequate allocation concealment ROR = 0.52 (0.29 to 0.92) (p = 0.024)	 + +
Balk et al., 2002[30]	Patients and either caregiver or outcome assessor blinded: Overall: ROR = 1.02 (0.79 to 1.24) Cardiovascular disease: ROR = 1.10 (0.90 to 1.33) Infectious disease: ROR = 0.71 (0.47 to 1.12) Pediatrics: ROR = 1.05 (0.56 to 1.61)	 = = = =
Sterne et al., 2002[16]	Composition of blinded groups not specified: Logistic regression Model-based SE: ROR = 0.68 (0.60 to 0.78) Robust SE: ROR = 0.68 (0.55 to 0.84) MA using logistic regression Fixed effects: ROR = 0.68 (0.60 to 0.78) Random effects: ROR = 0.67 (0.54 to 0.82) MA using random-effects meta-regression Fixed effects: ROR = 0.72 (0.61 to 0.85) Random effects: ROR = 0.71 (0.57 to 0.87)	 + + + + + +
Egger et al.,2003[34]	Composition of blinded groups not specified or assessor blinded: Overall: ROR = 0.88 (0.75 to 1.04) Infectious diseases: ROR = 0.91 (0.39 to 2.17) Neonatology: ROR = 0.88 (0.63 to 1.25) Neurology: ROR = 0.90 (0.61 to 1.33) Obstetrics & gynecology: ROR = 0.96 (0.66 to 1.39) Psychiatry: ROR = 0.47 (0.26 to 0.84) Miscellaneous: ROR = 0.97 (0.79 to 1.20)	 = = = = = + =
Als-Nielson et al., 2003[33]	Composition of blinded groups not specified: OR=2.9 (1.4-6.0) p =.004[b]	 +
Kyzas et al., 2005[38]	Composition of blinded groups and level of blinding not specified; could be single or double: Mortality Random effects Blinding stated: RR = 1.05 (0.86 to 1.28) Blinding not stated: RR = 1.32 (1.06 to 1.65) CIs overlap and so difference: (p > 0.05) Fixed-effects Blinding stated: RR = 1.05 (0.92 to 1.20) Blinding not stated: RR = 1.29 (1.11 to 1.50) CIs overlap and so difference: (p > 0.05) Lymph node metastasis Random effects Blinding stated: RR = 1.15 (1.02 to 1.30) Blinding not stated: RR = 1.18 (1.06 to 1.32) CIs overlap and so difference: (p > 0.05) Fixed-effects Blinding stated: RR = 1.16 (1.01 to 1.34) Blinding not stated: RR = 1.29 (1.17 to 1.42) CIs overlap and so difference: (p > 0.05)	 = = = =

Table 12. Summary of double blinding risk of detection/performance bias results by study (continued)

Study Identification	Double Blinding Specification: Results of Bias	Effect[a]
Derry et al., 2006[40]	Double blinding of patient and assessor, practitioner (acupuncturist) could not be blinded	
	Randomized and blinded	
	Nausea: RR = 0.78 (0.58 to 1.05)	=
	Vomiting: RR = 0.84 (0.62 to 1.14)	=
	Antiemetic consumption: RR = 0.83 (0.64 to 1.09)	=
	Randomized, blinded, over 100 patients	
	Nausea: RR = 0.82 (0.58 to 1.17)	=
	Vomiting: RR = 0.79 (0.55 to 1.14)	=
	Antiemetic consumption: RR = 0.83 (0.62 to 1.11)	=
	Randomized, blinded, over 100 patients; control event rate ≥20%	
	Nausea: RR = 0.82 (0.58 to 1.17)	=
	Vomiting: RR = 0.76 (0.54 to 1.05)	=
	Antiemetic consumption: RR = 0.89 (0.69 to 1.16)	=
Pildal et al., 2007[42]	Groups counted towards double blinding not specified; patients and caregivers reported as blinded; or placebo controlled without indication that treatments distinguishable or that investigators might have become unblinded. If only patient and assessor groups, not counted as double blinded.	
	Data from this study only: ROR = 0.94 (0.80 to 1.10)	=
	Data combined from this and 6 earlier MA studies: ROR = 0.91 (0.83 to 1.00)	=
Siersma et al., 2007[13]	Groups described as double or outcome assessor blinded	
	Logistic regression with random effect: ROR=1.09 (0.90 to 1.33)	=
	Weighted regression with random effect: ROR=1.14 (0.94 to 1.39)	=
Wood et al., 2008[45]	Groups as defined in Schulz et al., 1995,[22] Kjaergard, Villumsen, and Gluud, 2001,[28] and Egger et al., 2003[34] and described in this study as double blind and using adequate methods such as identical placebo tablets or including blinding of the outcome assessor	
	Overall: ROR=0.93 (0.83 to 1.04)	=
	Other than all-cause mortality: ROR=0.83 (0.70 to 0.98)	+
	All-cause mortality: ROR=1.04 (0.95 to 1.14)	=
	Interaction of blinding and outcome mortality (vs. not): (p =0.011)	+
	Subjective outcomes: ROR=0.75 (0.61 to 0.93)	+
	Objective outcomes: ROR=1.01 (0.92 to 1.10)	=
	Interaction of blinding and outcome subjective (vs. not): (p=0.01)	+
	Drug interventions: ROR=0.92 (0.81 to 1.05)	=
	Non-drug interventions: ROR=1.00 (0.71 to 1.39)	=
	Interaction of blinding and type of intervention: (p = 0.66)	=
	Trials with adequate allocation concealment: ROR=1.02 (0.92 to 1.14)	=
	Subjective outcomes: ROR=0.77 (0.65 to 0.91)[d]	+
	All outcomes other than mortality: ROR=0.85 (0.76 to 0.95)[d]	+
Hartling et al., 2009[46] supplemental analysis Memo 10	Blinding level (single or double) not specified	
	In comparison to adequate blinding:	
	Unclear (p < 0.828)[e]	=
	Inadequate (p < 0.952)[e]	=
Hamm et al, 2010,[52]	Described as double blind, groups not specified	
	High RoB studies: SES =0.43 (0.22 to 0.65)	=
	Unclear RoB studies: SES =0.39 (0.19 to 0.59)	=
	Low RoB: SES =0.37 (0.25 to 0.48)	=
Hempel et al., 2011[55]	Jadad definition: blinding score=2	
	Dataset 3	
	Random effects MA: ROR=1.08 (0.72 to 1.64)	=
	Fixed effects MA: ROR=1.05 (0.85 to 1.31)	=
	Schulz definition: blinded described as reported or not	
	Dataset 3	
	Random effects MA: ROR=1.13 (0.74 to 1.71)	=
	Fixed effects MA: ROR=0.92 (0.75 to 1.12)	=

Table 12. Summary of double blinding risk of detection/performance bias results by study (continued)

Study Identification	Double Blinding Specification: Results of Bias	Effect[a]
Savovic et al., 2012[68]; Savovic et al., 2012[62]	Described as double (groups not specified) vs. unclear or lack of double blinding	
	All outcomes ROR = 0.86 (0.73 to 0.98)[f]	=
	Mortality ROR = 1.07 (0.78 to 1.48)[f]	=
Savovic et al., 2010[12]	Other objective outcomes ROR = 0.91 (0.64 to 1.33)[f]	=
	Subjective or mixed outcome ROR =0.77 (0.61 to 0.93)[f]	+
Hartling et al., 2012[24]	Double blinding based on Cochrane RoB tool	
	High RoB studies: SMES=0.79 (0.40 to 1.18)	=
	Unclear RoB studies: SMES =0.77 (0.62 to 0.93)	+
	Low RoB studies: SMES=0.56 (0.44 to 0.68)	+
	Diff between RoB study groups: (p = 0.31)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias
[b] Adjusting for explicitly randomized, adequate concealment, complete followup.
[c] Adjusting for funding source and treatment effect (disease area and type of treatments)
[d] Adjusting for allocation concealment
[e] Adjusting for efficacy, crossover, factorial, binary outcome, and objective outcome
[f] Adjusting for sequence generation and allocation concealment

Abbreviation: Diff = difference; RoB = risk of bias; ROR = ratio of odds ratios; RR = risk ratio; SMES =standardized mean effect size.

In other studies or in relation to other types of outcomes, point estimates were generally in the expected direction, but double blinding was not associated with statistically significant differences in effect in most of the studies. We found several exceptions. In the oldest study we reviewed, Schulz et al. (1995)[22] found that odds ratios were exaggerated in obstetrics studies that lacked double blinding, using a meta-epidemiological design that adjusted for meta-analysis nesting and also adjusted for other bias components (allocation concealment, sequence generation, and exclusions from followup). Using data from the Schulz et al. study, Sterne et al.[16] reached a similar result, based on a variety of model specifications (logistic regression with model-based and robust standard errors, meta-analysis using logistic regression [fixed effects and random effects], and meta-analysis using random-effects meta-regression [fixed effects and random effects]), while not adjusting for other bias components. Kjaergard et al.[28] found a difference in effect in pharmaceutical trials from a variety of treatment areas using a meta-epidemiological design and, further, that the effect was significantly more pronounced in smaller studies. Als-Nielson et al.[33] also found a significant difference in effect in pharmaceutical trials, adjusting for funding source and treatment effect (disease area and type of treatments). Linde et al.[25] found an effect in homeopathy trials, adjusting for allocation concealment and completeness of followup, but not for the clustering of RCTs in meta-analyses.

Attrition Bias

Description of Included Studies

Study Characteristics

Fifteen studies evaluated how attrition bias influences the treatment effect in meta-analyses (Table 10).[6, 22, 23, 28, 30, 39, 49, 50, 52, 61, 64, 67, 74]

All studies included trials published before the first CONSORT statement in 1996. Two studies provided subgroup analysis to examine the effect of small versus large trials,[28] level of

heterogeneity, magnitude of effect, type of intervention, type of clinical area,[49] data source,[74]and outcome.[6, 47]

Some included studies focused on very specific methodological issues; the issue of attrition was secondary to the main analysis in these studies. One study evaluated a range of statistical techniques with and without accounting for random effects, other variables, and clustering at the meta-analyses level; the results below present data for the random effects multivariable multilevel model.[13] A second focused on whether published methodological quality in trials reflects the quality in protocols.[61]

Two studies described below are related.[50, 74] One study by van Tulder et al. calculated standardized mean differences in treatment effect in a meta-analysis of back pain trials identified from a review of 15 Cochrane reviews in 2005 that focused on conservative treatments for low back pain.[50] As noted in the section on performance bias, the second study reevaluated results from four datasets, including the van Tulder study.

Definition of Attrition Bias

Of the 15 studies, 13 used dichotomous and clearly described measures of risk of bias.[6, 22, 23, 28, 30, 39, 49, 50, 52, 61, 64, 67, 74] These measures varied in focus and evaluated reporting, conduct, or a combination of conduct, reporting, and analysis. Studies focusing on reporting evaluated whether individual trials provided numbers and reasons for dropouts.[28, 30, 61] Studies focusing on conduct evaluated the effect of the percentage of dropouts[30] or dropouts under a prespecified threshold[6, 50, 74] on treatment effect. Studies focusing on a combination of reporting, conduct, and analysis evaluated the effect of intention-to-treat analysis (yes versus no or unclear) on treatment effect;[23, 30, 50, 61, 64, 74] compared the effect of trials with and without exclusions (unclear combined with no exclusions[22] or unclear combined with exclusions.[39, 49]) on treatment effect; or compared treatment effects for ratings based on the Cochrane risk of bias tool domain on incomplete reporting, with one major additional specification of considering low or unclear risk of bias in studies with less than 90 percent followup.[52, 67]

Two studies did not clearly describe measures of completeness of data. One study presented the judgment on completeness of followup as a dichotomous measure but did not explain how this dichotomous judgment was assigned.[25] The underlying Linde Internal Validity Scale appears to be a three-point (rather than two-point) scale for studies that completely, partially, or do not report withdrawals and dropouts; no explanation was provided for how the three-point scale was translated to a two-point dichotomous judgment. One study[47] distinguished between "yes" and "unclear" disclosure of withdrawals and dropouts and cited the Cochrane Handbook in doing so. The Cochrane risk-of-bias tool, however, uses the term "incomplete outcome data" and categorized the risk of bias in three ways: yes (low risk of bias), no (high risk of bias), and unclear (uncertain risk of bias).[*] These judgments of risk of bias rely on more than just whether or not trialists disclosed information on withdrawals and dropouts: they require the reviewer to infer whether missing data were likely to relate to the true outcome and whether imputation methods, when used, were appropriate. The study did not explain how the authors arrived at judgments of "yes" or "unclear" and how these categories related to the Cochrane risk-of-bias categories.

Population, Interventions, Comparators, and Outcomes of Included Studies

The 15 studies that evaluate the effect of attrition bias on treatment effect vary substantially in breadth of inclusion, size, and scope.

Two studies defined populations and interventions narrowly. One study was a review of 25 trials of 5,919 patients that examined the efficacy of adjunctive devices to prevent distal embolization during acute myocardial infarctions.[47] A second study was a meta-analysis of 17 trials comparing low-molecular-weight heparin with standard heparin for the prevention of postoperative thrombosis that compared 25 different scales and components.[23]

One study focused on the intervention, homeopathy, but did not restrict populations or specific types of homeopathy. Of the 119 studies that met inclusion criteria, 89 had sufficient data to be included in quantitative analysis.[25]

Three studies focused on populations but did not restrict interventions. One study found 33 eligible meta-analyses in the Cochrane Pregnancy and Childbirth Database; these meta-analyses included 250 trials and 62,091 patients.[22] A second study searched the Cochrane Library, MEDLINE, EMBASE, and CINAHL for systematic reviewers and meta-analyses on osteoarthritis of the knee or hip.[49] The study included 14 eligible meta-analyses with 167 trials and 41,170 patients. A third study evaluated all 429 unique consecutive Phase III trials with protocols published by eight National Cancer Institute Groups up to 2006.[61]

Some studies selected the entire eligible corpus of specific databases.[64, 74] Other approaches to selecting the sample used criteria such as size, specific clinical areas, access to previous bodies of work, and so on. One study identified 14 eligible meta-analyses with 190 trials, and 136,164 patients in a search for meta-analyses of large trials on Cochrane and PubMed.[28] One study selected four medical areas (cardiovascular disease, infectious disease, pediatrics, and surgery) to ensure a variety of conditions.[30] The authors then selected eligible meta-analysis by searching MEDLINE and Cochrane for specific dates or issues for three of the four clinical areas; for the fourth, they relied on the yield from a previous analysis by their group. Their final database included 26 meta-analysis and 276 trials. One study focusing on individual patient data included 133 randomized controlled trials and 21,905 patients from 14 meta-analyses of various interventions for bladder, brain, lung, ovarian, and esophageal cancer. They conducted risk of bias analyses for individual trials and for meta-analyses.[39] Some studies incorporated or built on past databases or collections of meta-analyses.[6, 67, 74]

Findings

Within and across different measures of attrition, little evidence consistently emerges of a precise and consistent effect of attrition on treatment effect (Table 13). The three studies examining attrition bias solely through measures of reporting (did the study provide the numbers and reasons for withdrawals and dropouts for each group?) found inconsistent and imprecise differences.[28, 30, 61]

Likewise, studies focusing solely on the effect of the percentage of dropouts[30] or dropouts under a prespecified threshold[6, 50, 74] on treatment effect were inconsistent and imprecise in their findings. Studies focusing on a combination of reporting, conduct, and analysis that evaluated the effect of intention-to-treat analysis on treatment effect also found inconsistent effects with overlapping confidence intervals.[23, 30, 50, 61, 64, 74] The study that collapsed "no" and "unclear" responses for exclusions found an increased association between exclusions and treatment effect, but as with other studies, the effect was imprecise.[22]

Evidence from two studies that collapsed "yes" and "unclear" responses for exclusion found that meta-analyses with exclusions tend to have different treatment effects,[39, 49] but one study notes that the direction and extent of bias are unpredictable.[49] The same study found that overall meta-analyses with exclusions of patients show a more beneficial treatment effect, but the confidence intervals around that estimate spanned the line of no difference and indicated a pronounced degree of heterogeneity. Subgroup analysis offered some insight into the source of heterogeneity: the impact of exclusions on estimates of treatment effects seemed most pronounced in "meta-analyses with large treatment benefits, meta-analyses on complementary interventions, and meta-analyses with a high degree of heterogeneity between trials. The study with access to individual patient data found no effect of exclusions on individual trials, but it did find a small effect in meta-analyses that resulted in overestimates of treatment benefit.[39]

The two studies using the Cochrane risk-of-bias definition of incomplete reporting had overlapping confidence intervals for all categories of bias.[5, 24] The two studies using dichotomous but undefined measures of complete followup[25] or disclosure of withdrawals also found no statistically significant association between their measures and treatment effect.[47]

No pattern emerged between the dates of publication of included studies and the results.

Table 13. Summary of attrition bias results by study

Type of Analysis	Study Identification	Attrition Bias: Results of Bias	Effect[a]
Measure based on reporting: were numbers and reasons for withdrawals and dropouts described for each group?	Balk et al., 2002[30]	Dropouts recorded: Yes vs. no for whether the number of dropouts was explicitly recorded or could be calculated: ROR = 1.26 (95 % CI, 0.87 to 2.05)	=
		Reason for dropouts given: Yes vs. no for reasons reported for dropouts: ROR = 0.93 (95 % CI, 0.77 to 1.13)	=
	Kjaergard et al., 2001[28]	Adequate vs. inadequate followup: ROR = 1.50 (95% CI, 0.80 to 2.78)	=
		Subgroup analysis for large vs. small trials	
		Large trials vs. small trials for inadequate followup: ROR = 0.72 (95% CI, 0.30 to1.71)	=
		Large trials vs. small trials for adequate followup: ROR = 0.58 (95% CI, 0.32 to 1.02)	=
	Mhasker et al., 2012[61]	RHR for description of dropouts: Publication: RHR = 0.91 (95% CI, 0.72 to 1.15)	=
		Protocol plus publication: RHR = 0.76 (95% CI, 0.37 to 1.55)	=
Measure based on conduct: percentage dropouts or acceptable dropouts based on a threshold	Balk et al., 2002[30]	Percentage dropouts: results are ROR for 1 percentage-point increase in dropouts) : ROR = 1.02 (95% CI, 0.94 to 1.12)	=
	van Tulder et al., 2009[50]	Acceptable dropout (≥20% for short-term, ≥30% for long-term outcomes) SMD = -0.13 (95% CI, -0.29 to 0.02)	=
	Hempel et al., 2011[74b]	Acceptable dropout (≥20% for short-term, ≥30% for long-term outcomes) for dataset 2 (EPC reports): SMD = 0.15 (95% CI, 0.01 to 0.29)	=
		Dataset 3 ("pro-bias"): ROR = 0.72 (95% CI, 0.59 to 0.88)	-
		Dataset 4 ("heterogeneity"): ROR = 1.02 (95% CI, 0.81 to 1.27)	=
	Savovic et al., 2012[6]	Trials with 20% of patients with missing outcome data vs. <20%	
		All outcomes: ROR = 1.07 (95% CI, 0.92 to 1.25)	=
		Mortality outcomes ROR= 01.07 (95% CI, 0.80 to 1.42)	=
		Other objective outcomes ROR= 1.35 (95% CI, 0.63 to 2.94)	=
		Subjective or mixed ROR= 1.03 (95% CI, 0.79 to 1.36)	=

Table 13. Summary of attrition bias results by study (continued)

Type of Analysis	Study Identification	Attrition Bias: Results of Bias	Effect[a]
Measure based on a combination of reporting, conduct, and analysis: intention to treat conducted vs. no intention to treat conducted or not reported	Juni et al., 1999[23]	ITT analysis not performed: RRR = 1.37 (95% CI, 0.92 to 2.03)	=
	Balk et al., 2002[30]	Yes vs. no for whether all analyzed patients were analyzed in the group to which they were originally allocated ("dropouts were allowed so long as the reasons for withdrawal were not related to the group to which they were assigned"): ROR = 0.91 (95% CI, 0.70 to 1.13)	=
	Siersma et al., 2007[13]	Estimates from multivariable multilevel analyses of ITT Logistic regression with random effect ROR: 0.92 (95% CI, 0.79 to 1.06) Weighted regression with random effect ROR: 1.01 (95% CI, 0.88 to 1.15)	= =
	van Tulder et al., 2009[50]	ITT conducted SMD: -0.10 (95% CI, -0.24 to 0.04)	=
	Hempel et al., 2011[74 b]	ITT conducted Dataset 2 (EPC reports): SMD = 0.05 (95% CI, -0.10 to 0.20) Dataset 3 ("pro-bias"): ROR = 0.91 (95% CI, 0.74 to 1.12) Dataset 4 ("heterogeneity"): ROR = 0.93 (95% CI, 0.76 to 1.14)	= = =
	Mhasker et al., 2012[61]	ITT analysis: Publication: RHR = 0.98 (95% CI, 0.90 to 1.07) Protocol plus publication:RHR = 0.96 (95% CI, 0.88 to 1.05)	= =
Measure based on a combination of reporting, conduct, and analysis: comparisons of studies without or not reporting exclusions compared with studies reporting exclusions	Schulz et al., 1995[22]	ROR=1.07 (95% CI, 0.94 to 1.21) p=0.32	=
Measure based on a combination of reporting, conduct, and analysis: comparisons of studies without exclusions compared with studies with or not reporting exclusions	Tierney et al., 2004[39]	For individual patient data, no clear indication that the exclusion of patients altered the results more in one direction than another (t=1.5337, p=0.13) For meta-analyses, tendency for HR of included patients to favor the treatment rather than HR of all patients (t=2.401, p=0.03); differences ranged between 1 and 5%	= +

Table 13. Summary of attrition bias results by study (continued)

Type of Analysis	Study Identification	Attrition Bias: Results of Bias	Effect[a]
Measure based on a combination of reporting, conduct, and analysis: comparisons of studies without exclusions compared with studies with or not reporting exclusions (continued)	Nuesch et al., 2009[49]	SMD = -0.13 (-95% CI, 0.29 to 0.04) p=0.13, high variability (τ2=0.07, p<0.001) SMD for no exclusions vs. exclusions meta-analyses with low and high heterogeneity Low: 0.03 (95% CI, -0.09 to 0.15) High: -0.52 (95% CI, -0.76 to -0.27 Interaction: (P < 0.001) SMD for no exclusions vs. exclusions meta-analyses with small and large treatment benefit Small: -0.03 (95% CI, -0.16 to 0.09) Large: -0.74 (95% CI, -1.02 to -0.46) Interaction: (P < 0.001) SMD for no exclusions vs. exclusions meta-analyses with and without a drug intervention Yes: -0.16 (95% CI, -0.41 to 0.09) No: -0.05 (95% CI, -0.17 to 0.07) SMD for no exclusions vs. exclusions meta-analyses with and without complementary medicine Yes: -0.59 (95% CI, -0.87 to -0.31) No: -0.01 (95% CI, -0.14 to 0.12) P for interaction <0.001	= = + = + = = + =
Measure based on a combination of reporting, conduct, and analysis: high, unclear, or low risk of incomplete reporting based on Cochrane risk-of-bias tool	Harting et al., 2012[67]	SES High: 0.95 (95% CI, 0.64 to 1.27) Unclear: 0.72 (95% CI, 0.48 to 0.95) Low: 0.62 (95% CI, 0.51 to 0.73) Overall meta-regression (p = 0.16) Stratified analysis: No variables associated with a risk of bias for incomplete data reporting.	=
	Hamm et al, 2010,[52]	High RoB: SES = 0.23 (95% CI, 0.07 to 0.39) Unclear RoB: SES = 0.45 (95% CI, 0.30 to 0.60) Low RoB: SES = 0.34 (95% CI, 0.21 to 0.46) SMD: (P > 0.05)	=

Table 13. Summary of attrition bias results by study (continued)

Type of analysis	Study Identification	Attrition Bias: Results of Bias	Effect[a]
Measure based on reporting (undefined): complete followup or disclosure of withdrawals	Linde et al., 1999[25]	Complete followup ROR (univariate) = 1.31(95% CI, 0.88 to 2.00)	=
		Complete followup ROR (multivariate) = 1.23 (95% CI, 0.85 to 1.77)	=
	Inaba et al., 2009[47]	Results for disclosure of withdrawals and dropouts: (1) Mortality Unclear disclosure: ES = 0.74 (95% CI, 0.44 to 1.25) Yes: ES = 0.90 (95% CI, 0.55 to 1.47)	=
		(2) Incomplete ST resolution Unclear disclosure: ES = 0.79 (95% CI, 0.65 to 0.95) Yes: ES = 0.75 (95% CI, 0.63 to 0.89)	=
		(3) Impaired Myocardial Blush Grade Unclear disclosure: ES = 0.74 (95% CI, 0.44 to 1.25) Yes : ES = 0.90 (95% CI, 0.55 to 1.47)	=

[a] +: exaggerated odds ratio from bias; =: no significant effect from bias; -: reduced odds ratio from bias
[b] This row does not repeat results from dataset 1 because they appear in the row for van Tulder.

Abbreviations: CI = confidence interval; EPC = Evidence-based Practice Center; ES = effect size; ITT = intention-to-treat; SES = standardized effect size; SMD = standardized mean difference; HR = hazard ratio; RHR = ratio of hazard ratios; RoB = risk of bias; ROR = ratio of odds ratios; RRR = ratio of risk ratios.

Reporting Bias: Selective Outcome Reporting

Description of Included Studies

Study Characteristics and Definition of Selective Outcome Reporting Bias

Attempts to identify the effect of selective outcome reporting on the treatment effect require a comparison of the pooled effects from studies with a higher risk of bias with studies with a lower risk of bias. Identifying studies at high risk of selective outcome reporting is challenging, precisely because they are not reported. Nine studies[36, 38, 51, 53] examined the effect of selective outcome reporting bias on treatment effect using five different approaches to identify studies with suspected selective outcome reporting (Table 14). One study examined differences in treatment effect between studies that were appropriately indexed and studies that were not.[38] Two studies compared protocols with publications to examine the differences in statistical significance between studies that fully reported outcomes in publications relative to their protocols and studies that incompletely reported outcomes in publications relative to their protocols.[36, 37] Two studies[51, 53] used the Outcome Reporting Bias in Trials (ORBIT) approach as a means to distinguish between high and low or no risk-of-bias studies by comparing outcomes reported in individual studies to outcomes reported across the body of evidence. Using a combination of inference and judgment, these studies assigned a risk of high, low, or no risk of bias of selective outcome reporting based on clarity around the measurement and analysis of the outcome, and whether analysis and reporting of the outcome was likely to be influenced by statistical significance of the results. Three studies evaluated the Cochrane risk of bias tool.[52, 54, 67] Reviewers rate studies as low risk of bias if the protocol is available and the published results

include all prespecified outcomes. Infrequently, reviewers may rate studies without protocols as low risk of bias if the published results clearly indicate that all prespecified outcomes were reported. Reviewers report a high risk of bias of selective outcome reporting for studies with a mismatch between protocols and publications, for studies providing insufficient information for a meta-analysis, or for studies that did not report outcomes that should have been reported. Reviewers may choose an unclear rating when information is insufficient to arrive at a judgment. Two studies compared treatment effects for high, low, and unclear risk of bias.[67] One study testing the effect of collapsing three categories into two, specifically looking at treatment effect differences when pooling studies with high or unclear risk of bias versus low compared with pooling high versus low and unclear risk of bias.[54]

The fifth approach to measure the effect of selective outcome reporting evaluated the change in treatment effect with a decreasing proportion of trials in a systematic review contributing to the meta-analysis.[65]

Population, Interventions, Comparators, and Outcomes of Included Studies

The indexing study examined differences in treatment effect between studies in which the outcome in question (mortality in prognostic studies that look at the status of the tumor suppressor protein TP 53 in patients with head and neck squamous cell cancer) was published and indexed with studies that were (1) published but not indexed and (2) published studies that suggested that mortality data had been collected, provided no usable information in the publication, but then provided Kyzas and colleagues with data ("retrieved" studies). Their inclusion criteria resulted in a census of all eligible studies from MEDLINE and EMBASE, but of the eligible 64 studies that alluded to mortality information, only 42 were included in the final analysis: 22 studies could not be included because primary investigators did not respond, or because primary investigators were unable to retrieve the raw data.

One study comparing publications to protocols began with all available protocols for trials approved by the Danish Scientific-Ethical Committees for Copenhagen and Frederiksberg in 1994 and 1995 for which published articles were found.[37] The authors tabulated outcomes from 122 publications from 102 trials and their protocols in a 2X2 table relating the level of outcome reporting (full versus incomplete) to the level of statistical significance ($p<0.05$ versus $p\geq0.05$). Only 48 percent of trialists (49/102) responded to surveys about unreported outcomes, and only 11 trialists provided information on whether their unreported outcomes were statistically significant. The authors then pooled the odds ratio for trials with nonzero rows or columns: 50 trials provided information for efficacy outcomes, and 18 for harms. When trials were dichotomized to contrast fully or partially reported outcome with qualitatively or unreported outcomes, 35 trials provided information for efficacy outcomes, 15 for harms.

The second study used the same approach to review available protocols for trials approved for funding by the Canadian Institute for Health Research from 1990 to 1998. They tabulated 2X2 tables for 48 trials. They had a much higher response rates from primary investigators than in the Danish study: as many as 90 percent responded to their questionnaires (43/48), and 77 percent (48) provided information about unreported outcomes. As with the Canadian study, when authors pooled the odds ratio for trials with nonzero rows or columns, the number of trials dropped. Thirty trials provided information on efficacy, and 4 on harms. When trials were dichotomized to contrast fully or partially reported outcome with qualitatively or unreported outcomes, 20 trials provided information for efficacy outcomes and 4 for harms.[36]

The first of the two ORBIT studies is a meta-epidemiological study examining the prevalence of selective outcome reporting from 50 or 51 Cochrane collaboration review groups published in three issues of the Cochrane Library (Issue 4, 2005, Issue 1, 2007, and Issue 2, 2007).[53] The authors classified 538 of 712 trials as lacking information on whether the outcome of interest was measured or analyzed. They contacted 167 trialists (31 percent) for further information and obtained responses from 65 authors (12 percent). The authors analyzed 81 in detail for an assessment of the impact of missing trial data; of these trials, 25 underwent sensitivity analysis to assess the potential change in treatment effect from selective outcome reporting bias. The second study presents the results of the approach for a single systematic review for a systematic review with 24 trials on intravenous and nebulized magnesium sulphate for acute asthma.[51] Twenty-two of these trials reported on the two outcomes of interest: pulmonary function and hospital admissions, and two trials were excluded for lack of outcomes. From these numbers, the authors judged that 10 more trials could have added information on the outcomes of interest. Nine of 10 trialists responded to requests for more information. The study authors went on to report sensitivity analysis for 2 meta-analyses of hospital admissions, one with 3 included trials for children and another with 6 included trials for adults.

Of the 3 studies using the Cochrane risk-of-bias tool, one selected 154 trials at random from 616 published in December 2006 that had been previously evaluated for quality,[67] the second selected 10 percent of pediatric trials (N=300) from the Cochrane Central Register of Controlled Trials published in 2007,[52] and the third selected 107 trials included in a previous systematic review of combination long-acting beta-agonists or inhaled corticosteroids for maintenance therapy in persistent asthma.[54]

The single study looking at the proportion of trials in a systematic review contributing to a meta-analysis selected from the first 56 meta-analyses found 10 that met the eligibility criterion of containing 10 or more RCTs from Issue 4 of the 2005 Cochrane library.[65]

Findings

The indexing study found evidence that published and indexed data (i.e., selectively reported data) were more likely to show an association than unreported or inadequately indexed data, but confidence intervals for the estimates of effect overlapped.[38]

Both studies evaluating the association between statistical significance and outcome reporting provide evidence that statistical significance is closely associated with selective reporting of outcomes.[36, 37]

Two studies using the ORBIT tool offer evidence on the question of the impact of selective outcome on treatment effect (Table 14).[51, 53] The largest of these, by Kirkham and colleagues, found evidence that as many as a third of the evaluated reviews were at risk of having their findings overturned because of risk of bias.[53] The other study, of a single systematic review, did not find evidence that trial results were at risk of being overturned because of selective outcome reporting, but their findings suggested that for at least one outcome, a single unpublished trial would overturn the outcome.[51]

By contrast, studies using the Cochrane risk-of-bias tool by and large did not demonstrate the effect of selective outcome reporting on treatment effect. This lack of effect appears to be unrelated to how uncertainty is treated: the two studies generating effect estimates for high, low, and unclear risk of bias showed overlapping treatment effects,[52, 67] and the study creating a dichotomous measure of selective outcome reporting did not consistently find an effect.[54]

The study evaluating the changes in treatment effect with decreasing proportions of trials contributing to the meta-analysis found modest increases in treatment effect.[65]

Table 14. Summary of selective outcome reporting bias results by study

Measure of Selective Reporting	Study Identification	Selective Outcome Reporting: Results of Bias	Effect[a]
Studies with outcome published and appropriately indexed vs. published and not indexed or not published or indexed	Chan et al., 2004[37]	Full vs. incomplete reporting	
		Efficacy outcomes (N=50), pooled OR = 2.4 (95% CI, 1.4 to 4.0)	+
		Harms outcomes (N=18), pooled OR = 4.7 (95% CI, 1.8 to 12.0)	+
		Fully or partially vs. qualitatively or not reported	
		Efficacy outcomes (N=35), pooled OR = 3.1 (95% CI, 1.7 to 5.9)	+
		Harms outcomes (N=15), pooled OR = 7.6 (95% CI, 2.3 to 25.0)	+
Comparisons of publications and protocols	Chan et al., 2004[36]	Full vs. incomplete reporting	
		Efficacy outcomes (N=30), pooled OR = 2.7 (95% CI, 1.5 to 5.0)	+
		Harms outcomes (N=4), pooled OR = 7.7 (95% CI, 0.5 to 111)	=
		Fully or partially vs. qualitatively or not reported	
		Efficacy outcomes (N=20), pooled OR = 5.1 (95% CI, 2.5 to 10.0)	+
		Harms outcomes (N=4), pooled OR = 12.3 (95% CI, 1.5 to 99)	+
	Kyzas et al., 2005[38]	Fixed effects estimate for association between TP53 and HNSCC for published and indexed studies (N=18): RR = 1.23 (95% CI, 1.06 to 1.43)	+
		Fixed effects RR estimate for association between TP53 and HNSCC for published, not indexed (N=13): RR = 1.1 (95% CI, 0.99 to 1.42)	=
		Fixed effects RR estimate for association between TP53 and HNSCC for studies with information retrieved from investigators (N=11): RR = 0.98 (95% CI, 0.81, 1.19)	=

Table 14. Summary of selective outcome reporting bias results by study (continued)

Measure of Selective Reporting	Study Identification	Selective Outcome Reporting: Results of Bias	Effect[a]
Outcome Reporting Bias in Trials (ORBIT)	Dwan et al., 2010[51]	Intravenous magnesium: Children Hospital admission: SR results: RR = 0.69 (95% CI, 0.53 to 0.90) from 3 studies; I2 = 17.7%; Favors intervention Studies suspected of ORB: 3; No. participants missing from M-A (%): 117 (50%); Sensitivity analysis: ORB alone: RR = 0.76 (95% CI, 0.58 to 0.99) Nebulised Magnesium: Adults Hospital admission: SR results: RR = 0.68 (95% CI, 0.46 to 1.02) from 6 studies, I22 = 0; Favors intervention Studies suspected of ORB: 1; No. participants missing from M-A: 74 (17%); Sensitivity analysis results w/ORB alone: RR = 0.76 (95% CI, 0.51 to 1.13); Study publication bias: 8	+ +
	Kirham et al., 2010[53]	SRs which included a single MA of the review primary outcome (and were therefore assessed of the impact of ORB): 81 Reviews which included at least one trial that had a high suspicion of ORB: 52/81 Of those 52 reviews with at least one trial with high suspicion of ORB: Reviews not assessed for ORB: 27/52 Reviews assessed for ORB using maximum bias bound sensitivity analysis because c primary outcome was measured, did not have 0 events, and did not have missing studies for reasons not related to ORB: 25/52 Of those 25 reviews assessed for ORB using sensitivity analysis: SRs with conclusions that were robust to ORB: 8 SRs (significant effect estimate remained significant). SRs with conclusions that were not robust to ORB: 8	One-third of evaluated reviews were at risk of inflated estimates of effect because of risk of bias
Cochrane risk of bias tool	Harting et al., 2012[67]	High: SES = 0.64 (95% CI, 0.23 to 1.05) Unclear: SES = 1.18 (95% CI, 0.61 to 1.75) Low: SES = 0.64 (95% CI, 0.54 to 0.74) Overall meta-regression (p = 0.46) Stratified analysis: For selective outcome reporting, surgical trials showed higher risk of bias (p= 0.02) than other types of interventions and studies with industry support had higher risk of bias (p = 0.01) than studies without industry support	=

Table 14. Summary of selective outcome reporting bias results by study (continued)

Measure of Selective Reporting	Study Identification	Selective Outcome Reporting: Results of Bias	Effect[a]
	Hamm et al, 2010,[52]	High RoB: SES = 0.52 (95% CI, 0.21 to 0.83) Unclear RoB: SES = 0.22 (95% CI, -0.19 to 0.64) Low RoB: SES = 0.37 (95% CI, 0.27 to 0.47) All CIs overlap	=
	Hartling et al., 2011[54]	Mean differences in effect estimates for FEV1 across domains and overall risk of bias, high/unclear vs. low risk of bias: MD = 0.01 (95% CI, -0.02 to 0.03)	=
		Mean differences in effect estimates for FEV1 across domains and overall risk of bias, high vs. low/unclear risk of bias: MD = 0.01 (95% CI, -0.04 to 0.06)	=
		Mean differences in effect estimates for symptom-free days across domains and overall risk of bias, high/unclear vs. low risk of bias: MD = 7.22 (95% CI, 3.68 to 10.75)	+
		Mean differences in effect estimates for symptom-free days across domains and overall risk of bias, high vs. low/unclear risk of bias: MD = 3.67 (95% CI, -8.5 to 15.84)	=
Proportion of trials in a systematic review contributing to meta-analysis estimates	Furukawa et al., 2009[65]	When outcomes favored the intervention, for each 10-point decrease in the percentage of contributing trials, increase in OR = 1.046 (95% CI, 1.004 to 1.090) SMD: = 0.041 (95% CI, 0.002 to 0.079 Percentage of RCTs contributing to the meta-analyses: odds ratio, mean (95% CI) <20: OR = 2.67 (1.81 to 3.94) 20 to <40: OR = 2.38 (1.72 to 3.28) 40 to <60: OR = 2.22 (1.64 to 3.01) 60 to <80: OR = 1.61 (1.36 to 1.90) ≥ 80: OR = 1.87 (1.52 to 2.31) Percentage of RCTs contributing to the meta-analyses: SMD (95% CI) <20: SMD = 0.64 (0.36 to 0.91) 20 to <40: SMD = 0.48 (0.3 to -0.62) 40 to <60: SMD = 0.56 (0.34 to 0.78) 60 to <80: SMD = 0.36 (0.22 to 0.51) ≥ 80: SMD = 0.31 (0.19 to 0.43)	+

* +: exaggerated effect from bias; =: no significant effect from bias; -: reduced effect from bias

Abbreviations: CI = confidence interval; FEV_1 = forced expiratory volume in 1 second; HNSCC = ; MA = meta-analysis; N = number; ORB = ORBIT; RCT = randomized controlled trial; RoB: risk of bias; RR = risk ratio; SMD = standardized mean difference; SR = systematic review.

Discussion

This systematic review has attempted to evaluate the impact of systematic error due to study conduct and reporting on the treatment effects estimated in randomized controlled trials (RCTs). We qualitatively synthesized the methods and results of 38 studies that separately considered approaches to evaluating the impact of protection against all or some of six bias domains as follows:

1. Selection bias at baseline through randomization (sequence generation and allocation concealment)
2. Confounding through design or analysis
3. Detection bias through assessor blinding and adequate statistical methods
4. Performance bias through fidelity to the protocol (protection against unintended interventions) and blinding (patient/provider and double)
5. Attrition bias through intention-to-treat or other approaches to accounting for loss of participants
6. Reporting bias through detection of selective outcome reporting

In summary, although a trend toward exaggeration of treatment effects was seen across the bodies of evidence for most biases, the magnitude and precision of the effect varied widely across studies. We generally found evidence that was precise and consistent in direction of effect for assessor and double blinding among studies with subjective outcomes, and for selective outcome reporting. Evidence was generally consistent in direction of effect but with variable precision across studies for allocation concealment, sequence generation, and assessor blinding of objective or mixed outcomes. In contrast, evidence was generally inconsistent and imprecise in relation to confounding, adequate statistical methods, fidelity to the protocol, patient/provider blinding, and attrition bias.

Key Considerations Across Studies

Our synthesis was complicated by a number of factors. First, studies employed different comparisons in evaluating the risk of a bias. This conclusion was also noted by Dechartres et al.,[77] who recently conducted a systematic review intended as an overview of studies (n=177) assessing the methodological quality or reporting of randomized trials; excluding studies that assessed the impact of methodological quality on treatment effect. They observed that the criteria used to assess bias was markedly heterogeneous and often insufficiently defined.[77] Our systematic review, focusing only on studies that evaluated the impact of a bias on treatment effect, observed a similar problem: different studies employed different definitions of the approach to protecting against a bias. For example, only some considered inadequate study conduct separately from lack of sufficient reporting. Among the three studies that combined previous meta-epidemiological datasets,[6, 42, 45, 62, 63] only one, Savovic et al.,[6, 62, 63] attempted to establish some general agreement across datasets in approach to assessment of bias. The varying methods used to assess bias may be an important source of heterogeneity that resulted in the inconsistency of finding statistically significant results.

Further, assessing bias must also rely on sufficient reporting in addition to the planned and actual conduct of the study. A recent study by Mhaskar et al.[61] compared risk-of-bias assessments in trial protocols relative to those reported in the subsequent publication from 429

cancer-related RCTs and found that the methods reported in the published manuscripts did not reflect the methods described in the protocols of the trials. These study authors found that associations between methodological quality and treatment effect did not persist when they evaluated study conduct based on protocols rather than published trials.[61] Because meta-epidemiological studies, such as those included in our review, generally rely on assessments of the risk of bias from meta-analyses in included studies, rather than more direct observations of the trials themselves, they risk conflating the relationship between reporting of bias and treatment effect with that of actual bias and treatment effect.

Standards for reporting study conduct have changed over time. The majority of studies included in our review obtained evidence from trials published prior to 1990, before the publication and adoption of the CONSORT reporting standards. To be in conformance with these standards, published study results are more likely to reflect actual trial conduct in RCTs published after their dissemination. Even so, two recent meta-epidemiological studies evaluating more recently published trials continued to show problems with reporting. One study[78] sampled trials published for a 1-month interval in 2011 in journals that subscribe to the CONSORT reporting standards and found that 78 percent were rated as unclear or insufficient to assess allocation concealment. Their attempt to contact authors for clarification to minimize misclassification yielded a poor response. A second study[79] evaluated RCTs in top-rated orthopedic journals for a 4-year period (2006 to 2010) and found inadequate reporting to judge the approach to sequence generation and allocation concealment in approximately half the studies; however, they also showed some improvement as a function of increasingly later year of publication.

Studies differed in the instruments used to evaluate a bias and because of this, their approach to making the evaluation. For example, the Jadad scale does not provide an explicit question for allocation concealment, even though this instrument is known to otherwise be reliable. Therefore, earlier studies generally added items to the Jadad scale to address allocation concealment. Some of the more recent meta-epidemiological studies in our systematic review used other instruments (and different criteria) to assess risk of bias, most notably the risk-of-bias tool developed by the Cochrane Collaboration. The Cochrane tool continues to evolve and improve with respect to operationalizing criteria to adequately measure bias categories. A number of studies have evaluated the reliability and validity of the Cochrane tool and suggested improvements in the criteria and approach to assessing risk of bias in general.[67, 69] These variations and ongoing improvements suggest that the different tools used to assess risk of bias are themselves a source of heterogeneity. Future research should attempt to establish consensus on the criteria for determining adequate, inadequate, and unclear status for the bias being evaluated.

Because this body of literature is primarily based on meta-epidemiological studies, it builds on multiple units of analysis (patients, trials, and meta-analyses). Each of these levels of combination has the potential for substantial heterogeneity. Many of the earlier studies combined trials or meta-analyses that examined different interventions, diseases, or outcomes. While this approach would increase the power of the analysis by increasing the sample size, the approach assumes that combining disparate trials would not negatively impact the estimate of the bias. Our results demonstrate that most of these sources of heterogeneity were not accounted for consistently in the meta-epidemiological studies and it could potentially obscure the relationship between a bias and treatment effect.

Further regarding patient-level heterogeneity, conclusions from meta-analyses about the absence of effect between the proportion of dropouts and treatment effect may be incorrect. Meta-epidemiological studies, by combining aggregate measures of dropouts at the meta-analysis level, may fail to account for variations within and between trials in which improved treatment effect and lower dropout rates were associated.[30]

Some of the more recent studies showed differential magnitude based on the type of outcome (mortality versus morbidity, objective vs. subjective) and intervention (drug versus nondrug, conventional versus complementary) categories. Some studies attempted to evaluate the impact of particular biases, adjusting for other factors, such as trial sample size, source of funding, and magnitude of effect (large versus small); these issues were studied infrequently and inconsistently and no consistent impact was observed. Further evaluation of the role of differing interventions, populations (diseases), outcomes, and other factors will be an important area to explore in future research.

Additional considerations that differed across studies and that may have caused accuracy and precision of results to differ include how comparisons were measured (ratio of odd ratios, difference in effect estimates, or some other approach) and approaches to modeling.

An additional limitation is the lack of evaluation of statistical power. Power calculations in this setting are challenging because of the hierarchical nature of the meta-epidemiological study design (patients within trials within meta-analyses within meta-epidemiological studies). The fact that meta-epidemiological studies may be underpowered was a motivation for the Bias in Randomised & Observational Studies (BRANDO) project (personal correspondence, Jonathan Sterne, 2012). If the meta-epidemiological studies are underpowered, nonsignificant findings may be because of lack of power, not lack of effect. Hempel et al.[35] presented evidence and concluded that although bias components may explain some heterogeneity, the degree of heterogeneity due to other factors is key in parsing out the effect of bias. The authors concluded that most meta-epidemiological studies are underpowered to detect the effects of bias. Furthermore, the Hempel study was conducted on individual trials, rather than at the meta-epidemiological level as was most of our included evidence. Even with individual trial data, Hempel et al. were unable to clearly and consistently discern the impact of individual bias components across multiple datasets. One conclusion is that larger, carefully constructed datasets of individual trials must be assembled and analyzed to thoroughly understand the effect of bias as distinct from other sources of treatment effect heterogeneity. Our results should be viewed in light of this caveat.

Meta-analysis level heterogeneity in the types of biases in included studies and failure to account for potential interactions and confounding among various sources of bias may have resulted in an overstatement or understatement of the extent that bias influences treatment effect. Only 7 of 35 studies examined interactions between various sources of bias: Schulz et al.,[22] Linde et al.,[25] Nuesch et al.,[48, 49]Wood et al.,[45] Liu et al.,[59] Hrobjartsson et al.,[60] and Savovic et al.[6, 62] For example, studies with inadequate allocation concealment may also be prone to problems with blinding. The meta-epidemiological studies also varied in the modeling approaches used to evaluate the potential relationship between bias and treatment effect, and it was often difficult to discern the exact specifications of the models that were fit. Thus, comparison across studies was difficult. The important consideration is that biases may be associated with each other and future research should attempt to disentangle possible relationships using appropriate statistical methods.

Finally, we attempted to evaluate the quality of meta-epidemiological or other included studies to understand whether the design, conduct, or analysis of these studies could have influenced whether they themselves found differences in treatment effects for biased primary trials. Our criteria focused on risk of bias (from selection of studies, adjustments for confounding, validity and reliability of bias measure), applicability (whether findings apply to other clinical areas), and precision (power calculations). Given the wide variability in study designs, populations, measures of risk of bias, and approaches to analyses, we were unable to discern patterns in treatment effect differences associated with the quality of the meta-epidemiological and other studies included as evidence.

Specific Biases: Key Considerations

Selection Bias: Sequence Generation and Allocation Concealment

We identified more studies evaluating allocation concealment than sequence generation as a means of adjusting for potential selection bias. Both biases showed a consistent trend toward exaggeration of treatment effect when these biases were present. However, differences reached statistical significance in less than half of the studies for allocation concealment and in very few for sequence generation. The magnitude of overestimation of the treatment effect was greatly variable across studies, ranging from 50 to 5 percent. Overall, our findings suggest that across studies inadequate sequence generation and allocation concealment may exaggerate the estimate of effect, but the magnitude varies considerably and evidence of precision is poor.

The studies evaluating the impact of allocation concealment and sequence generation bias were prone to all the challenges previously noted in this discussion, particularly with respect to the consistency of the bias definitions (and specific tools used to assess), the outcomes used to assess the impact of the bias, the year of trial publication, and other sources of heterogeneity (outcomes, interventions, populations, sample size, other characteristics). Future research in assessing the impact of allocation concealment and sequence generation on effect estimates would benefit from a more explicit description and presentation of the statistical modeling approach undertaken. For example, many studies did not clarify if all trials measured the outcome in the same direction (so that the odds ratios or difference in treatment effect showed the same direction of the effect of the intervention). As a further example, the factors included in the models should be clearly defined. Finally, assumptions about the model form should be made clear, such as whether the effect of a trial characteristic is assumed to be constant across meta-analyses, and evaluations of such assumptions should be conducted.[16]

Confounding

Only two studies evaluated the effect of confounding.[30, 50] Neither found a precise estimate of effect, but both focused on trials only. When imbalance in prognostic arms may occur by chance in trials, this is random error and therefore not related to bias; it may not be linked with treatment effect.

Performance Bias: Fidelity to Protocol, Unintended Interventions or Cointerventions

Two studies evaluating four databases found inconsistent effects in magnitude, direction, and statistical significance for fidelity to protocol and unintended interventions.[35, 50] Of the four

databases, only one found a statistically significant difference in treatment effects as a result of compliance or cointerventions.

Performance Bias: Patient/Caregiver or Provider Blinding

We found the evidence of an effect of either patient/caregiver or provider blinding to be inconsistent across studies. This, coupled with variation in the approach across studies, limits our reaching conclusions in relation to these potential protections against performance bias.

Detection Bias: Assessor Blinding

We anticipated that although outcome assessment may be vulnerable to assessor bias in measuring any outcome, it would be particularly vulnerable when the outcome being measured is subjective. We found some compelling evidence that lack of assessor blinding could lead to exaggerated effects. Most notably, evidence of a larger effect from lack of assessor blinding was found in Hrobjartsson et al.[60] who may have been better able to observe differences because the authors compared the effect within individual studies. Their subgroup findings, limited to studies with subjective outcomes, found more exaggerated differences. A second study measuring subjective outcomes (muscle strength) also found a significant and exaggerated effect across studies.[59]

However, our confidence in our conclusion is tempered by inconsistencies across the larger body of evidence; two large studies with multiple analyses did not find a consistent relationship, in significance or direction of effect. In Balk et al.[30] results were stratified by clinical area, not by whether the outcome was measured subjectively or objectively. The appropriate role of the Hempel et al.[35] study on our conclusions is more elusive because of different results across the four datasets. Effect estimates were not in a consistent direction, and only one of four separate datasets found a significant relationship; it was not the dataset limited to a subjective outcome (pain).

Detection Bias: Valid Statistical Measures

Only two studies considered the role of valid statistical measures on treatment effect and their approach to measurement differed. Neither measured the validity of the statistical measures against a gold standard, such as an independent evaluation of the difference in the results based on an approach that would be recommended by methodological experts and the one that was used by the researchers. Instead, one relied on reviewers' general notions of the quality of the approach, and the second considered prespecification of the criteria for evaluating statistical significance and power. Therefore, we consider the evidence to be insufficient for reaching any conclusions.

Performance/Detection Bias: Double Blinding

The empirical evidence was mixed in relation to supporting whether lack of double blinding significantly affects differences in the direction, size, or significance of outcomes in trials. However, our evaluation of double blinding could only be as good as the information presented in journal articles across a large number of studies. We found that not only did the definition of double blinding vary across studies (such as whether the combination that constituted the "double" included assessor, patient, and/or provider), but that the comparison categories were inconsistent as well (ranging from clearly not blinded to blinding unknown or not reported). We

cannot know if the general lack of statistically significant association we found between blinding groups is partially due to vaguely described or different definitions used across studies.

Lack of double blinding, similar to lack of assessor blinding, was related to an exaggeration of the intervention effect estimates when subjective outcomes were estimated. These findings suggest that in circumstances that provide greater room for individual judgment or preferences, blinding is critical.

Lack of double blinding was also more likely to be related to significantly exaggerated treatment effect in studies that were limited to particular clinical topic areas (such as obstetrics, pharmaceutical trials, or homeopathy). This may imply that combining studies into large databases, while improving the power for statistical testing, may also be creating unmeasured error or "noise" because of this heterogeneity across studies; this may be less likely to affect more limited subject areas. We do not have sufficient data to support either this explanation or an alternative, that double blinding could be more important in those specific clinical areas.

Lastly, much of the evidence on this bias came from studies that relied on the same databases of trials or various combinations of a relatively small number of databases of trials. If studies included in these databases are inherently different from studies generally, then the results are not generalizable.

Attrition Bias

The lack of precision in findings for attrition bias calls attention to issues of power, as discussed earlier in this section. Additionally, the empirical evidence on whether aspects of attrition bias influence treatment effect appears to be linked to the measures used. Most estimates of attrition yielded inconsistent and imprecise results. Measures of reporting (did the study provide the number and reasons for withdrawal and dropout for each group?[28, 30, 61]) and conduct (percentage of dropouts[30] or dropouts under a prespecified threshold[6, 50, 74]) did not result in consistent or precise estimates of treatment effect. Similarly, studies focusing on a combination of reporting, conduct, and analysis that evaluated the effect of intention-to-treat analysis on treatment effect also found inconsistent estimate of effects with overlapping confidence intervals,[23, 30, 50, 61, 64, 74] as did studies using the Cochrane risk-of-bias tool,[52, 67] dichotomous but undefined measures of complete followup,[25] or disclosure of withdrawals.[47] One study collapsing "no" and "unclear" responses for exclusions reported imprecise estimates of effect of exclusions on treatment effect.[22]

Only one measure of attrition bias found precise estimates for differences in treatment effect: both studies that collapsed "yes" and "unclear" responses for exclusion found that meta-analyses with exclusions tend to have different treatment effects.[39, 49] These effects may be more pronounced for meta-analyses with large treatment effects or with a high level of heterogeneity or for specific types of interventions.[49]

Meta-analyses of efficacy routinely evaluate whether studies have accounted for patient exclusions. The limited evidence from this analysis suggests focusing on measures of analysis of attrition, rather than reporting or study conduct alone.

The CONSORT extension for harms suggests that results for harms should also be analyzed based on intention-to-treat because such an approach maintains the original random assignment.[80] Intention-to-treat may be the most conservative approach when evaluating superiority trials that test a null hypothesis of no difference between study arms. Meta-analysis of equivalence trials (that test a null hypothesis of a difference between study arms) may require

a close examination of both per-protocol analysis and intention-to-treat analysis to ensure that the most conservative analysis is chosen.[46]

Reporting Bias: Selective Outcome Reporting

The findings of studies on outcome reporting bias suggest that authors of systematic reviews should expect, unless otherwise confirmed, that every review is at risk of bias from selective outcome reporting in the original studies. Four of five approaches to measuring selective outcome reporting suggest a difference in treatment effect as a result of selective outcome reporting; no study using the Cochrane risk-of-bias tool consistently found effects. One of these four approaches (testing the effect of the proportion of trials contributing to the meta-analysis) holds limited relevance for prospective application to new reviews. The other three approaches to identify the extent of the risk (comparing protocols to publications, evaluating the association between statistical significance tests and extent of reporting, using a combination of inference and judgment in comparing outcomes from individual studies to outcomes in the body of evidence) can be applied to future systematic reviews but have significant constraints. All approaches require outreach to authors of included studies. Response rates from authors varied in five included studies from 11 percent to 90 percent. Other studies evaluating the overall risk of bias have found similarly low response rates.[52] Another alternative, at least for trials, is to standardize data fields in registries such as ClinicalTrials.gov so that the required information is made public routinely, particularly if space limitations in journals constrain the complete reporting of outcomes. At the present time, existing data in ClinicalTrials.gov offer limited utility for evaluating selective outcome reporting.[81]

Each approach to address selective outcome reporting has constraints. Expanding searches to sidestep the potential bias from inadequate indexing (influenced by the choice of definitions by trialists) and including reviews of protocols could result in a vast expansion of the scope of work for systematic reviews; the gains from this expansion of scope of work remain unclear. Studies that evaluate the likelihood of full versus incomplete reporting based on statistical significance offer utility for assessing the prevalence of selective outcome reporting and its likely effect in aggregate. They do not offer a practical approach to assessing the risk of selective outcome reporting for individual studies. The ORBIT approach, which relies on a diversity of outcomes in the body of evidence to help identify holes in reporting for individual studies, may not work as well in relatively homogenous bodies of evidence that nonetheless may be at risk of selective outcome reporting (as, for instance, in a body of evidence on a single intervention supported by a single funder). As noted earlier, the task of classifying studies as high, low, or no risk of selective outcome reporting bias relies on inference and judgment; response rates from trial authors to validate these judgments are highly variable.

Limitations of the Review Process

As noted by other studies evaluating the effect of bias,[17] we may not have constructed efficient (highly sensitive and specific) search strategies because some eligible studies may not be indexed with the terms that were used in this systematic review. As a result, we may have missed other relevant studies. Additionally, we restricted our search to capture studies assessing bias within randomized trials. We excluded studies that focused solely on nonrandomized studies or compared nonrandomized studies with randomized studies.

We attempted to evaluate the quality of included studies. Our criteria focused on risk of bias (from selection of studies, adjustment for confounding, validity and reliability of bias measure), applicability (whether findings apply to other clinical areas), and precision (power calculations). Given the wide variability in study designs, populations, measures of risk of bias, and approaches to analyses, we were unable to discern patterns in outcomes associated with study quality.

Research Gaps and Implications for Future Research

These findings may assist review teams in developing approaches to evaluating risk of bias and strength of evidence and support other methods development projects. Although empirical evidence should be the foundation for the methods used within systematic reviews, in relation to protection from bias, we cannot conclude that lack of a finding of a significant relationship implies that a relationship does not exist. Reviewers should consider the magnitude of potential effects of bias when interpreting effects (both direction and magnitude) in reviews and meta-analyses. Future research advancing systematic review methods needs to address the gaps we have identified in this review. Risk-of-bias assessments are considered the cornerstone of the four key factors affecting the grading of strength of evidence, and validation of the impact of these assessments remains key.

Future meta-epidemiological studies should not assume that differing interventions, outcomes, or populations (diseases) will not affect the magnitude of impact of the presence of biases on treatment effects. Although the magnitude and direction of an effect are important for drawing conclusions from a review, for future meta-epidemiological studies to have the most impact on future trial design and implementation, these studies will need to be adequately powered to be able to account for various sources of heterogeneity. Power calculations would have to take into account the intra-cluster correlations at the trial and meta-analysis levels, akin to the approach taken to design a cluster-randomized trial with two levels of clustering, for example, for the design of a cluster-randomized trial of patients within clinician within hospital. Current meta-epidemiological studies could provide a range of correlation estimates with which to conduct these power calculations.

Future meta-epidemiologic studies evaluating the impact of these biases should have access to and build on a database of existing trials and observational studies such as the one developed by the BRANDO project. Such future studies would thus be conducted with the unit of analysis equal to the individual trial or observational study, rather than built from meta-analyses or at the meta-epidemiological study level. Future studies would increase their reliability by using methods that control for consistency in evaluations across trials, beyond differences in trial conduct. For example, the “noise” in an analysis could be lessened by limiting the risk of bias assessment to one instrument implemented through a dual review process, distinguishing between information that is not reported and not adequately performed, including only new studies that were subject to stricter reporting standards, and not attempting to evaluate complex concepts such as double blinding, where meeting the standard may not be consistent across studies.

Future studies in this area should also account for the potential for patient-level heterogeneity and ecological fallacy. Such fallacy can result from using average patient-level values at the study level, thereby ignoring patient-level variation; as such, meta-analysis of individual patient-data may be required.

Lastly, future studies should focus, when possible, on trials published after the dissemination of reporting standards. Trials or meta-analyses published after 1996 are less likely to have limitations in reporting; assessment of the presence or absence of the bias is more likely to be valid as a result.

Implications for Reviewers

Of greatest concern are the implications of our findings for systematic reviewers. Although we only found statistically significant differences by studies employing a minority of protections against bias, we do not conclude that reviewers should abandon these evaluations. We discussed a number of considerations that would have resulted in imprecision and noise in study findings, so generally we conclude that additional evidence and evaluation are needed.

In those instances where we did find relatively consistent and precise differences in effect (allocation concealment, sequence generation, lack of assessor blinding in evaluating subjective outcomes, and selective outcome reporting), reviewers should consider being more cautious in rating risk of bias in relation to these concerns. Reviewers may also consider the lack of these protections against potential biases on treatment effect by conducting sensitivity analyses prior to integrating estimates from studies with these concerns in their strength of evidence grades. Lastly, of particular interest were findings from the BRANDO study that failings in study conduct concerning two biases increased the size of the difference in the effect estimate beyond the additive effect of each of the individual biases.[6] This may imply a value in more cautiously evaluating studies that lack protections for multiple sources of bias, including biases without consistent or precise evidence of having an individual effect.

References

1. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott, Williams, & Wilkins; 2008.
2. West S, King V, Carey TS, et al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute- University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality; Apr 2002. http://archive.ahrq.gov/clinic/epcsums/strengthsum.pdf.
3. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii-iix, 1-173. PMID: 14499048.
4. Song F, Eastwood AJ, Gilbody S, et al. Publication and related biases. Health Technol Assess. 2000;4(10):1-115. PMID: 10932019.
5. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]: The Cochrane Collaboration; 2008.
6. Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012 Sep 18;157(6):429-38. PMID: 22945832.
7. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615-25. PMID: 15308962.
8. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol. 2012 Feb;65(2):163-78. PMID: 21959223.
9. Ioannidis JP. Excess significance bias in the literature on brain volume abnormalities. Arch Gen Psychiatry. 2011;68(8):773-80. PMID: 21464342.
10. Schwarzer G, Carpenter J, Rucker G. Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. J Clin Epidemiol. 2010;63(3):282-8. PMID: 19836925.
11. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. [Review] [46 refs]. BMJ. 2001;323(7304):101-5. PMID: 11451790.
12. Palma S, Delgado-Rodriguez M. Assessment of publication bias in meta-analyses of cardiovascular diseases. J Epidemiol Community Health. 2005 Oct;59(10):864-9. PMID: 16166360.
13. Relevo R, Balshem H. Finding evidence for comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1168-77. PMID: 21684115.
14. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. AHRQ Publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/
15. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ. 1997 Sep 13;315(7109):617-9. PMID: 9310553.
16. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11):1513-24. PMID: 12111917.
17. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. [Review][Update of Cochrane Database Syst Rev. 2007;(2):MR000012; PMID: 17443633]. Cochrane Database of Systematic Reviews. 2011;4:MR000012. PMID: 21491415.

18. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007;7:10. PMID: 17302989.

19. Bausell RB, Lee WL, Soeken KL, et al. Larger effect sizes were associated with higher quality ratings in complementary and alternative medicine randomized controlled trials. J Clin Epidemiol. 2004;57(5):438-46. PMID: 15196613.

20. Contopoulos-Ioannidis DG, Gilbody SM, Trikalinos TA, et al. Comparison of large versus smaller randomized trials for mental health-related interventions. Am J Psychiatry. 2005 Mar;162(3):578-84. PMID: 15741476.

21. Bafeta A, Dechartres A, Trinquart L, et al. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. BMJ. 2012;344:e813. PMID: 22334559.

22. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12. PMID: 7823387.

23. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282(11):1054-60. PMID: 10493204.

24. Nurmohamed MT, Rosendaal FR, Buller HR, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet. 1992 Jul 18;340(8812):152-6. PMID: 1352573.

25. Linde K, Scholz M, Ramirez G, et al. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol. 1999;52(7):631-6. PMID: 10391656

26. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess. 1999;3(12):i-iv, 1-98. PMID: 10374081.

27. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352(9128):609-13. PMID: 9746022.

28. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135(11):982-9. PMID: 11730399.

29. Kjaergard LL, Villumsen J, Glund C. Correction: reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2008 5 August;149(3):219.

30. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA. 2002;287(22):2973-82. PMID: 12052127.

31. Balk EM, Bonis PAL, Lau J, et al. Allocation concealment in clinical trials (response). JAMA. 2002;288(19):2406-9.

32. Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? Results of a pilot study. BMC Health Services Research. 2002;2(1):18. PMID: 12213183.

33. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA. 2003;290(7):921-8. PMID: 12928469.

34. Egger M, Juni P, Bartlett C, et al. 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. PMID: 12583822

35. Hempel S, Miles J, Suttorp M, et al. Detection of Associations between Trial Quality and Effect Sizes. Methods Research Report. Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I. AHRQ Publication No. 12-EHC010-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2012. PMID: 22359777

36. Chan AW, Krleza-Jeric K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ. 2004;171(7):735-40. PMID: 15451835.

37. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291(20):2457-65. PMID: 15161896.

38. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst. 2005;97(14):1043-55. PMID: 16030302.

39. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol. 2005;34(1):79-87. PMID: 15561753.

40. Derry CJ, Derry S, McQuay HJ, et al. Systematic review of systematic reviews of acupuncture published 1996-2005. Clin Med. 2006;6(4):381-6. PMID: 16956145.

41. Mohammed S, Goodacre S. Intravenous and nebulised magnesium sulphate for acute asthma: systematic review and meta-analysis. Emerg Med J. 2007 Dec;24(12):823-30. PMID: 18029512.

42. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol. 2007;36(4):847-57. PMID: 17517809.

43. Als-Nielsen B, Chen W, Gluud LL, et al. Are trial size and reported methodological quality associated with treatment effects? Program & Abstract Book, 12th Cochrane Colloquium 2004. 2004:102-3.

44. Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol. 2008;35(9):775-82. PMID: 18840153.

45. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5. PMID: 18316340.

46. Treadwell J, Uhl S, Tipton K, et al. Assessing Equivalence and Noninferiority. Methods Research Report. (Prepared by the EPC Workgroup under Contract No. 290-2007-10063.). AHRQ Publication No. 12-EHC045-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov

47. Inaba Y, Chen JA, Mehta N, et al. Impact of single or multicentre study design on the results of trials examining the efficacy of adjunctive devices to prevent distal embolisation during acute myocardial infarction. Eurointervention. 2009;5(3):375-83. PMID: 19736164.

48. Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum. 2009;61(12):1633-41. PMID: 19950329.

49. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339:b3244. PMID: 19736281.

50. van Tulder MW, Suttorp M, Morton S, et al. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine (Phila Pa 1976). 2009 Jul 15;34(16):1685-92. PMID: 19770609.

51. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials [Electronic Resource]. 2010;11:52. PMID: 20462436.

52. Hamm MP, Hartling L, Milne A, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. BMC Pediatr. 2010;10:96. PMID: 21176224.

53. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365. PMID: 20156912.

54. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. [Review]. PLoS ONE [Electronic Resource]. 2011;6(2):e17242. PMID: 21390219.

55. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations Between Trial Quality and Effect Sizes. Methods Research Report (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I). AHRQ Publication No. 11-EHC045-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2011. http://effectivehealthcare.ahrq.gov. PMID: 21834174

56. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. [Review] [60 refs]. J Clin Epidemiol. 2006;59(12):1249-56. PMID: 17098567.

57. Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. J Clin Epidemiol. 2011;64(10):1070-5. PMID: 21474279.

58. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol. 2006;59(12):1249-56. PMID: 17098567

59. Liu CJ, LaValley M, Latham NK. Do unblinded assessors bias muscle strength outcomes in randomized controlled trials of progressive resistance strength training in older adults? Am J Phys Med Rehabil. 2011 Mar;90(3):190-6. PMID: 21173683.

60. Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012;344:e1119. PMID: 22371859.

61. Mhaskar R, Djulbegovic B, Magazin A, et al. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012 Jun;65(6):602-9. PMID: 22424985.

62. Savović J, Harris RJ, Wood L, et al. Development of a combined database for meta-epidemiological research. Research Synthesis Methods. 2010;1(3-4):212-25.

63. Savovic J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess. 2012 Sep;16(35):1-82. PMID: 22989478.

64. Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat.Med. 2007;26(14):2745-58. PMID: 17117373

65. Furukawa TA, Watanabe N, Omori IM, et al. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. JAMA. 2007;297(5):468-70. PMID: 17284696

66. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ. 2009;339:b4012. PMID: 19841007.

67. Hartling L, Hamm M, Milne A, et al. Validity and inter-rater reliability testing of quality assessment instruments. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021-I.) AHRQ Publication No. 12-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm PMID: 22536612

68. Lim B, Manheimer E, Lao L, et al. Acupuncture for treatment of irritable bowel syndrome. Cochrane Database Syst Rev. 2006;Issue 4(4):CD005111. PMID: 17054239.

69. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol. 2013 Sep;66(9):973-81. PMID: 22981249.

70. Linde K, Ramirez G, Mulrow CD, et al. St John's wort for depression--an overview and meta-analysis of randomised clinical trials. BMJ. 1996 Aug 3;313(7052):253-8. PMID: 8704532.

71. Agarwal R, Aggarwal AN, Gupta D. Role of noninvasive ventilation in acute lung injury/acute respiratory distress syndrome: a proportion meta-analysis. Respir Care. 2010;55(12):1653-60. PMID: 21122173.

72. Oakley JE, Brennan A, Tappenden P, et al. Simulation sample sizes for Monte Carlo partial EVPI calculations. J Health Econ. 2010;29(3):468-77. PMID: 20378190.

73. Dave M, Elmunzer BJ, Dwamena BA, et al. Primary sclerosing cholangitis: meta-analysis of diagnostic performance of MR cholangiopancreatography. Radiology. 2010;256(2):387-96. PMID: 20656832.

74. Hempel S, Miles J, Suttorp M, et al. Detection of Associations between Trial Quality and Effect Sizes. Methods Research Report. Prepared by the Southern California Evidence-based Practice Centerunder Contract No. 290-2007-10062-I. AHRQ Publication No. 12-EHC010-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2012. PMID: 22359777

75. Furlan AD, Tomlinson G, Jadad AA, et al. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. Spine (Phila Pa 1976.). 2008;33(3):339-48. PMID: 18303468

76. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996 Feb;17(1):1-12. PMID: 8721797.

77. Dechartres A, Charles P, Hopewell S, et al. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol. 2011;64(2):136-44. PMID: 20705426

78. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ. 2013;346:f1135. PMID: 23518273.

79. Chess LE, Gagnier J. Risk of bias of randomized controlled trials published in orthopaedic journals. BMC Med Res Methodol. 2013;13:76. PMID: 23758875.

80. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004;141(10):781-8. PMID: 15545678.

81. Norris SL, Holmer HK, Ogden LA, et al. Selective outcome reporting as a source of bias in reviews of comparative effectiveness [Internet]. Agency for Healthcare Research and Quality. Report No.: 12-EHC110-EF. Rockville, MD: Aug 2012. PMID: 22993870

Appendix A. Empirical Evidence of Bias Search Strategies

All databases were searched from 1980 to September 24, 2012.

Medline-OVID

January 19 2012

1. attrition bias*.tw.
2. selection bias*.tw.
3. detection bias*.tw.
4. performance bias*.tw.
5. reporting bias*.tw.
6. data bias*.tw.
7. mortality bias*.tw.
8. contamination bias*.tw.
9. differential treatment effect?.tw.
10. ecologic bias*.tw.
11. aggregate bias*.tw.
12. cross-level bias*.tw.
13. information bias*.tw.
14. publication bias*.tw.
15. suppression bias*.tw.
16. significance bias*.tw.
17. ((observer or intra-observer or intraobserver or inter-observer or interobserver) adj (bias* or variation*)).tw.
18. Co-intervention bias*.tw.
19. timing bias*.tw.
20. Compliance bias*.tw.
21. Withdrawal bias*.tw.
22. Proficiency bias*.tw.
23. performance bias*.tw.
24. or/1-23
25. exp in vitro study/
26. simulation/ or computer simulation/
27. simulation.tw.
28. 26 or 27
29. exp "bias (epidemiology)"/
30. 24 or 29
31. computer simulation/
32. simulation.tw.
33. 31 or 32
34. meta-analysis.pt,ti,ab,sh.
35. (meta anal$ or metaanal$).ti,ab,sh.
36. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ti.
37. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ab.
38. ((pool$ or combined or combining) adj (data or trials or studies or results)).ti,ab.
39. (medline or embase or cochrane or pubmed or pub med).ti,ab.

40. or/37-39
41. review.pt,sh.
42. 40 and 41
43. or/34-36
44. 42 or 43
45. exp DNA/
46. exp Genetics/
47. (genom* or genetic).ti.
48. or/45-47
49. "Quality Control"/
50. *"Meta-Analysis as Topic"/
51. 49 and 50
52. *"Randomized Controlled Trials as Topic"/
53. 49 and 52
54. 51 or 53
55. 30 not (25 or 48)
56. limit 55 to meta analysis
57. 33 and 55
58. 44 and 55
59. 56 or 57 or 58
60. limit 59 to english language
61. animals/
62. 60 not 61
63. limit 62 to (comment or editorial or in vitro or letter or video-audio media or webcasts)
64. 62 not 63
65. 54 or 59
66. limit 65 to english language
67. animals/
68. 66 not 67
69. limit 68 to (comment or editorial or in vitro or letter or video-audio media or webcasts)
70. 68 not 69

Embase-OVID
January 19 2012
1. attrition bias*.tw.
2. selection bias*.tw.
3. detection bias*.tw.
4. performance bias*.tw.
5. reporting bias*.tw.
6. data bias*.tw.
7. mortality bias*.tw.
8. contamination bias*.tw.
9. differential treatment effect?.tw.
10. ecologic bias*.tw.
11. aggregate bias*.tw.
12. cross-level bias*.tw.

13. information bias*.tw.
14. publication bias*.tw.
15. suppression bias*.tw.
16. significance bias*.tw.
17. ((observer or intra-observer or intraobserver or inter-observer or interobserver) adj (bias* or variation*)).tw.
18. Co-intervention bias*.tw.
19. timing bias*.tw.
20. Compliance bias*.tw.
21. Withdrawal bias*.tw.
22. Proficiency bias*.tw.
23. performance bias*.tw.
24. or/1-23
25. *systematic error/ or external bias/ or internal bias/ or interview bias/ or nonresponse bias/ or observer bias/ or recall bias/
26. 24 or 25
27. exp in vitro study/
28. simulation/ or computer simulation/
29. simulation.tw.
30. 28 or 29
31. meta analysis/
32. meta-analysis.ti,ab.
33. (meta anal$ or metaanal$).ti,ab.
34. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ti.
35. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ab.
36. ((pool$ or combined or combining) adj (data or trials or studies or results)).ti,ab.
37. (medline or embase or cochrane or pubmed or pub med).ti,ab.
38. or/35-37
39. review.pt,sh.
40. 38 and 39
41. or/31-34
42. 40 or 41
43. exp *DNA/
44. exp genetics/
45. polymorphism.ti.
46. (laboratory or animal).ti.
47. or/43-46
48. "quality control"/
49. "meta analysis"/
50. 48 and 49
51. "statistical analysis"/
52. 50 and 51
53. 30 or 42
54. 27 or 47
55. 26 not 54
56. 53 and 55

57. 52 or 56
58. limit 55 to (meta analysis or "systematic review")
59. 57 or 58
60. limit 59 to english language
61. animals/
62. 60 not 61
63. limit 62 to (book or book series or conference abstract or editorial or letter or note)
64. 62 not 63

Cochrane-Wiley
January 19, 2012
All reviews by Cochrane Methods Group

Appendix B. Screening and Abstraction Forms

B-1. Screening Forms

Level 1 Title and Abstract Screening Form

1. Does this study aim to assess the effect of bias (quality) on study effect size OR the validity of it conclusions?
   ***note: quality is sometimes used as a synonym for bias**

   - o Yes/Maybe (include)
   - o No (exclude)

2. Keep this paper as background

   - o Yes
     Clear Response

3. Comment:

Full Text Screening Level 2

1. This citation should be excluded for the following reason:

   - o Not in English (stop, submit)
   - o Published before 1980 (stop, submit)
   - o In vitro study only (stop, submit)
   - o Medical tests only (screening, diagnosis, genetic testing) (stop, submit)
   - o Simulation study only
   - o None of the above (continue)

2. Does this paper evaluate one or more of the following: (check all that apply)

   - o Selection bias- e.g. randomization, allocation, concealment, confounders (continue)
   - o Attrition bias- e.g. dropout completeness of data (continue)
   - o Detection bias- e.g. blinding outcome assessors, valid statistical methods (continue)
   - o Performance bias- e.g. blinding participants and personnel, fidelity to protocol, unintended interventions (continue)
   - o Reporting bias- e.g. selective outcome reporting (continue)
   - o Publication bias only (stop and submit)
   - o Other bias-identify (continue)
   - o Other global bias (stop and exclude)
   - o Wrong design: not designed to primarily look at effect of bias on effect size (stop and exclude)

3. Does this paper include the following outcomes:

   - o Change in effect size and directions
   - o Changes in estimates and variance
   - o Estimates of heterogeneity ONLY

4. Is this paper useful for background?

   - o Yes

5. Comments?

B-2. Full Text Abstraction Form – ID, Study chx, baseline chx section

RefID	**First author's last name**	**Year**	**Publication title**	**Describe goal of the study**	**Name of disease(s) or condition(s)** (list all)	**Source(s) of bias examined** (list all)	**Outcomes selected** (list outcomes, include details on how outcomes were selected)	**Inclusion/exclusion Criteria** (e.g, databases searched, dates of inclusion, type of clinical condition, size of meta-analysis, type of outcome [dichotomous outcomes required], level of heterogeneity)

Unit of analysis (study with MA nesting, or study without accounting for MA nesting)	**N MA Eligible** **N Trial Eligible** **If N varies by clinical condition then list by clinical condition**	**N MA Included** **N Trial Included** **If N varies by clinical condition then list by clinical condition**	**N MA Analyzed** **N Trial Analyzed** **If N varies by clinical condition then list by clinical condition**	**Funding source**	**Add comments or specify "other" entries** here (clarify which column the "other" entry belongs to)

B-3. Full Text Abstraction Form – Results section

RefID	**First author's last name**	**Year**	**Publication title**	**Selection bias: adequate randomization (Yes/No)**	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value** (Include any subanalyses)	**Were the results stratified by clinical condition, outcome, or other variable? OR Did the study account for other sources of bias?** (e.g., did the study consider blinding only for studies which had adequate allocation concealment ?)

Selection bias: allocation concealment (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Selection bias: Confounders accounted for (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Detection bias: blinding (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Detection bias: Valid statistical methods (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Performance bias: fidelity to protocol (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Performance bias: unintended interventions or co-interventions (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)	**Inter-rater reliability**	**Statistics used to measure outcome (e.g., OR, Ratio of odd ratios)**	**Modeling approach (e.g, Bayesian)**	**Results 95% CI P value**	**Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? (e.g., did the study consider blinding only for studies which had adequate allocation concealment?)**

Reporting bias: selective outcome reporting (Yes/No)	**Measurement of bias** (Scale or question used)	**Method of measurement** (dual review or single review)Inter-rater reliability Statistics used to measure outcome (e.g., OR, Ratio of odd ratios) Modeling approach (e.g., Bayesian) Results 95% CI P value Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? **(e.g., did the study consider blinding only for studies which had adequate allocation concealment?)** Attrition bias: **ITT; dropouts; attrition bias; completeness of data (lumped together) (Yes/No) Measurement of bias** (Scale or question used) Method of measurement (dual review or single review) Inter-rater reliability Statistics used to measure outcome (e.g., OR, Ratio of odd ratios) Modeling approach (e.g., Bayesian) Results 95% CI P value Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? **(e.g., did the study consider blinding only for studies which had adequate allocation concealment?)** Other bias (Describe): Multicentre study; study country; etc. **Measurement of bias** (Scale or question used) Inter-rater reliability Statistics used to measure outcome (e.g., OR, Ratio of odd ratios) Modeling approach (e.g, Bayesian) Results 95% CI P value Were the results stratified by clinical condition, outcome, or other variable or did the study account for other sources of bias? **(e.g., did the study consider blinding only for studies which had adequate allocation concealment?)** Comments

B-3. Full Text Abstraction Form – Quality questions + AMSTAR section

RefID	First author's last name	Year	Trial name (if applicable)

Were studies selected by a census? If no, how was the sample selected? (options: Randomly, purposefully, or unknown). Describe.	Selection bias: Did the study account for confounding or interaction among sources of bias? (Y/N)	Detection bias: Did the study use dichotomous measures for high vs. low quality? (Y/N)	Detection bias: If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	Detection bias: IRR of risk of bias measure	Detection bias: validity of risk of bias measure	Precision: How was sample size calculated?

Do the findings of this meta-epidemiological study apply to multiple clinical areas?	**Analysis: did the study account for duplication of trials?**	**IS THIS STUDY A SYSTEMATIC REVIEW**	**Was an 'a priori' design provided?** The research question and inclusion criteria should be established before the conduct of the review.	**Was there duplicate study selection and data extraction?** There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	**Was the status of publication (i.e. grey literature) used as an inclusion criterion?** The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.	**Was a list of studies (included and excluded) provided?** A list of included and excluded studies should be provided.	**Were the characteristics of the included studies provided?** In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.	**Was the scientific quality of the included studies assessed and documented?** 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	**Were the methods used to combine the findings of studies appropriate?** For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).	**Was the likelihood of publication bias assessed?** An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).	**Was the conflict of interest stated?** Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

Appendix C. Excluded Full-Text Article List

1. Adams AS, Soumerai SB, Lomas J, et al. Evidence of self-report bias in assessing adherence to guidelines. [Review] [28 refs]. Int J Qual Health Care. 1999;11(3):187-92. PMID: 10435838.
2. Alexander DD, Mink PJ, Cushing CA, et al. A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer. [Review]. Nutrition Journal. 2010;9:50. PMID: 21044319.
3. Anderson CA, Shibuya A, Ihori N, et al. Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review. [Review] [79 refs]. Psychol Bull. 2010;136(2):151-73. PMID: 20192553.
4. Baird DD, Weinberg CR, Rowland AS. Reporting errors in time-to-pregnancy data collected with a short questionnaire. Impact on power and estimation of fecundability ratios. Am J Epidemiol. 1991;133(12):1282-90. PMID: 2063836.
5. Baldi I, Ponti A, Zanetti R, et al. The impact of record-linkage bias in the Cox model. J Eval Clin Pract. 2010;16(1):92-6. PMID: 20367819.
6. Barden J, Derry S, McQuay HJ, et al. Bias from industry trial funding? A framework, a suggested approach, and a negative result. Pain. 2006;121(3):207-18. PMID: 16495012.
7. Baron G, Ravaud P, Samson A, et al. Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: a simulation study. Arthritis Rheum. 2008;59(1):25-31. PMID: 18163406.
8. Barone BB, Yeh HC, Snyder CF, et al. Postoperative mortality in cancer patients with preexisting diabetes: systematic review and meta-analysis. [Review] [39 refs]. Diabetes Care. 2010;33(4):931-9. PMID: 20351229.
9. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. [Review] [22 refs]. JAMA. 2010;303(12):1180-7. PMID: 20332404.
10. Bassler D, Ferreira-Gonzalez I, Briel M, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. [Review] [9 refs]. J Clin Epidemiol. 2007;60(9):869-73. PMID: 17689802.
11. Berard A, Andreu N, Tetrault J, et al. Reliability of Chalmers' scale to assess quality in meta-analyses on pharmacological treatments for osteoporosis. Ann Epidemiol. 2000;10(8):498-503. PMID: 11118928.
12. Biondi-Zoccai GG, Lotrionte M, Abbate A, et al. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. [Review] [43 refs]. BMJ. 2006;332(7535):202-9. PMID: 16415336.
13. Bjordal JM. A quantitative study of bias in systematic reviews. Advances in Physiotherapy. 2003;5(2):83-96.
14. Boccia S, De FE, Galli P, et al. A systematic review evaluating the methodological aspects of meta-analyses of genetic association studies in cancer research. Eur J Epidemiol. 2010;25(11):765-75.
15. Boyer K, Wies J, Turkelson CM. Effects of bias on the results of diagnostic studies of carpal tunnel syndrome. [Review] [41 refs]. Journal of Hand Surgery - American Volume. 2009;34(6):1006-13. PMID: 19446966.
16. Braunholtz DA, Edwards SJ, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". [Review] [26 refs]. J Clin Epidemiol. 2001;54(3):217-24. PMID: 11223318.
17. Briel M, Lane M, Montori VM, et al. Stopping randomized trials early for benefit: a protocol of the Study Of Trial Policy Of Interim Truncation-2 (STOPIT-2). Trials [Electronic Resource]. 2009;10:49. PMID: 19580665.
18. Brooks JM, Fang G. Interpreting treatment-effect estimates with heterogeneity and choice: simulation model results. Clin Ther. 2009;31(4):902-19. PMID: 19446162.

19. Brunoni AR, Amadera J, Berbel B, et al. A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. International Journal of Neuropsychopharmacology. 2011;14(8):1133-45.

20. Canter PH, Ernst E. Sources of bias in reviews of spinal manipulation for back pain. Wi_n Klin Wochenschr. 2005;117(9-10):333-41. PMID: 15989112.

21. Carpenter JR, Schwarzer G, Rucker G, et al. Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses. J Clin Epidemiol. 2009;62(6):624-31. PMID: 19282148.

22. Chootrakool H, Shi JQ, Yue R. Meta-analysis and sensitivity analysis for multi-arm trials with selection bias. Stat Med. 2011;30(11):1183-98. PMID: 21538449.

23. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. [Review] [81 refs]. J Clin Epidemiol. 2010;63(5):502-12. PMID: 18823754.

24. Colliver JA, Kucera K, Verhulst SJ. Meta-analysis of quasi-experimental research: are systematic narrative reviews indicated? Med Educ. 2008;42(9):858-65. PMID: 18715482.

25. Cook RJ, Wei W. Selection effects in randomized trials with count data. Stat Med. 2002;21(4):515-31. PMID: 11836733.

26. Cronin AM, Vickers AJ. Statistical methods to correct for verification bias in diagnostic studies are inadequate when there are few false negatives: a simulation study. BMC Medical Research Methodology. 2008;8:75. PMID: 19014457.

27. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. III: The issue of carry-over. Stat Med. 2002;21(15):2161-73. PMID: 12210631.

28. De VC, Manzoli L, Marzuillo C, et al. A systematic review evaluating the potential for bias and the methodological quality of meta-analyses in vaccinology. [Review] [123 refs]. Vaccine. 2007;25(52):8794-806. PMID: 18035456.

29. Dechartres A, Boutron I, Trinquart L, et al. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med. 2011;155(1):39-51. PMID: 21727292.

30. Dechartres A, Charles P, Hopewell S, et al. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. [Review]. J Clin Epidemiol. 2011;64(2):136-44. PMID: 20705426.

31. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii-iix, 1-173. PMID: 14499048.

32. Devereaux PJ, Choi PT, El-Dika S, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004;57(12):1232-6. PMID: 15617948.

33. Diamond GA. Affirmative actions: can the discriminant accuracy of a test be determined in the face of selection bias? Med Decis Making. 1991;11(1):48-56. PMID: 2034075.

34. Dias S, McNamee R, Vail A. Bias in frequently reported analyses of subfertility trials. Stat Med. 2008;27(27):5605-19. PMID: 18693327.

35. Doucet M, Sismondo S. Evaluating solutions to sponsorship bias. J Med Ethics. 2008;34(8):627-30. PMID: 18667655.

36. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. [Review] [45 refs]. PLoS ONE [Electronic Resource]. 2008;3(8):e3081. PMID: 18769481.

37. Engels EA, Schmid CH, Terrin N, et al. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. [Review] [40 refs]. Stat Med. 2000;19(13):1707-28. PMID: 10861773.

38. Fergusson D, Laupacis A, Salmi LR, et al. What should be included in meta-analyses? An exploration of methodological issues using the ISPOT meta-analyses. [Review] [33 refs]. Int J Technol Assess Health Care. 2000;16(4):1109-19. PMID: 11155831.

39. Fisk WJ, Eliseeva EA, Mendell MJ. Association of residential dampness and mold with respiratory tract infections and bronchitis: a meta-analysis. [Review]. Environmental Health: A Global Access Science Source. 2010;9:72. PMID: 21078183.

40. Gartlehner G, Morgan L, Thieda P, et al. The effect of study sponsorship on a systematically evaluated body of evidence of head-to-head trials was modest: secondary analysis of a systematic review. [Review] [72 refs]. J Clin Epidemiol. 2010;63(2):117-25. PMID: 19880289.

41. Gluud LL. Bias in clinical intervention research. Am J Epidemiol. 2006;163(6):493-501.

42. Groenwold RH, Van Deursen AM, Hoes AW, et al. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. [Review] [25 refs]. Ann Epidemiol. 2008;18(10):746-51. PMID: 18693038.

43. Hahn S, Puffer S, Torgerson DJ, et al. Methodological bias in cluster randomised trials. [Review] [19 refs]. BMC Medical Research Methodology. 2005;5:10. PMID: 15743523.

44. Hahn S, Williamson PR, Hutton JL, et al. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med. 2000;19(24):3325-36. PMID: 11122498.

45. Hawkins N, Scott DA, Woods BS, et al. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. Value in Health. 2009;12(6):996-1003. PMID: 19402854.

46. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37. PMID: 16549855.

47. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. [Review] [60 refs]. J Clin Epidemiol. 2006;59(12):1249-56. PMID: 17098567.

48. Hewitt CE, Kumaravel B, Dumville JC, et al. Assessing the impact of attrition in randomized controlled trials. J Clin Epidemiol. 2010;63(11):1264-70. PMID: 20573482.

49. Hunter PR. Household water treatment in developing countries: comparing different intervention types using meta-regression. Environmental Science & Technology. 2009;43(23):8991-7. PMID: 19943678.

50. Ioannidis JP. Excess significance bias in the literature on brain volume abnormalities. Arch Gen Psychiatry. 2011;68(8):773-80. PMID: 21464342.

51. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. [Review] [100 refs]. J Clin Epidemiol. 1996;49(2):235-43. PMID: 8606325.

52. Jones R, Younie S, Macallister A, et al. A comparison of the scientific quality of publicly and privately funded randomized controlled drug trials. J Eval Clin Pract. 2010;16(6):1322-5. PMID: 20738476.

53. Jorgensen AW, Maric KL, Tendal B, et al. Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: differences in methodological quality and conclusions. BMC Medical Research Methodology. 2008;8:60. PMID: 18782430.

54. Joseph L, Belisle P, Tamim H, et al. Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. J Clin Epidemiol. 2004;57(2):147-53.

55. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. [Review] [37 refs]. BMJ. 2001;323(7303):42-6. PMID: 11440947.

56. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. Int J Epidemiol. 2002;31(1):115-23.

57. Katerndahl DA, Lawler WR. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: a meta-meta-analysis. Am J Epidemiol. 1999;149(5):429-41. PMID: 10067902.

58. Keogh-Brown MR, Bachmann MO, Shepstone L, et al. Contamination in trials of educational interventions. [Review] [276 refs]. Health Technology Assessment (Winchester, England). 2007;11(43):iii-107. PMID: 17935683.

59. Khan KS, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. Arch Intern Med. 1996;156(6):661-6. PMID: 8629879.

60. Kilwein JH. Biases in medical literature. J Clin Pharm Ther. 1999;24(6):393-6. PMID: 10651971.

61. Kondo N, Bessho H, Honda S, et al. Complement factor H Y402H variant and risk of age-related macular degeneration in Asians: a systematic review and meta-analysis. [Review]. Ophthalmology. 2011;118(2):339-44. PMID: 20869121.

62. Kunz R, Neumayer HH, Khan KS. When small degrees of bias in randomized trials can mislead clinical decisions: an example of individualizing preventive treatment of upper gastrointestinal bleeding. Crit Care Med. 2002;30(7):1503-7. PMID: 12130970.

63. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ. 1998;317(7167):1185-90. PMID: 9794851.

64. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. [Review] [113 refs][Update in Cochrane Database Syst Rev. 2011;4:MR000012; PMID: 21491415]. Cochrane Database of Systematic Reviews. 2007(2):MR000012. PMID: 17443633.

65. Lee PN. Difficulties in assessing the relationship between passive smoking and lung cancer. [Review] [99 refs]. Stat Methods Med Res. 1998;7(2):137-63. PMID: 9654639.

66. Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. [Review] [26 refs]. BMJ. 2003;326(7400):1167-70. PMID: 12775614.

67. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests.[Erratum appears in JAMA 2000 Apr 19;283(15):1963]. JAMA. 1999;282(11):1061-6. PMID: 10493205.

68. Lin IF, Myunghee CP. Matched case-control data analysis with selection bias. Biometrics. 2001;57(4):1106-12.

69. Lo BW, Kyu HH, Jichici D, et al. Meta-analysis of randomized trials on first line and adjunctive levetiracetam. [Review]. Can J Neurol Sci. 2011;38(3):475-86. PMID: 21515509.

70. MacLehose RR, Reeves BC, Harvey IM, et al. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. [Review] [192 refs]. Health Technology Assessment (Winchester, England). 2000;4(34):1-154. PMID: 11134917.

71. Marshall RJ. An empirical investigation of exposure measurement bias and its components in case-control studies. J Clin Epidemiol. 1999;52(6):547-50. PMID: 10408994.

72. Marsoni S, Torri W, Taiana A, et al. Critical review of the quality and development of randomized clinical trials (RCTs) and their influence on the treatment of advanced epithelial ovarian cancer. Ann Oncol. 1990;1(5):343-50. PMID: 2148106.

73. Matt GE, Navarro AM. What meta-analyses have and have not taught us about psychotherapy effects: a review and future directions. [Review] [114 refs]. Clin Psychol Rev. 1997;17(1):1-32. PMID: 9125365.

74. Mhaskar R, Djulbegovic B, Magazin A, et al. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012 Jun;65(6):602-9. PMID: 22424985.

75. Michaud S, Suzuki S, Harbarth S. Effect of design-related bias in studies of diagnostic tests for ventilator-associated pneumonia. [Review] [50 refs]. Am J Respir Crit Care Med. 2002;166(10):1320-5. PMID: 12421741.

76. Miller F, Friede T, Kieser M. Blinded assessment of treatment effects utilizing information about the randomization block length. Stat Med. 2009;28(12):1690-706. PMID: 19340815.

77. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000;53(9):964-72. PMID: 11004423.

78. Moher D, Pham B, Lawson ML, et al. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. [Review] [72 refs]. Health Technology Assessment (Winchester, England). 2003;7(41):1-90. PMID: 14670218.

79. Morissette K, Tricco AC, Horsley T, et al. Blinded versus unblinded assessments of risk of bias in studies included in a systematic review. Cochrane Database of Systematic Reviews: Reviews 2011 Issue 9 John.Wiley & Sons. 2011.

80. Morris RK, Selman TJ, Zamora J, et al. Methodological quality of test accuracy studies included in systematic reviews in obstetrics and gynaecology: sources of bias. BMC Women's Health. 2011;11:7. PMID: 21426545.

81. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ. 1997 Sep 13;315(7109):617-9. PMID: 9310553.

82. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. [Review][Update of Cochrane Database Syst Rev. 2007;(2):MR000012; PMID: 17443633]. Cochrane Database of Systematic Reviews. 2011;4:MR000012. PMID: 21491415.

83. Oliveri RS, Wetterslev J, Gluud C. Transarterial (chemo) embolisation for unresectable hepatocellular carcinoma. [Review]. Cochrane Database of Systematic Reviews. 2011;3:CD004787. PMID: 21412886.

84. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. [Review]. J Clin Epidemiol. 2011;64(10):1060-9. PMID: 21454050.

85. Perry M, Faes M, Reelick MF, et al. Studywise minimization: a treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. J Clin Epidemiol. 2010;63(10):1118-22. PMID: 20304606.

86. Peters J, Mengersen K. Selective reporting of adjusted estimates in observational epidemiology studies: reasons and implications for meta-analyses. [Review] [54 refs]. Eval Health Prof. 2008;31(4):370-89. PMID: 19000980.

87. Petscavage JM, Richardson ML, Carr RB. Verification bias an underrecognized source of error in assessing the efficacy of medical imaging. Acad Radiol. 2011;18(3):343-6. PMID: 21145764.

88. Pigott HE, Leventhal AM, Alter GS, et al. Efficacy and effectiveness of antidepressants: current status of research. [Review]. Psychother Psychosom. 2010;79(5):267-79. PMID: 20616621.

89. Polit DF, Gillespie BM, Griffin R. Deliberate ignorance: a systematic review of blinding in nursing clinical trials. [Review]. Nurs Res. 2011;60(1):9-16. PMID: 21127453.

90. Potter S, Brigic A, Whiting PF, et al. Reporting clinical outcomes of breast reconstruction: a systematic review. [Review]. J Natl Cancer Inst. 2011;103(1):31-46. PMID: 21131574.

91. Rief W, Nestoriuc Y, von Lilienfeld-Toal A, et al. Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: a systematic review and meta-analysis. [Review] [185 refs]. Drug Saf. 2009;32(11):1041-56. PMID: 19810776.

92. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med. 1071;8(9):1051-69. PMID: 2799131.

93. Sampson M, Barrowman NJ, Moher D, et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol. 2003;56(10):943-55. PMID: 14568625.

94. Sanchez-Ramos L, Delke I, Zamora J, et al. Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients: a meta-analysis. [Review] [54 refs]. Obstet Gynecol. 2009;114(3):631-40. PMID: 19701045.

95. Savard LA, Thompson DR, Clark AM. A meta-review of evidence on heart failure disease management programs: the challenges of describing and synthesizing evidence on complex interventions. [Review]. Trials [Electronic Resource]. 2011;12:194. PMID: 21846340.

96. Schmidt LM, Gotzsche PC. Of mites and men: reference bias in narrative review articles: a systematic review. [Review] [20 refs]. J Fam Pract. 2005;54(4):334-8. PMID: 15833223.

97. Schulz KF, Chalmers I, Grimes DA, et al. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA. 1994;272(2):125-8. PMID: 8015122.

98. Schwarzer G, Antes G, Schumacher M. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. Stat Med. 2002;21(17):2465-77. PMID: 12205693.

99. Schwarzer G, Carpenter J, Rucker G. Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. J Clin Epidemiol. 2010;63(3):282-8. PMID: 19836925.

100. Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. [Review] [32 refs]. Am J Epidemiol. 2000;151(10):939-45. PMID: 10853631.

101. Shimada YJ, Shiota T. A meta-analysis and investigation for the source of bias of left ventricular volumes and function by three-dimensional echocardiography in comparison with magnetic resonance imaging. Am J Cardiol. 2011;107(1):126-38. PMID: 21146700.

102. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. [Review] [537 refs]. Health Technology Assessment (Winchester, England). 2001;14(8):iii-ixi. PMID: 20181324.

103. Soonawala D, Middelburg RA, Egger M, et al. Efficacy of experimental treatments compared with standard treatments in non-inferiority trials: a meta-analysis of randomized controlled trials. Int J Epidemiol. 2010;39(6):1567-81. PMID: 20837637.

104. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. [Review] [46 refs]. BMJ. 2001;323(7304):101-5. PMID: 11451790.

105. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000;53(11):1119-29. PMID: 11106885.

106. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002. PMID: 21784880.

107. Thompson S, Ekelund U, Jebb S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. Int J Epidemiol. 2011;40(3):765-77.

108. Tricco AC, Tetzlaff J, Pham B, et al. Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. J Clin Epidemiol. 2009;62(4):380-6. PMID: 19128940.

109. Tricco AC, Tetzlaff J, Sampson M, et al. Few systematic reviews exist documenting the extent of bias: a systematic review. [Review] [81 refs]. J Clin Epidemiol. 2008;61(5):422-34. PMID: 18394534.

110. van TE, van de Pol RJ, Oostendorp RA, et al. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: a systematic review. [Review]. Journal of Physiotherapy. 2010;56(4):223-35. PMID: 21091412.

111. Villar J, Mackey ME, Carroli G, et al. Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: comparison of fixed and random effects models. Stat Med. 2001;20(23):3635-47. PMID: 11746343.

112. Whiting P, Rutjes AW, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. [Review] [61 refs]. Ann Intern Med. 2004;140(3):189-202. PMID: 14757617.

113. Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. Trials. 2007;8 , 2007. Article Number: 9. Date of Publication: 06 Mar 2007.

114. Wilson DB, Lipsey MW. The role of method in treatment effectiveness research: evidence from meta-analysis. Psychological Methods. 2001;6(4):413-29. PMID: 11778681.

115. Yengopal V, Mickenautsch S. Caries-preventive effect of resin-modified glass-ionomer cement (RM-GIC) versus composite resin: a quantitative systematic review. [Review]. European Archives of Paediatric Dentistry: Official Journal of the European Academy of Paediatric Dentistry. 2011;12(1):5-14. PMID: 21299939.

116. Zeng Y, Duan X, Xu J, et al. TPO receptor agonist for chronic idiopathic thrombocytopenic purpura. [Review]. Cochrane Database of Systematic Reviews. 2011(7):CD008235. PMID: 21735426.

117. Turner EH, Knoepflmacher D, Shapley L. Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration database. PLoS Med. 2012;9(3):e1001189. PMID: 22448149.

Appendix D. Detailed Study Characteristics

Table D-1. Characteristics of all included studies

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Schulz et al., 1995[1]	Goal: To determine if selected biases are associated with estimates of treatment effect. Control for bias: Allocation concealment, Generation of allocation schedule (randomization), Exclusions after randomization, Double-blinding	Conditions: care during pregnancy, preterm labor and delivery, induction of labor, labor and delivery, prophylactic antibiotics for cesarean delivery, puerperium, early neonatal period Outcomes: Binary, specific outcomes: NR	Include: RCTs of pregnancy and childbirth; MAs of 5+ RCTs with at least 25 outcome events among control group and at least one trial with adequate concealment and one without; duplicative trials dropped by including most homogeneous MA. Exclude: Unpublished and non-English language studies	MA level 33 MA (250 trials) Search: Cochrane database of SRs published by the Pregnancy and Childbirth group (1955-1992); Based on Oxford database of perinatal trials; dates of trials dk
Juni et al.,1999[2]	Goal: To determine if the type of quality assessment scale affects the conclusions of MA. Control for bias: Concealment of randomization Blinding of outcome assessor Handling of drop-outs and withdrawals (ITT performed)	Condition: Prevention of postoperative thrombosis comparing low molecular-weight heparin to regular heparin in general surgery trials Outcomes: thromboembolitic events (bleeding and deep vein thrombosis)	Inclusion/exclusion: NR in this article; heterogeneity in surgical procedures not described	RCT level Data from 17 RCTs included in 1 MA Search: MA by Nurmohamed et al., 1992[3]

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Linde et al., 1999[4]	Goal: To compare three different approaches to investigating the impact of quality aspects on outcomes in a published MA of placebo-controlled trials: (1) the influence of single-quality components on the outcome, (2) using cut-off points in quality scores as inclusion criterion, and (3) entering trials into MAs consecutively according to quality scores (cumulative MA). Control for bias: Randomization (explicitly stated) Double blinding (patients and evaluators, txs indistinguishable) Full description of handling of drop-outs and withdrawals	Condition: homeopathic intervention for tx or prevention Outcomes: NR	From 186 trials evaluated, 119 selected for original MA (Linde et al., 1997)[5]. Of these, 89 selected for sensitivity analysis of quality Include: controlled trials on tx or prevention; parallel control group receiving placebo; explicit statement of random assignment to tx and placebo groups, or that the trial involved double-blind conditions for participants, therapists, and outcome evaluators, making unbiased tx allocation likely; presented in a written report, published or unpublished; abstract, full report, or book section; sufficient information after data extraction to have outcome rates calculated for both groups. Exclude: homoeopathic "provings" in which remedies are given to healthy volunteers to assess effects; studies of healthy participants not aimed at tx or prevention; single-case reports; a reasonable outcome measure for data synthesis could not be determined.	Trial level Data from 89 trials included within a single MA on homeopathic interventions compared with placebo Search: Medline, Embase and CAM registries for trials evaluating homeopathy, contacts with researchers, bibliographies of identified articles.

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Moher et al., 1999,[6] Moher et al., 1998[7]	Goal: To determine the effect that the quality of RCTs included in a MA has on estimates of intervention effectiveness Control for bias: • Randomization sequence • Allocation concealment • Double blinding • Adequate follow-up	Conditions: 3 MAs each from the areas of digestive diseases, circulatory diseases, and mental health; 3 MA randomly chosen on stroke, 2 on pregnancy and childbirth. Outcomes: varied across studies	Random selection of 12 MAs (1 excluded post hoc) from larger database of 491 MAs of RCTs. Inclusion: published in English; no formal incorporation of quality scores in quantitative analysis; binary outcomes reported using an overall quantitative summary result. Exclusion: MAs that did not provide references for included trials.	MA level 11 MA: data from 127 trials Search: Cochrane Database of Systematic Reviews
Kjaergard, Villumsen and Gluud, 2001[8]	Goal: To explore whether methodologic quality affects estimated intervention effects in RCTs and contributes to differences between large and small RCTs in MA. Control for bias: • Allocation sequence • Allocation concealment • Double blinding • Adequate follow-up	Conditions: NR Outcomes: mortality, neonatal mortality, cesarean section, deep vein thrombosis, dropouts, endocervical cells, resumed smoking,	Inclusion (SRs): MAs that included at least one large trial (≥ 1000 subjects) Exclusion (SRs): MA with RCTs that were also included in larger eligible MA, lacking references to the primary trials, or low-quality Exclusion (RCTs): unpublished, quasi-randomized, published as abstracts, language (not English or German)	MA level 14 MA: data from 190 RCTs (23 large, 167 small) Search: Medline and Cochrane library
Balk et al., 2002[9]	Goal: To determine if quality measures are associated with tx effect sizes in RCTs 24 controls for biases, including: • allocation concealment, • randomization, • attrition & loss to follow-up, • blinding (double & component) • valid statistical methods • confounding	Conditions: Cardiovascular diseases, infectious diseases, pediatrics, surgery Outcomes: mortality in cardiovascular disease studies; varied in other clinical areas. If multiple outcomes, included those examined by the largest number of studies or those most clearly defined.	Inclusion: MA with at least 6 RCTs, dichotomous outcomes; sig between-study heterogeneity within MA. Exclusion: abstracts, letters, unavailable articles, detailed outcome data not available.	MA level 26 MAs, data from 276 trials (included "85%" of the 325 trials from the MAs) Cardiovascular: 8 MA (93 trials) Infectious disease: 6 MA (56 trials) Pediatrics: 5 MA (60 trials) Surgery: 7 MA (67 trials) Search: Medline and Cochrane databases

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Clifford et al., 2002[10]	Goal: To examine the relationships between funding source, trial outcome and reporting quality--particularly allocation concealment. Controls for biases: • allocation concealment	Conditions: pharmaceutical intervention Outcomes: primary outcome as defined by authors or if not defined, most clinically relevant	Inclusion: RCT published as full report; pharmaceutical intervention.	RCT level 100 RCTs (convenience sample) Search: Hand search of 1/99-10/00 issues of 5 high-impact medical journals (Annals; BMJ; JAMA, Lancet; NEJM) until 20 articles found in a journal.
Sterne et al., 2002[11]	Goal: To compare methods for assessing the influence of trial characteristics on estimated tx effects in data sets containing collections of MAs: (1) fixed effects logistic regression, (2) a meta-analytic approach that combines separate logistic regressions, and (3) meta-regression approach Controls for biases: • Allocation concealment, • Double-blinding	Conditions and Outcomes: See Schulz et al., 1995[1] description	Re-analysis of Schulz et al. (1995)[1] database	MA level 33 MA, data from 250 trials Search: As per Schulz et al. (1995).[1]
Als-Nielsen et al., 2003[12]	Goal: To examine whether an association between funding and conclusions in randomized drug trials. Also, to explore the impact of methodological quality, type of control intervention, trial size, year of publication, or publication in high-impact journals on this association. Bias examined: • Double blinding	Conditions: intensive care, smoking cessation, respiratory disease, ob/gyn, gastroenterology, neurology, psychiatry, infectious disease, rheumatology, nephrology, dermatology	Include: all RCTs in eligible meta-analyses from a random sample of Cochrane reviews obtained in May 2001	RCT level 370 RCTs in 25 MA Source: Cochrane reviews obtained in May 2001

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Egger et al.,2003[13]	Goal: To compare within MAs the characteristics of trials that are difficult to locate (unpublished, published in languages other than English, published in journals not indexed in MEDLINE database), and of lower quality and to assess the impact of excluding trials from pooled effect estimates, based on these characteristics. Controls for biases: • Allocation concealment • Double blinding	Conditions: therapeutic or preventive interventions. Outcomes: binary, specifics: NR	Exclude: MAs that did not have quality information or showed no differences in quality between included RCTs; unpublished, and non-English RCTs.	MA level Allocation concealment: 39 MA, (304 trials) Blinding: 22 MA (399 trials) Search: Issue 1 of Cochrane Database of SRs (1998), SRs included in Database of Abstracts of Reviews of Effectiveness (1994-1998), handsearch: *Health Technology Assessment* and 8 medical journals that regularly publish SRs (1994-1998).
Chan et al., 2004[14]	Goal: To study empirically the extent and nature of outcome reporting bias in a cohort of RCTs. The odds of having statistically sig results if the results were fully or partially reported compared with results that were qualitatively reported or unreported. Bias examined: • Selective outcome reporting	Condition: NR Outcomes: N=2175 for efficacy and N=605 for harms	Include: Completed RCTs with at least 1 published result; outcomes compared between protocol and publication and statistical sig for missing outcome sought from trialists by survey Exclude: if entire rows or columns for trial were empty in a 2x2 table of statistical sig (N of outcomes with p<0.5 vs. p>=0.05) with reporting (fully or partially/qualitatively or unreported)	RCT level (based on comparisons of individual trials and protocols) 35 trials for efficacy 15 trials for harms Sources: PubMed, EMBASE and Cochrane Controlled Trials Register using investigator names and keywords (final search January 2003) protocols for RCTs that were approved for funding (1990-1998) by the Canadian Institutes of Health or the Medical Research Council of Canada

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Chan et al., 2004[15]	Goal: To study empirically the extent and nature of outcome reporting bias in a cohort of RCTs Bias examined: • Selective outcome reporting	Condition: NR Outcome: N=1039 efficacy and N=145 for harms.	Include: Completed RCTs with available protocol and at least 1 published result Exclude: if entire rows or columns for trial were empty in a 2x2 table of statistical sig (N of outcomes with p<0.5 vs. p>=0.05) with reporting (fully or partially/qualitatively or unreported)	RCT level (comparisons of published studies and protocols) 30 trials for efficacy, 4 trials for harms Sources: PubMed, EMBASE and Cochrane Trials Register using investigator names and keywords (final search in January 2003) for protocols
Kyzas et al., 2005[16]	Goal: To assemble empirical evidence on the importance of selective reporting biases for prognostic evidence in malignant diseases Biases examined: • Outcome reporting • Blinding	Condition: Association between the tumor suppressor protein TP53 and head and neck squamous cell cancer (HNSCC), Outcome: all cause mortality and lymph node status	Inclusion: All English language MAs that examined potential prognostic factors for any malignancy and their association with mortality.	Trial level 42 trials Sources: PubMed and EMBASE
Tierney et al., 2004[17]	Goal: To investigate how excluding patients from RCTs can affect the results of trials and MAs Bias examined: • Attrition (ITT vs. not)	Condition: Cancer (bladder, brain, lung, esophagus, ovary, lung, and soft tissue sarcoma) Outcome: survival	Inclusion/Exclusion: No other details reported	MA and trial level 14 MA 92 RCTs with at least one patient exclusion (21905 patients) Sources: SRs and MAs of patient-level data from RCTs that addressed cancer therapies

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Derry et al., 2006[18]	Goal: To review the efficacy of SRs to accurately assess the evidence for acupuncture. Biases examined: • Randomization • Blinding	Condition: those treated with acupuncture including various painful conditions (18 SRs), stroke (2 SRs), nausea and vomiting (2SRs), depression (2 SRs) and other including insomnia, smoking cessation, weight loss, and asthma (11 SRs) Outcomes: those relevant to topic area (e.g. patient pain scoring, number of headache-free days, long-term outcomes for chronic conditions)	Inclusion: English, examining the efficacy of traditional Chinese or mechanical acupuncture, electro-acupuncture, laser acupuncture or acupressure, electrical nerve stimulation. Exclusion: Transcutaneous or dry needling, reviews of adverse event from acupuncture. Where one SR clearly updated a previous review, only the most recent was used. If more than one SR covered same trials for the same outcome and indication, the most recent was included.	SR level 35 SRs Sources: PubMed, AMED, Cochrane library of SRs of acupuncture for any conditions in humans, published 1/1996-8/2005 using terms 'acupuncture' and 'systematic review OR meta-analysis'.
Furukawa et al., 2007 [19]	Goal: To evaluate the extent Cochrane MAs include only a proportion of identified RCTs when estimating tx effect and whether the proportion of RCTs included in a MA is associated with its pooled effect size	Conditions: NS Outcomes: Primary but NS	Inclusion: SRs with 10 or more RCTs from a set of 500 SRs selected by random-number generator from Issue 4 of the Cochrane Library 2005.	RCT level 156 trials Source: Cochrane Library, Issue 4 (2005)

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Pildal et al., 2007[20]	Goal: To estimate the fraction of conclusions based on statistically significant results in MAs that would no longer be supported if only trials with reported adequate allocation concealment were included, and to assess the impact of absence vs. presence of reported adequate allocation concealment on the effect estimates of trials. Biases examined: • Allocation concealment • Double blinding	Conditions: NS Outcomes: NS	Inclusion: SRs where authors concluded that one of assessed interventions was superior to the other and if this preference was supported by first statistically sig result of a MA reported in the abstract. Exclusion: The first statistically sig result of a MA reported in the abstract was not for a binary outcome; substantial uncertainty concerning what authors of the SR perceived as experimental and conventional tx, more than 40 trials in the first statistically sig MA reported in the abstract, genuine MA not performed; abstract of the SR stated that it was partly based on non-randomized trials; a mix of adequate and inadequate concealment	MA level MA level 34 MA for allocation concealment (283 trials), 20 MA for blinding (182 trials) Sources: Cochrane Library, PubMed
Siersma et al., 2007[21]	Goal: To investigate the properties of multivariable meta-epidemiological analyses based on logistic regression and weighted regression models. Biases examined: • Sequence generation, • allocation concealment, • double blinding, • intention-to-treat	Conditions: NS Outcomes: Primary binary but NS	Inclusion: 167 SRs from the 1081 SRs published in The Cochrane Library selected by a computer-generated list of random numbers	Trial level 523 RCTs included in 48 MA Source: Cochrane Library

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Fenwick et al., 2008[22]	Goal: To examine the impact of allocation concealment and assessor blinding on the size of clinical outcomes in periodontology Biases examined: • Allocation concealment • Assessor blinding	Condition: periodontology Outcomes: probing depth, and clinical or probing attachment level	Inclusion: RCTs published up to 1/2007, SRs with RCTs with outcomes of probing depth or clinical or probing attachment level, English language	RCT level 5 SRs (50 RCTs), 34 RCTs allocation concealment, 33 RCTs blinding Sources: Cochrane library
Wood et al., 2008[23]	Goal: To examine whether the association of inadequate or unclear allocation concealment and lack of blinding with biased estimates of intervention effects varies with the nature of the intervention or outcome (objective vs. subjective measures) Biases examined: • Allocation concealment • Blinding • Combination of the two	Condition: varied Outcomes: Objectively assessed: All cause mortality, Laboratory measurement including surgical/instrumental (caesarean, instrumental delivery, epidural analgesia, manual removal of placenta), Other (birth weight, timing of delivery, hemorrhage or blood loss, non-cephalic birth, continuing lactation one week after birth, deep venous thrombosis, live birth, failed delivery, episiotomy, retention in school grade) Subjectively assessed: patient reported outcomes, physician reported	Data from 3 meta-epi studies: Schulz et al, 1995:[24] 33 MA from Pregnancy and Childbirth Group of Cochrane. Each MA included at least 5 trials with a combined total of at least 25 outcome events Kjaergard et al, 2001:[8] 14 MA from 11 SRs, including at least 1 trial of at least 1000 participants Egger et al., 2003:[25] 122 MA from Cochrane Database of SRs that contained at least 5 randomized trials Removed duplicate MA, retained trials that contributed to >1 MA because had >1 intervention arm or outcome.	RCT level MA included: 146 of 169 eligible (1346 trials of 1615 eligible) Allocation concealment: 102 MA (804 trials); Blinding: 76 MA (746 trials) Sources: Authors of previously published meta-epidemiological studies: Schulz et al (1995):[24], Kjaergard et al (2001):[8], Egger et al (2003):[25]
Hartling et al., 2009 [26]	Goal: To evaluate the reliability and concurrent validity of the Cochrane risk of bias tool compared with the Jadad and Schulz approach to risk of bias assessment. Biases examined: • Allocation concealment • Sequence generation • Randomization • Blinding	Condition: related to pediatric health Outcomes: Primary but NS	Inclusion: Convenience sample of RCTs in child health.	RCT level 163 RCTs Source: Manuscripts resulting from abstracts presented at the annual scientific meetings of the Society for Pediatric Research between 1992-1995

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Inaba et al., 2009[27]	Goal: To evaluate differences in study design between single and multicenter trials on study outcomes. Specifically, the impact of adequate randomization and blinding of the outcome assessors. Biases examined: • Randomization • Blinding of outcome assessor • Disclosure of withdrawals	Condition: heart procedures Outcomes: mortality/ survival, impaired myocardial blush grade, incomplete ST resolution	Inclusion: RCTs, all languages, had to examine the use of adjunctive mechanical devices compared with percutaneous coronary intervention in patients with acute myocardial infarction. Exclusion: if adjunctive mechanical devices were used in saphenous vein grafts, duplicate data from other studies, or insufficient data for MA.	Study-level without accounting for MA. 25 trials, two subgroup analyses (or MA) but the number of trials into these groups was not specified. The subgroups were based on different outcomes Sources: Searched 9 bibliographic databases, slide and oral presentations.
Nuesch et al., 2009[28] Nuesch et al., 2009[29]	Goal: To evaluate the association of adequate allocation concealment and patient blinding with estimates of tx benefits in osteoarthritis trials (whether excluding patients from the analysis was associated with biased estimates of tx effects and with increased heterogeneity between trials in MAs) Biases examined: • Allocation concealment • Patient blinding • Attrition	Condition: Pain in osteoarthritis of the knee or hip Outcomes: non-binary subjective outcomes from interventions (Patient reported pain)	Inclusion: MAs of randomized or quasi-randomized trials. Trials using an unpredictable allocation sequence considered randomized, potentially predictable allocation mechanisms, such as alternation or the allocation of patients according to their date of birth considered quasi-randomized. MAs eligible if assessed patient reported pain, comparing any intervention with placebo, sham, or a non-intervention control. No language restrictions applied.	MA Concealment: 14 MA (158 trials and 40,437 patients) Blinding: 10 MA (122 trials and 27,452 patients) Sources: Cochrane, PubMed, Embase and CINAHL (Within specified databases, combinations of keywords and text words related to osteoarthritis were combined with validated filters for SRs and MAs (Last update: 11/20/2007).

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Van Tulder et al. 2009 [30]	Goal: To assess the validity of the criteria list recommended for evaluating internal validity by the Cochrane Back Review Group Editorial Board by evaluating whether individual items and a total score are associated with effect sizes in RCTs of back pain interventions. Biases evaluated: • method of randomization • allocation concealment • groups similar at baseline • patient, care provider, assessor blinding • effect of co-intervention • differential compliance • drop-out rate • timing of outcome assessment across groups • ITT	Condition: Low-back pain Outcomes: Pain, function, or similar improvement measure	Inclusion: RCT, comparison between tx and either placebo, usual care, no tx or another tx. All Cochrane Back Review Group (CBRG) reviews of nonsurgical tx for nonspecific low back pain in the Cochrane library 2005, issue 3. Exclusion: Data presented in such a way that effect size could not be calculated	RCT N = 216 Comparison to placebo/tx/usual care: N = 122 Comparison to active intervention: N = 128 Some studies included both types of comparisons Sources: Cochrane Library 2005, issue 3: all Cochrane Back Review Group reviews of a nonsurgical treatment for nonspecific low back pain.
Dwan et al., 2010[31]	Goal: To assess a SR for outcome reporting bias Bias evaluated: • Outcome reporting	Condition: Acute asthma Outcomes: Pulmonary function tests, including peak expiratory flow rate (PEFR) and forced expiratory volume in one second (FEV1), and hospital admission	Inclusion: Trials that were eligible for inclusion in the SR because they reported a measure of pulmonary function or hospital admission	RCT-level 24 trials Sources: All studies included in the a SR concerning intravenous and nebulized magnesium sulfate for acute asthma'

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Number Included Sources Searched for Studies or MA
Hamm et al., 2010 [32]	Goal: To review a sample of RCTs in the Cochrane pediatric registry and assess the validity of the results. Biases evaluated: • sequence generation, • allocation concealment, • blinding, • incomplete outcomes, • selective outcome reporting	Conditions: Unspecified Outcomes: Unspecified (primary RCT outcome)	Inclusion: Random selection of pediatric trials published in Cochrane Registry in 2007	RCT-level 300 trials (236 analyzed) Source: See inclusion criteria
Kirkham et al., 2010[33]	Goal: To examine the prevalence of outcome reporting bias and its impact on Cochrane reviews Bias evaluated: • Outcome reporting	Condition: not specified; Outcomes: Primary outcomes of SRs. If SR did not specify a single primary outcome, authors were contacted to select one. When no contact could be established, two investigators independently selected and agreed upon one from the outcomes listed in the SR.	Excluded: conducted by Cochrane methodology group; ill-defined primary outcome; no RCTs identified; fully reported review with primary outcomes from all included trials; multiple MAs of primary outcome; non-English; no MA; primary outcome measured in different ways; longitudinal study; studies not combined due to clinical heterogeneity Inclusion: single MA of primary outcome; primary outcome of interest not fully reported in MA or tabulated form in at least one trial Trial exclusion: non-English; primary outcome fully reported; primary outcome clearly not reported; unclear whether primary outcome was measured, but no suspicion of reporting bias or likely to have not been reported because of zero events	SR level 81 SRs for assessment of impact (RCTs: NR) 25 SRs for sensitivity analysis (RCTs: NR) Sources: SRs in 3 issues of the Cochrane Library representing 50 of the 51 Cochrane Collaboration review groups: Issue 4, (2006); Issues 1 & 2 (2007).

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Included Sources Searched for Studies or MA
Hartling et al., 2011[34]	Goal: To evaluate the relationship between risk of bias and effect estimates in RCTs and to examine differences when unclear RoB is grouped with low RoB and when it is grouped with high RoB Biases examined: • Sequence generation • Allocation concealment • Blinding • Incomplete data • Selective reporting	Conditions: Adults with persistent asthma Outcomes: Forced expiratory volume in 1 second (FEV_1)	Inclusion: RCTs included in a SR of combination long-acting beta-agonists and inhaled corticosteroids for maintenance therapy in persistent asthma.	RCT level 107 RCTs Sources: electronic databases and grey literature

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Included Sources Searched for Studies or MA
Hempel et al., 2012 [35], 2011 [36]	Goal: to examine the empirical evidence for associations between a set of proposed quality criteria and estimates of effect sizes in RCTs using multiple datasets representing a variety of clinical fields and to explore variables potentially influencing the association (effect moderators or confounders); specifically, whether (1) the overall size of the observed treatment effect, (2) the condition being treated, (3) the type of outcome, and (4) the variance in effect sizes across studies moderates or confounds the association between quality and effect sizes. Biases examined • Randomization (sequence generation and allocation concealment) • Assessor blinding • Patient blinding • Care provider blinding • Similar co-interventions/ compliance/timing • Drop-out rate	Conditions: Various including back pain, complementary and alternative medicine, mental health (Alzheimer's, obsessive compulsive-disorder), diabetes, digestive diseases, pregnancy and childbirth, and infectious diseases. Outcomes: Mostly continuous but some categorical such as death, pregnancy	Include: RCTs from four different data sets that contained MA. Exclude: SRs that did not contain MA	RCT level 481 studies
Herbison 2011[37] Herbison 2006[38]	Goal: To determine how much bias is associated with different methods of tx allocation concealment (grouped into 6 categories). Bias examined: • Allocation concealment	Condition: Not limited by condition Outcome: Any binary outcome	Inclusion: SRs with binary outcomes, at least 10 included trials, and at least one trial with more than 500 people randomized to each arm.	Meta-analysis 65 MAs from 18 SRs; 389 studies Source: Issue 1 (2001) of Cochrane Library

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Included Sources Searched for Studies or MA
Liu, LaValley & Latham, 2011[39]	Goal: To determine the differential effects of progressive resistance strength training on lower limb muscle strength in older adults between RCTs that used blinded outcome assessors and those that did not. As a further step, to determine the influence of ITT analysis while estimating the effect of blinding Bias examined: • Blinding	Condition: Not limited by condition; Outcome: Continuous outcome measuring lower limb muscle strength	Include: Must measure lower limb muscle strength, participants ≥60 years of age, progressive resistance training was main intervention	RCT level 73 studies
Hartling et al., 2012 [3]	Goal: To assess the reliability of the Cochrane ROB tool for RCTs and the Newcastle-Ottawa Scale (NOS) for cohort studies between individual raters, and between consensus agreements of individual raters for the ROB tool; assess the validity of the Cochrane ROB tool and NOS by examining the association between study quality and treatment effect size; examine the impact of study-level factors (e.g., outcomes, interventions and conditions) on scale reliability and validity. Bias examined: • randomization • allocation concealment • double blinding • Selective outcome reporting • Attrition	Conditions: Varied across studies. The most frequently represented categories were circulatory and respiratory health (18 percent), nutrition, metabolism, and diabetes (17 percent), and musculoskeletal health and arthritis (15 percent). The primary outcomes were objective in 48 percent of trials and subjective in 52 percent. Source of outcome assessment was primarily by clinician (35 percent), laboratory measure (23 percent), or self-report (23 percent). Outcomes: 1) Objective include all cause mortality, measures based on a recognized laboratory procedure, surgical or instrumental outcomes and other objective measures. 2) Subjective include patient reported, physician assessed disease outcomes, measures combined from several outcomes, and withdrawals or study dropouts.	Inclusion: RCTs elected from previous study by Hopewell 2010 and selected random sample (154/ 616) representing approximately 25% of the original trials.	RCT level 154 studies Source: previous study by Hopewell 2010

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Included Sources Searched for Studies or MA
Hrobjartsson et al., 2012 [40]	Goal: To review RCTs with blinded and non-blinded assessors of binary outcomes to evaluate the impact of non-blinded outcome assessment on estimated treatment effects and to examine reasons for the variation. Bias examined: • Assessor blinding	Condition: Not searched in relation to specific conditions but included general surgery, orthopedic surgery, plastic surgery, cardiology, gynecology, anesthesiology, neurology, psychiatry, dermatology, otolaryngology, infectious diseases, and ophthalmology. Outcome: One, primary, binary outcome from each RCT.	Inclusion: RCT, blinded and non-blinded assessment of the same binary outcome, if more than one outcome met criteria, the primary outcome was chosen. Exclude: 1) Unclear which group was experimental and which control, 2) only a subgroup of patients evaluated by blinded and non-blinded assessors, unless they were selected at random, 3) blinded and non-blinded assessors had access to each other's results (e.g., blinded assessments were provided to non-blinded assessors as a quality enhancement procedure), 4) initially blinded assessors clearly had become unblinded, 5) blinded end point committees adjudicating the assessments made by non-blinded clinicians because such adjudication often involved previous knowledge of the non-blinded assessment or is restricted to adjudication of events only.	RCT level 25 RCTs included and of these, 21 RCTs analyzed Searched: PubMed, Embase, PsycINFO, CINAHL, Cochrane Register of Controlled Trials, full text databases (High Wire Press and Google Scholar). Last search: January 26, 2010

Table D-1. Characteristics of all included studies (continued)

Study Identification	Goal of Study/Approach to Controlling for Biases in Study Conduct	Conditions Evaluated/ Outcomes Reported	Inclusion/Exclusion Criteria	Unit of Analysis and Included Sources Searched for Studies or MA
Mhaskar et al., 2012 [41]	Goal: To assess whether the reported methodological quality of RCTs reflects the actual methodological quality and to evaluate the association of effect size and sample size with methodological quality Biases examined: • Selective outcome reporting: • generation of randomization sequence • allocation concealment • intention-to-treat and description of dropouts • description of blinding procedures and appropriate statistical methods	Conditions: Cancer trials Outcomes: Survival	Inclusion: All National Cancer Institute sponsored Cooperative Group trials between 1968 and 2006 with protocols and publications available with unique RCTs.	RCT level 429 RCTs Search: eight National Cancer Institute (NCI) sponsored Cooperative Groups (COG)
Savovic et al., 2012 [42]; Savovic et al., [43]	Goal: To combine trials and MAs across databases from previous meta-epidemiological studies and assess the impact of 3 different types of biases. Bias examined: • sequence generation, • allocation concealment, • double blinding	Conditions: 8 clinical areas based on the ICD-10-CM coding: Pregnancy and Childbirth, Mental and Behavioral Health, Circulatory System, Digestive System, Other factors (factors influencing health status and contact with health services), Respiratory System, Other ICD-10, Unclassified Outcomes: All cause mortality; Other objective; Objective but influenced by clinician judgment; Subjective; Mixture of objective and subjective	Data from 10 earlier meta-epidemiological studies combined into one database: 1) Als-Nielson et al., 2004,[44] Siersma et al., 2007[45] 2) Balk et al., 2002[9] 3) Contopoulos-Ionnidis et al., 2005[46] 4) Egger et al., 2003[13] 5) Kjaergard, Villumsen, and Gluud; 2001[8] 6) McAuley et al., 2000 7) Pildal et al., 2007[20] 8) Royle, 2003 9) Sampson et al., 2003 10) Schulz et al., 1995[1]	MA level 234 MA, including 1793 trials (not all MA or RCTs used for all analyses) Sources: see inclusion criteria

Abbreviations: BMJ = British Medical Journal; CAM = complementary and alternative medicine; ITT = intent to treat; MA = meta-analysis; N = number; NEJM=New England Journal of Medicine; NR = not reported; ob/gyn: obstetrics and gynecology; RCT = randomized controlled trial; ROB = risk of bias; SR = systematic review; tx = treatment

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Schulz 1995[1]	Pregnancy related: induction of labor, labor and delivery, prophylactic antibiotics for cesarean delivery, puererium, early neonatal condition Comparators: NS	Pregnancy and child-birth	Binary but otherwise NS	No, but attempted to assemble homogenous groups of interventions.	Controlled for other biases (sequence generation, double-blinding, post-allocation exclusions)
Juni 1999[2]	Low molecular weight heparin Comparator: standard heparin	General Surgery	Thromboemolitic events (bleeding and DVT)	NA: all RCTs included same intervention and outcome	Did not control for any other bias components or other factors.
Linde 1999[4]	Various homeopathic remedies Comparators: placebo	Included but is not limited to: Allergic asthma, pollinosis, warts, minor burns, skin lesions, pyodemia, dermatoses, anal fissure, diarrhea, gastritis, choelcystopathia, IBS, sprains, heamarthrosis, cramps, dental neuralgia, migraine, seasickness, aphasia, stroke, menopause, vaginal discharge, PMS, childbirth, menopausal complications, mastodyia, cystitis, cough, URI, OM, pharyngitis, chronic sinusitis, asthma, RA, osteoarthritis, fibromyalgia	Diverse	No	Controlled for other biases (sequence generation, double-blinding, randomization and losses to follow-up)

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Moher, 1998[7] Moher, 1999[6]	Rectal amniosalicylate, medical tx, sulfazine, low molecular weight heparin, family interventions, anti-depressants, anti-coagulants, bromocropin, cesarean delivery Comparators: NS	Digestive diseases (distalulcerative colitis, reflux esphogitis, colitis) circulatory diseases (transient ischemic attacks, DVT, peripheral arterial disease), mental health (schizophrenia, depression, prenatal smoking cessation), stroke, and childbirth and pregnancy (infertility and caesarian delivery with small baby)	NS	Intervention effect	No
Kjaergard, Villumsen and Gluud, 2001[8]	Thromboembolitic therapy, ACE inhibitors, beta blockers, low molecular weight heparin, pharmacotherapy, Doppler ultrasound, devices for cervical cytology, prostaglandins, risperidone and olanzapine, nicotine replacement, electronic heart rate monitoring Comparators: NS	Heart failure, general surgery, preterm labor, schizophrenia, cervical disorders, hypertension, high risk pregnancy	Primary measure within MA. Included mortality, neonatal mortality, c-section, DVT, endocervical cells, resumed smoking.	Stratified by small vs. large trials (> than 1000 subjects)	Did not control for other bias components or other factors. Controlled for clustering within MAs.
Balk 2002[9]	Interventions: varied Comparators: varied	Cardiovascular diseases, infectious diseases, pediatrics, and surgery.	Mortality in cardiovascular trials, various outcomes for other clinical areas	Stratified by 4 disease areas.	Controlled for clustering in MAs. Did not control for other bias components or other factors.

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Clifford 2002[10]	NS Comparators: NS	NS	Trial primary outcome	Stratified by the direction of the trial outcome: (favored new intervention; favored conventional; neutral; unclear)	Controlled for funding source
Sterne 2002[11]	As per Schultz (1995)[1] Comparators: NS	As per Schultz (1995)[1]	As per Schultz (1995)[1]	No	Controlled for clustering in MAs but no control for other biases.
Egger 2003[13]	Drug interventions (82%) Active control interventions (25%) Comparators: NS	Infectious diseases, neurology, obstetrics and gynecology, other	NS	Intervention: drug vs. nondrug, active vs. non-active control, complementary vs. conventional medicine.	No
Als-Nielsen et al., 2003[12]	Drug interventions	Intensive care, smoking cessation, respiratory disease, ob/gyn, gastroenterology, neurology, psychiatry, infectious disease, rheumatology, nephrology, dermatology	Trial primary outcome	No	Controlled for funding and treatment effect (disease area and type of treatments being compared)
Chan et al., 2004[14]	Drug interventions (74%); surgery or procedure, (11%); counseling or lifestyle, (12%); and equipment, (2%) Comparators: NS	NR	Efficacy and harms outcomes, details: NR	No	No
Chan et al., 2004[15]	Drug interventions (56%); surgery or procedure (21%); counseling or lifestyle (17%); and equipment (6%) Comparators: NS	Cardiology (10 trials), obstetrics and gynecology (8), surgery (7) and pediatrics (6).	Efficacy and harms outcomes, details NR	No	No

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Tierney et al., 2005[17]	Chemotherapy of various kinds Comparators: NS	Cancers of the bladder, brain, lung, esophagus, ovary, lung, and soft tissue sarcoma	Hazard of survival and recurrence	Stratified by each included review	No
Kyzas et al., 2005[16]	Test for TP53 status Comparators: NA	Head and neck squamous cell cancer	Death, lymph node metastases	Location of cancer	No
Derry et al., 2006[18]	Traditional Chinese or mechanical acupuncture, electro-acupuncture, laser acupuncture or acupressure, electrical nerve stimulation or dry needling. Comparator: NS for all, sham acupuncture in some trials	Stroke, various painful conditions, nausea and vomiting, depression other conditions (insomnia, smoking cessation, weight loss, asthma)	NS	Stratified by type of bias (randomization and blinding) and sample size and control event rate.	No
Pildal et al., 2007[20]	Antiretrovirals, assertive community tx, calcium channel blockers, chemotherapy and radiotherapy, systemic corticosteroids, elastic compressive stockings, exercise, stem cell therapy/ autologous bone marrow, low protein diets, non-oxynol 9, postnatal phenobarbital, etc. Comparators: NS	HIV, severe mental disorders, preterm labor, esophageal carcinoma, acute asthma, PVD, coronary heart disease, breast cancer, renal failure, preterm babies	NS	No	No
Siersma et al., 2007[21]	Interventions: NS Comparators: NS	Not limited by disease/population	NS	No	No

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Furukawa et al., 2007[19]	Interventions: NS Comparators: NS	Not limited by disease/population	Outcomes: 2 binary and 2 continuous of greatest patient importance based on the review authors' description; alternatively, Furukawa et al. selected primary outcomes	No	No
Fenwick et al., 2008[22]	Scaling, surgery, guided tissue regeneration, antibiotic, surgery of enamel matrix, tooth brushing, mouthwash, routine care, periodontal surgery, gel, variable frequency maintenance program Comparators: NS	Periodontal disease	Probing depth, clinical attachment level (CAL), probing attachment level (PAL).	Outcome: Probing depth and CAL/PAL	No
Wood et al., 2008[23]	Based on Schultz, 1995,[1] Kjaergard, 2001,[8] Egger, 2003[13] Drug (68%), rehabilitation or psychosocial (5%), prevention and screening (12%), surgery or radiotherapy (3%), communication, organizational, or educational (2%), alternative medicine (1%), other (9%) Comparators: NS	Based on Schultz, 1995[1] Kjaergard, 2001[8] Egger, 2003[13]	Based on Schultz, 1995[1] Kjaergard, 2001[8] Egger, 2003[13] Considered objective vs. subjective outcomes Objectively assessed outcomes: All-cause mortality, laboratory, surgical/instrument, other Subjective Outcomes: patient reported, physician assessed, combined	Intervention: Drug interventions vs. non-drug interventions Outcome: All-cause mortality vs. Other outcomes, Objective versus subjective outcomes	Controlled for other biases (allocation concealment, blinding) Controlled for clustering in one set of analyses but presented logistic regression results described as similar.

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Inaba et al., 2009[27]	Adjunctive devices Comparator: standard percutaneous coronary intervention (PCI)	Acute myocardial infarction	Mortality/ survival, post PCI impaired myocardial blush grade, incomplete ST segment resolution	Stratified by Single vs. multicenter, study regions (n=3), study size, conflict of interest, type of device used	Controlled for other factors (number of centers, age, publication year, % male, % smokers, % glycoprotein Ib/IIz, % anterior wall infarction, % direct stenting, difference in procedure times, pain to balloon time, and door to balloon time)
Hartling et al., 2009[26]	Interventions: NS Comparators: NS Varied across studies	Child health	Primary study outcome	No	Controlled for other factors: Study type (efficacy v equivalence), study design (crossover, factorial, or parallel), outcome type (binary v continuous, objective v subjective). Did not control for other types of bias.
Neusch et al., 2009[28]	Exercise, visco-supplementation, self-management, glucosomine, diacerin, opioids, aquatic exercise, Oral NSAIDS, Topical NSAIDS, pulsed electromagnetic fields, static magnets, weight reduction, acupuncture, chondroitin. Comparators: placebo, sham or non-intervention group	Osteoarthritis	Limited to trials evaluating self-reported pain	Stratified by: Size of tx benefit (small: effect size > -0.5 vs. large: effect size ≤ -0.5) Heterogeneity in overall MA (low: tsq < 0.06 and high: tsq ≥ to 0.06) Intervention: Pharmacologic vs. other and complementary vs. conventional medicine	No

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Van Tulder et al. 2009[30]	Intervention: Any nonsurgical tx Comparators: Placebo, usual care, "no tx" or another tx	Nonspecific low back pain	Pain, function, or similar improvement measure	No	No
Hamm et al., 2010[32]	General surgery, orthopedic surgery, plastic surgery, cardiology, gynecology, anesthesiology, neurology, psychiatry, dermatology, otolaryngology, infectious diseases, and ophthalmology.	Children, not limited by specific disease	Not limited by type of outcome, primary outcome in trial	No	No
Dwan et al., 2010[31]	Intravenous and nebulised magnesium sulphate Comparator: placebo or salbutamol	Acute asthma	Pulmonary function tests and hospital admission	Stratified by intervention and outcome	No
Kirkham et al., 2010[33]	NR Comparators NR	Hepato-biliary (21 reviews), pregnancy and childbirth (18), neonatal (14), oral health (13), menstrual disorders and subfertility (12).	NR	No	No
Hartling et al., 2011[34]	Long-acting beta agonists (LABA) with inhaled corticosteroids Monotherapy	Adults with persistent asthma	Forced expiratory volume in 1 second (FEV_1)	No	No
Liu, LaValley, and Latham, 2011[39]	Progressive muscle resistance Comparators not specified	Older adults	Difference in lower limb muscle strength	No	Analysis controlled for intention-to-treat

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions: Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Hempel et al., 2012[35] Hempel et al., 2011[36]	Range of interventions including dataset 1) non-surgical treatment for back pain: Comparisons: placebo, usual care, or no treatment or comparisons between treatments. Dataset 2) complementary and alternative medicine/dietary supplements; pharmacological therapies, drugs for arthritis, and behavioral interventions (such as self-monitoring of blood glucose (SMBG), diet and weight loss, chronic disease self-management (CDSM); interventions to manage and treat diabetes (chromium, SMBG, CDSM). Dataset 3) variety of interventions and comparisons Dataset 4) cardiovascular disease and pediatrics	Conditions: Various including back pain, complementary and alternative medicine, mental health (Alzheimer's, obsessive compulsive-disorder), diabetes, digestive diseases, pregnancy and childbirth, and infectious diseases.	Outcomes: Mostly continuous but some categorical such as death, pregnancy	No	No

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Herbison et al., 2011[38]	Antifibrinolytic use, antiplatelet agents, antibiotics, caregiver support, calcium antagonists (nimopodine), Doppler ultrasound, haemodilation, hypothermia to prevent neurological damage after coronary artery bypass surgery, intravenous immunoglobulin, maximal androgen blockade, nicotine replacement, oxtocyin, prophylactic corticosteroids, women's position Comparators: NS	Perioperative blood transfusions, preeclampsia, acute sinusitis, childbirth, acute ischemic stroke, fetus in high risk pregnancy, coronary artery disease, sepsis and septic shock, infection in preterm or low weight infants, prostate cancer, smoking cessation, prelabor rupture of membranes, second stage of labor	Number exposed to allogenic blood, pregnancy induced hypertension, proteinuric preeclampsia, C-section, fetal or neonatal death, gestational age, preterm delivery, dropouts related to adverse events, clinical cure or improvement, oxytocin augmentation, analgesia during labor, operative vaginal delivery, cesarean delivery, death or poor outcomes, diff in adverse events, perinatal deaths, stillbirths, neonatal deaths, normally formed, case-fatality, nonfatal strokes, perioperative death, nonfatal MI, low output syndrome, intra-aortic ballon pumpuse, pooled "bad" outcomes, all cause mortality, infection, infant mortality, smoking cessation, overall survival, chorioamniontis, neonatal infection, mode of delivery, episiotomy	No	No
Hrobjartsson et al., 2012[40]	Interventions: surgery, medications, type of dressing, and others Comparators: Similar to primary interventions, usual care, no treatment	Wound/ulcer, fractured bone, angina pectoris, facial folds, other	Varied by study, mostly subjective, such as patient function	Stratified by clinical problem, whether study arms were same type of procedure,	Sensitivity analyses controlling for funding, whether blinding procedure considered effective, whether patients seen by one or two outcome assessors

Table D-2. List of interventions, diseases, outcomes and factors controlled for in the analyses (continued)

Author, Year	Interventions Comparator	Diseases/Populations	Outcomes	Stratified Analyses by Intervention or Disease or Outcome	Statistical Adjustments to Control for Other Biases
Hartling et al., 2012[3]	Interventions: drug interventions (53 percent), behavioral/psychological (11 percent) and surgical (12 percent). Comparators: placebo (36 percent), the rest were similar to primary interventions.	Varied across studies	Primary outcome, varied by study	Stratified by type of outcome (objective or subjective)	No
Mhaskar et al., 2012 [41]	Interventions: NR Comparators: NS	Cancer treatment	NR	No	No
Savovic et al., 2012[42]; Savovic et al., 2012[43]	Pharmacologic; surgical; psychosocial, behavioral, or educational; or other Comparators: Placebo or no tx; other inactive ("standard care"); active comparison; mixture of active and inactive within MA	Pregnancy & childbirth, mental and behavioral, circulatory system, digestive system, other factors, respiratory system, Other ICD-10, unclassified	All-cause mortality, other objective, objectively measured but potentially influenced by clinician judgment, subjective, mixture of subjective and objective	Stratified by outcome grouping (mortality, other objective, subjective or mixed)	Estimation of effect of one bias, controlled for effect of other two biases

Abbreviations: DVT = deep venous thrombosis; IBS = irritable bowel syndrome; MA = meta-analysis; NR=not reported; NA = not applicable; NSAID = nonsteroidal anti-inflammatory drug; OM = otitis media; PMS = premenstrual syndrome; RA = rheumatoid arthritis; tx = treatment; URI = urinary tract infection; vs. = versus

Table D-3 Allocation concealment definitions

	Adequate	Inadequate	Unclear/ Not Reported
Studies reporting Odds Ratios (OR) or Ratio of Odds Ratios (ROR) or Ratio of Relative Risks = (RRR)			
Schulz et al., 1995[1] Schultz 1995[1] Sterne 2002[11]	Measures including central randomization, numbered or coded bottles or containers, drugs prepared by the pharmacy serially numbered opaque sealed envelopes; or other description that contains elements of concealment	INADEQUATE: Concealment was inadequate (such as alternation, or reference to case record numbers or date of birth).	Concealment is unclear because the authors either did not report an allocation concealment approach at all or reported an approach that did not fall into one of the categories above.
Juni 1999, [2]	Scale dependent	Scale dependent	Scale dependent
Linde 1999[4]	Jadad Scale: explicit statement that allocation was randomized Internal Validity Scale: one statement of random allocation	Not specified	Specified as not reported
Moher 1999[6] Moher 1998[7]	(From Schultz component) Trials that report using either central randomization, numbered or coded bottles or containers, a statement indicating that drugs were prepared by a pharmacy, or serially numbered, opaque	Sealed envelopes without mention of "opaque" are not adequate.	Specified as not reported
Kjaerard 2001[8]	Includes central independent unit, sealed envelopes, or similar) or inadequate (not described or open table of random numbers or similar	Not described or open table random numbers or similar.	N/A
Balk 2002[9]	Was allocation fully concealed? If randomization site was a central or random method was performed by computers, blinded code or blinded medicine vials, or opaque envelopes, allocation was adequately concealed.	Tables, cards, etc were not adequate concealed. Randomization by birth, year, registration, number was not adequately concealed despite where the randomization was performed.	N/A
Sterne et al., 2002[11]	As per Schultz 1995[1]		
Eggar 2003[13]	Includes central randomization, coded drug packs, assignment envelopes, etc.	Reported an inadequate approach, or lacked a statement on concealment	N/A

Table D-3 Allocation concealment definitions (continued)

	Adequate	Inadequate	Unclear/ Not Reported
Pildal 2007[20]	Based on the definition by Schultz 1995[1] Included central randomization including pharmacy-controlled randomization (where a pharmacy remote from the clinical ward allocated the treatment); numbered or coded bottles or containers; serially numbered, opaque, sealed envelopes; or the trialists presented other descriptions that implied convincing concealment Inadequate concealment.	Methods were deemed to provide 'inadequate concealment' if it was obvious to which treatment the next patient would be allocated (alternation, use of case record numbers, dates of birth, etc.).	Trials with 'unclear concealment' did not report on allocation concealment or reported an approach that did not clearly fall into one of the other categories
Siersma 2007[21]	Central randomization, sealed envelopes, identically coded drug boxes, or similar	Based on an open allocation sequence, alternation, or not described	N/A
Wood 2008[23]	As described in each of the 3 included meta-epidemiologic studies: Schultz 1995[1] Kjaergard 1999[8] Egger 2003[13]A		
Herbison 2011[37,38]	Categorized into 6 categories describing aspects of "Adequate concealment" : (G1) Trials that used some form of central randomization that clearly should hide the allocation (e.g., remote telephone service or randomization by a pharmacy).	(G2) Trials that used sealed envelopes with security enhancement (e.g., opaque and numbered); "Inadequate or unclear concealment" included.	(G3) Trials that used sealed envelopes without further details. (G4) Trials that were reported as randomized without details, and also "double-blind". (G5) Trials that simply said they were randomized with no further details. (G6): Trials where the allocation was clearly not hidden (e.g., based on an open list, odd/even days of the week, participant's birth date, team member on duty at enrollment).

Table D-3 Allocation concealment definitions (continued)

	Adequate	Inadequate	Unclear/ Not Reported
Hempel et al., 2011[44,47]	Dataset 1 and 2: Van Tulder 2003: Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the people included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient. Dataset 3: Jadad and Schultz: Jadad: A method to generate the sequence of randomization will be regarded as appropriate if it allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which treatment was next. Schultz: Adequately concealed trial (i.e. central randomization; numbered or coded bottles or containers; drugs prepared by the pharmacy; serially numbered; opaque, sealed envelopes; or other description that contained elements convincing of concealment	Dataset 1 and 2: Van Tulder 2003: Not specified Dataset 3: Jadad Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should be not regarded as appropriate. Schultz: Inadequately concealed trial (i.e. alternation or reference to case record numbers or dates of birth	Dataset 1 and 2: Van Tulder 2003: Not specified Dataset 3: Schultz: Unclearly concealed trial (authors did either not report an allocation concealment approach at all or reported an approach that did not fall into the categories above
Savovic et al., 2012[48-50]	Accepted all definitions of adequate allocation concealment Als-Nielsen vs Balk 2002 Egger 2003, Kjaerkard , Pildal , Royle and Milne , and Schultz Egger 2003 vs Contopoulos-Ionnidis, Pildal Kjaergard vs Balk, 2002, egger 2003, Contopoulos-Ionnidis, Schultz 1995 vs Eggar 2003	Accepted all definitions of inadequate allocation concealment	Accepted all definitions of unclear or not reported allocation concealment (6 different definitions). Evaluated the level of agreement between authors for adequate and inadequate.

Table D-3 Allocation concealment definitions (continued)

	Adequate	Inadequate	Unclear/ Not Reported
Mhaskar et al., 2012 [41]	Cochrane risk of bias tool Central Randomization: The central randomization office was remote from patient recruitment centres. Participant details were provided, for example, by phone, fax or email and the allocation sequence was concealed to individuals staffing the randomization office until a participant was irreversibly registered. Sequentially numbered containers: Drug containers prepared by an independent pharmacy were sequentially numbered and opened sequentially. Containers were of identical appearance, tamper-proof and equal in weight. Sequentially numbered opaque, sealed envelopes: Envelopes were sequentially numbered and opened sequentially only after participant details were written on the envelope. Pressure sensitive or carbon paper inside the envelope transferred the participant's details to the assignment card. Cardboard or aluminium foil inside the envelope rendered the envelope impermeable to intense light. Envelopes were sealed using tamper-proof security tape.	Not specified	Not specified

Table D-3 Allocation concealment definitions (continued)

	Adequate	Inadequate	Unclear/ Not Reported
Studies reporting Effect Size differences, median effect sizes, and effect sizes			
Fenwick et al., 2008[22]	Based on the Cochrane Handbook 2006 defining adequate, unclear and inadequate. Included centralized or pharmacy-controlled randomization; coded identical containers administered serially; onsite computer system combined allocations kept in locked unreadable computer file; sequentially numbered sealed opaque envelopes and similar schemes ensuring the patient and the clinician were unaware of the allocation, along with reassurance that the person who generated the allocation scheme did not administer it.	Included alternation of patients; use of patient data to assign patients to a treatment group, such as the use of case record numbers or dates of birth. Similarly, any procedure that was entirely transparent before allocation such as day of the week or open list of random numbers to allocate to a treatment group.	Trials with 'unclear concealment' included studies that did not report any concealment approach.
Nuesch et al., 2009[29] Nuesch et al., 2009[28]	Central randomization, use of sequentially numbered, sealed, and opaque assignment envelopes, or coded drug packs.		Lack of specific statement was coded as unclear
Van Tulder et al 2009[30]	Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the people included in the trail and has no influence on the assignment sequence or on the decision about eligibility of the patient		
Hamm et al., 2010 [32]	Jadad Cochrane ROB 2008* 2001 Consort	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version
Hartling et al., 2009 [26]	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version
Hartling et al. 2011[34]	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version
Hartling et al., 2012 [3]	Cochrane ROB tool 2008 version Additional decision rules were developed	Cochrane ROB tool 2008 version	Cochrane ROB tool 2008 version

Table D-3 Allocation concealment definitions (continued)

	Adequate	Inadequate	Unclear/ Not Reported
Inaba et al., 2009[27]	References Cochrane Handbook 2008 for all biases. Not detailed but combined BOTH adequate sequence generation and allocation concealment and called this adequate randomization.		Not detailed but considered both adequate sequence generation and allocation concealment and called this adequate randomization.
Clifford et al., 2002[10]	Definitions as per Schulz 1994[1]		
Derry et al., 2006[18]	Not specified		Not specified

* Modified Cochrane ROB 2008 with some additional decision rules. Use these decision rules in addition to the guidelines outlined in the Cochrane criteria. If the randomization is conducted by central telephone, pharmacy, etc., assume this is adequate and answer YES.

Table D-4. Sequence generation definitions

Study Identification	Adequate	Inadequate	Unclear
Studies reporting Odds Ratios (OR) or Ratio of Odds Ratios (ROR) or Ratio of Relative Risks = (RRR)			
Schulz et al., 1995[1]	Reported using random number table, computer random number generator, coin tossing, or shuffling.	Did not report one of the adequate approaches	Not specified
Moher et al., 1998[6,7]	Jadad scale: Receive a point if described as randomized. Receives another point if the trial describes the method of randomization, such as random number, computer generated. Schultz components. Trials that report using a random numbers table, computer random number, coin tossing, dice throwing, and shuffling	Jadad scale: If describes as randomized but was inappropriate such as date of birth, hospital numbers	Not specified
Linde et al., 1999[4]	Based on the following scales: Jadad Scale: explicit statement that trial was randomized	Not specified	Specified as not reported
Kjaergard, et al. 2001[8]	Includes computer generated random numbers or similar	Not described	Not specified
Balk et al., 2002[9]	Question: “Any description of how randomization (allocation among treatment arms) was achieved”	Question: “or did the article say only “randomization?”	
Siersma et al., 2007 [21]	Computer-generated, random number table, or similar	Quasi-randomized or not described	

Table D-4. Sequence generation definitions (continued)

Study Identification	Adequate	Inadequate	Unclear
Hempel et al 2011[44,47]	Group 1 & 2: Van Tulder: Was the method of randomization adequate? 2) Were the groups similar at baseline regarding most of the most important prognostic indicators? **Randomization sequence** A random (unpredictable) assignment sequence. Examples of adequate methods are coin toss (for studies with two groups), rolling a dice (for studies with two or more groups), drawing of balls of different colours, drawing of ballots with the study group labels from a dark bag, computer generated random sequence, preordered sealed envelopes, sequentially-ordered vials, telephone call to a central office, and preordered list of treatment assignments Examples of inadequate methods are: alternation, birth date, social insurance/security number, date in which they are invited to participate in the study, and hospital registration number **Baseline comparability** In order to receive a "yes", groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s). Group 3: Jadad and Schultz Jadad: Was the study described as randomized (this includes the use of words such as randomly, random, and randomization)? = 1 point Schultz: Adequately sequence generation (random-number table, computer random-number generator, coin tossing, or shuffling)	Group 3: Jadad: Deduct 1 point if: For question 1, the method to generate the sequence of randomization was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc.)	

Table D-4. Sequence generation definitions (continued)

Study Identification	Adequate	Inadequate	Unclear
Savovic et al., 2012[48-50]	Accepted all definitions (6 different ones) of adequate generation from included groups of studies. Evaluated the level of agreement between authors for adequate and inadequate ratings.	Accepted all definitions of inadequate generation.	Accepted all definitions of unclear or not reported sequence generation.
Mhaskar et al., 2012[41]	Cochrane Handbook 5.1.0 The use of a random component should be sufficient for adequate sequence generation. Simple methods such as repeated coin-tossing, throwing dice or dealing previously shuffled cards. More usually it is achieved by referring to a published list of random numbers, or to a list of random assignments generated by a computer. In trials using large samples (usually meaning at least 100 in each randomized group, simple randomization generates comparison groups of relatively similar sizes.	Systematic methods, such as alternation, assignment based on date of birth, case record number and date of presentation are sometimes referred to as 'quasi-random'. Alternation (or rotation, for more than two intervention groups) might in principle result in similar groups, but many other systematic methods of sequence generation may not. Concealing the allocation schedule is usually impossible, which allows foreknowledge of intervention assignment among those recruiting participants to the study, and biased allocations	A simple statement such as 'we randomly allocated' or 'using a randomized design' is insufficient and there is doubt, then the adequacy of sequence generation should be considered to be unclear. In the case of blocked randomization was used, but the process of selecting the blocks, such as a random number table or a computer random number generator, was not specified. The adequacy of sequence generation should then be classified as unclear.
Studies reporting Effect Size differences, median effect sizes, and effect sizes			
Van Tulder et al 2009[30]	A random (unpredictable) assignment sequence. Examples of adequate methods are computer generated random No. table and sue of sealed opaque envelopes.	Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.	
Hamm et al., 2010[32]	Use these decision rules in addition to the guidelines outlined in the Cochrane criteria. If blocked randomization, permutation, or stratification is specified, assume the randomization sequence was computer-generated and answer YES.	Use these decision rules in addition to the guidelines outlined in the Cochrane criteria.	Use these decision rules in addition to the guidelines outlined in the Cochrane criteria. If the description only includes 'random', 'randomly generated', 'randomized', etc., do not assume additional details and answer UNCLEAR.

Table D-4. Sequence generation definitions (continued)

Study Identification	Adequate	Inadequate	Unclear
Hartling et al., 2009[26]	Cochrane risk of bias tool 2008	Cochrane risk of bias tool 2008	Cochrane risk of bias tool 2008
Hartling et al. 2011[51]	Cochrane risk of bias tool 2008	Cochrane risk of bias tool 2008	Cochrane risk of bias tool 2008
Hartling et al., 2012[3]	Cochrane Risk of Bias tool 2008 "YES" if any mention of how the randomization list was generated "YES" if authors mention use of stratification or permuted blocking (use of computer implied)	Cochrane Risk of Bias tool 2008	Cochrane Risk of bias tool 2008 "UNCLEAR" if description only includes 'random', 'randomly generated', 'randomized', etc.

References for Appendix D

1. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12. PMID: 7823387.
2. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282(11):1054-60. PMID: 10493204.
3. Nurmohamed MT, Rosendaal FR, Buller HR, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet. 1992 Jul 18;340(8812):152-6. PMID: 1352573.
4. Linde K, Scholz M, Ramirez G, et al. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol. 1999;52(7):631-6.
5. Tang JL. Weighting bias in meta-analysis of binary outcomes. J Clin Epidemiol. 2000;53(11):1130-6. PMID: 11106886.
6. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess. 1999;3(12):i-iv, 1-98. PMID: 10374081.
7. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352(9128):609-13. PMID: 9746022.
8. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135(11):982-9. PMID: 11730399.
9. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA. 2002;287(22):2973-82. PMID: 12052127.
10. Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? Results of a pilot study. BMC Health Services Research. 2002;2(1):18. PMID: 12213183.
11. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11):1513-24. PMID: 12111917.
12. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA. 2003;290(7):921-8. PMID: 12928469.
13. Egger M, Juni P, Bartlett C, et al. 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.
14. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291(20):2457-65. PMID: 15161896.
15. Chan AW, Krleza-Jeric K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ. 2004;171(7):735-40. PMID: 15451835.
16. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst. 2005;97(14):1043-55. PMID: 16030302.
17. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol. 2005;34(1):79-87. PMID: 15561753.
18. Derry CJ, Derry S, McQuay HJ, et al. Systematic review of systematic reviews of acupuncture published 1996-2005. Clin Med. 2006;6(4):381-6. PMID: 16956145.
19. Mohammed S, Goodacre S. Intravenous and nebulised magnesium sulphate for acute asthma: systematic review and meta-analysis. Emerg Med J. 2007 Dec;24(12):823-30. PMID: 18029512.

20. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol. 2007;36(4):847-57. PMID: 17517809.

21. Relevo R, Balshem H. Finding evidence for comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1168-77. PMID: 21684115.

22. Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol. 2008;35(9):775-82. PMID: 18840153.

23. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5. PMID: 18316340.

24. Wartenberg D. Residential magnetic fields and childhood leukemia: a meta-analysis. Am J Public Health. 1998;88(12):1787-94. PMID: 9842375.

25. Feng J, Wan G, Zhu X, et al. The apolipoprotein e (APOE) gene and the risk of diabetic nephropathy (DN): A meta-analysis in East Asian populations. Asian Biomedicine. 2010;4(2):329-35.

26. Treadwell J, Uhl S, Tipton K, et al. Assessing Equivalence and Noninferiority. Methods Research Report. (Prepared by the EPC Workgroup under Contract No. 290-2007-10063.) AHRQ Publication No. 12-EHC045-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov.

27. Inaba Y, Chen JA, Mehta N, et al. Impact of single or multicentre study design on the results of trials examining the efficacy of adjunctive devices to prevent distal embolisation during acute myocardial infarction. Eurointervention. 2009;5(3):375-83. PMID: 19736164.

28. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339:b3244. PMID: 19736281.

29. Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum. 2009;61(12):1633-41. PMID: 19950329.

30. van Tulder MW, Suttorp M, Morton S, et al. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine (Phila Pa 1976). 2009 Jul 15;34(16):1685-92. PMID: 19770609.

31. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials [Electronic Resource]. 2010;11:52. PMID: 20462436.

32. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]: The Cochrane Collaboration; 2008.

33. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365. PMID: 20156912.

34. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. [Review]. PLoS ONE [Electronic Resource]. 2011;6(2):e17242. PMID: 21390219.

35. Hempel S, Miles J, Suttorp M, et al. Detection of Associations between Trial Quality and Effect Sizes. Methods Research Report. Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I AHRQ Publication No. 12-EHC010-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2012.

36. Hempel S, Suttorp MJ, Miles JNV, et al. Empirical Evidence of Associations Between Trial Quality and Effect Sizes. Methods Research Report (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I) AHRQ Publication No. 11-EHC045-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2011. http://effectivehealthcare.ahrq.gov.

37. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. [Review] [60 refs]. J Clin Epidemiol. 2006;59(12):1249-56. PMID: 17098567.

38. Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. J Clin Epidemiol. 2011;64(10):1070-5. PMID: 21474279.

39. Liu CJ, LaValley M, Latham NK. Do unblinded assessors bias muscle strength outcomes in randomized controlled trials of progressive resistance strength training in older adults? Am J Phys Med Rehabil. 2011 Mar;90(3):190-6. PMID: 21173683.

40. Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012;344:e1119. PMID: 22371859.

41. Mhaskar R, Djulbegovic B, Magazin A, et al. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012 Jun;65(6):602-9. PMID: 22424985.

42. Savovic J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess. 2012 Sep;16(35):1-82. PMID: 22989478.

43. Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012 Sep 18;157(6):429-38. PMID: 22945832.

44. Als-Nielsen B, Chen W, Gluud LL, et al. Are trial size and reported methodological quality associated with treatment effects? Program & Abstract Book, 12th Cochrane Colloquium 2004. 2004:102-3.

45. Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat.Med. 2007;26(14):2745-58.

46. Contopoulos-Ioannidis DG, Gilbody SM, Trikalinos TA, et al. Comparison of large versus smaller randomized trials for mental health-related interventions. Am J Psychiatry. 2005 Mar;162(3):578-84. PMID: 15741476.

47. Linde K, Ramirez G, Mulrow CD, et al. St John's wort for depression--an overview and meta-analysis of randomised clinical trials. BMJ. 1996 Aug 3;313(7052):253-8. PMID: 8704532.

48. Lim B, Manheimer E, Lao L, et al. Acupuncture for treatment of irritable bowel syndrome. Cochrane Database Syst Rev. 2006;Issue 4(4):CD005111. PMID: 17054239.

49. Palma S, Delgado-Rodriguez M. Assessment of publication bias in meta-analyses of cardiovascular diseases. J Epidemiol Community Health. 2005 Oct;59(10):864-9. PMID: 16166360.

50. Savović J, Harris RJ, Wood L, et al. Development of a combined database for meta-epidemiological research. Research Synthesis Methods. 2010;1(3-4):212-25.

51. Lin DY. Cox regression analysis of multivariate failure time data: the marginal approach. Stat Med. 1994;13(21):2233-47. PMID: 7846422.

Appendix E. Study Quality Assessment

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
Schulz 1995[1]	Census of Cochrane database of perinatal trials	Yes	Yes	Specific classifications were defined, e.g., if allocation sequence was of a certain type, then generation was adequate.	Not given	No justification for validity of bias measures provided. Ten trials were evaluated by two reviewers for reliability.	Not reported	No, only to pregnancy and childbirth	Yes
Juni 1999[2]	Studies were selected from a single systematic review which had 17eligible trials and all were assessed for quality using 25 different quality checklists.	Yes All three [concealed randomization, blinding, and treatment of withdrawals] were modeled in a multivariate model	Yes	Unclear how the ratings on 25 scales were distilled into component scores.	Intraclass correlation >0.9 for 12 scales (48%), 0.8 to 0.9 for 10 scales (40%), and >0.8 for 3 scales (12%)	Unclear	The trials included in a previous systematic review. All 17 trials were included in the analysis.	Results apply to multiple types of general surgery.	Not applicable (N/A) as a single MA was used.
Linde 1999[3]	Census	Yes	Measured bias categories dichoto-mously	Reporting in article	Not given	Unknown	Census	Yes (homeopathy was applied to many different	NR
Moher 1998, 1999[4,5]	12 meta-analyses were randomly selected (1 was subsequently excluded) from a series of MA	No	Yes	Adequate vs. inadequate based on explicit criteria	NR	Based on Jadad and Schulz's component for biases of interests	NR	Yes (multiple interventions) and multiple clinical conditions	Yes

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
	selected from the literature. The trials within the 11 eligible meta-analyses were used for the analysis.								
Kjaegard 1999[6]	Studies were selected from MA that had at least one trial with a sample size of greater than 1000 subjects during a specified time interval	Yes	Yes	Adequate vs. inadequate	Intraclass correlation for inter-observer reliability = 0.96 (95% CI, 0.92 to 0.98) Test-retest reliability: 0.98 (95% CI 0.97-0.99)	Based on Jadad questions	NR	Yes but not disease areas were not specified.	NR
Balk 2002[7]	Census of all MAs that met their inclusion criteria available in MEDLINE (1966-2000) and Cochrane (2000; issue 3)	No	Yes	Unclear	Not given though they did pilot study and calibrated reviewers; Dual review with disagreements resolved by third reviewer	28 items following a review of quality measures items	NR	Applicable to four clinical areas (Cardiovascular, Infectious diseases Pediatrics and General Surgery)	Not clear
Clifford 2002[8]	No	No	No	NA	No	Yes: used Jadad scale	Convenience sample	Clinical areas not specified	N/A
Sterne 2002[9]	As per Schultz	No	Yes	NR	NR	NR	NR	Unknown	NR
Als-Nielson, 2003[10]	Random sample of Cochrane reviews obtained in May 2001	No	Yes	Based on whether study reported that it was double-blinded	Not specified	Not provided	Not calculated	Yes	Not relevant
Egger 2003[11]	Yes	Yes	Yes	Presence or absence of blinding and	NR	Detailed definition	NR	Yes	NR

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
	Included 122 meta-analyses from 1998 issue of CDSR. From this excluded MA that did not adequate asses the bias. From this selected trials with and without bias of interest.			adequate vs inadequate allocation concealment					
Chan 2004[12]	Yes	No	Yes	Distinction between fully/partially and qualitatively/ unreported is whether studies provided any information on results more substantive than just P values and statements of statistical significance	NA	NA	NR	Applicability of results may be somewhat limited by the period of protocol approval (1990-1998); more recent trials may list more detailed results	N/A
Chan 2004[13]	Yes	No	Yes	Distinction between fully/partially and qualitatively/ unreported is whether studies provided any information on results more substantive than just P values and statements of statistical significance	NA	NA	NR		N/A
Kyzas 2005[14]	Yes	No	Yes	Blinded vs. not stated	NR	NR	NR		Yes
Tierney 2005[15]	Unknown	No	NA	NA	NA	NA	NR	All data came from cancer	Yes

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
								trials.	
Derry 2006[16]	Yes, -studies searched for in literature search from root to time interval.	Yes--randomization, blinding, sample size, and overall quality	NA	N/A	NR	Can't answer	Arbitrarily specified 4 trials and/or 200 patients as a minimum for sufficient quality/validity for calculating statistical significance	Yes, multiple clinical conditions but a single intervention--acupuncture	No
Furukawa 2007[17]	No, random selection applied to census	no	Yes	Contribute to meta-analysis or not	Not specified	Not provided	"A power calculation based on their regression coefficients indicated that detecting statistically significant coefficients of this magnitude required approximately 100 reviews."	Yes	Unclear
Pildal 2007[18]	Random sample of Cochrane reviews in 2003; PubMed reviews chosen in publication order	No	Yes	See measurement of bias column	NR	Schulz approach to define adequate concealment of allocation	Convenience	Unclear, a random sample of Cochrane reviews ordered by publication sample of PubMed reviews	Only 2 trials were duplicated
Siersma 2007[19]	Random sample of Cochrane reviews	No	Yes	Based on reporting in the study	Not specified	Not provided	Not calculated	Yes	Not relevant
Fenwick	Census from a	No	3 categories	Commonly used	No	Good	From universe	No	Yes

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
2008[20]	single database Cochrane		for allocation and 2 for blinding	criteria			of studies (Cochrane is not the universe of MA and studies)		
Wood 2008[21]	Census from three existing meta-epidemiological databases	Yes	NA	Based on data from three previous studies (adequate vs. inadequate/ unclear	NA	Based on previous studies	Not calculated	Yes	Yes
Inaba et al., 2009[22]	The sample was not a census. Trials were selected from a wide bibliographic search of related studies.	Yes	Yes	Presence or absence of the bias	NR	NR	NR	No only some cardiac procedures	Unclear
Nuesch 2009[23,24]	Yes	Yes: Some	No: N/A	N/A	Yes dual review reconciliation	Yes	NR	No	Yes
Van Tulder 2009[25]	Census	No	Yes	Yes vs. no or don't know	NR	Good	used the universe of available studies	No	Yes
Dwan 2010[26]	Yes All trials included in the systematic review 'Intravenous and nebulised magnesium sulphate for acute asthma' were assessed.	No	No	N/A for selective outcome reporting	NR	N/A	NR	No Limited to asthma	No
Kirkham 2010[27]	Yes--all SRs in issue 4, 2006,	No	No	NA	NR	NA	NR	Yes	Yes, by conducting

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
	Issue 1, 2007, and Issue 2, 2007 were searched for inclusion criteria								sensitivity analyses only for reviews that had a single M-A of the review primary outcome.
Hartling 2011[28]	Yes, census of studies meeting criteria	No	No	Low vs. high or unclear; Low or unclear vs. high	Interrater agreement for the majority of domains and overall risk of bias was moderate (k = 0.41–0.60).	Cochrane risk of bias tool	Not calculated	No	Yes
Herbison 2011[29,30]	Yes	No	Yes, in some analyses	By type of allocation approach (6 types)	Yes	Valid	NR	Yes	Yes
Liu 2011[31]	Census from review	Yes	Yes (bias measures)	Blinded, ITT	NR	Based on report in article	NR	No	No

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
Savovic 2012[32]	Yes, from the census of all previously published meta-epidemiological studies AND that provided data to the database	Yes, an extensive methodology was used to remove duplicates, check for completeness of the trial characteristics, and tested reliability of the rating for risk of bias	Yes	Presence, unclear, or absence of the bias	Kappa statistics ranged from 0.55 to 1.00 (median 0.87).	Data was obtained across a large number of studies and so it likely to be varied.	Not calculated	Yes	Yes
Hartling 2012[33]	Census from Hopewell study and then sample of cohort studies	Yes, it accounted for other biases through meta-regression	Yes	Yes vs. no. vs. unclear	Yes for raters and different study designs	Risk of Bias and Newcastle Ottawa scales are valid measures	Yes, but indicated these were arbitrary	Yes	Yes
Hróbjartsson 2012[34]	Census of studies that met criteria	Yes	Yes	Presence or absence of blinded assessors	No	Based on original author info	Census of available studies	Not clear, a small number of studies was used that covered a variety of treatment areas	NA

Study ID	Were studies selected by a census? If no, how was the sample selected?	Did the study account for confounding or interaction among sources of bias?	Did the study use dichoto-mous measures for high vs. low quality?	If the study used dichotomous measures for high vs. low quality, how was the threshold determined?	IRR of risk of bias measure	Validity of risk of bias measure	How was sample size calculated?	Do the findings of this meta-epidemio-logical study apply to multiple clinical areas?	Did the study account for duplication of trials?
Mhaskar 2012[35]	Yes	No	yes	Presence or absence of the bias	Not specified	Not provided	Not calculated	No	Not applicable, but accounted for duplication of citations for each trial

Abbreviations: ITT = intent to treat; MA = meta-analysis; NA = not applicable; NR = not reported; SR = systematic review;

References for Appendix E

1. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12. PMID: 7823387.
2. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282(11):1054-60. PMID: 10493204.
3. Linde K, Scholz M, Ramirez G, et al. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol. 1999;52(7):631-6.
4. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352(9128):609-13. PMID: 9746022.
5. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess. 1999;3(12):i-iv, 1-98. PMID: 10374081.
6. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135(11):982-9. PMID: 11730399.
7. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA. 2002;287(22):2973-82. PMID: 12052127.
8. Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: are they related? Results of a pilot study. BMC Health Services Research. 2002;2(1):18. PMID: 12213183.
9. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11):1513-24. PMID: 12111917.
10. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA. 2003;290(7):921-8. PMID: 12928469.
11. Egger M, Juni P, Bartlett C, et al. 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.
12. Chan AW, Krleza-Jeric K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ. 2004;171(7):735-40. PMID: 15451835.
13. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291(20):2457-65. PMID: 15161896.
14. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst. 2005;97(14):1043-55. PMID: 16030302.
15. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol. 2005;34(1):79-87. PMID: 15561753.
16. Derry CJ, Derry S, McQuay HJ, et al. Systematic review of systematic reviews of acupuncture published 1996-2005. Clin Med. 2006;6(4):381-6. PMID: 16956145.
17. Furukawa TA, Watanabe N, Omori IM, et al. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. JAMA. 2007;297(5):468-70.
18. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol. 2007;36(4):847-57. PMID: 17517809.
19. Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat.Med. 2007;26(14):2745-58.
20. Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol. 2008;35(9):775-82. PMID: 18840153.

21. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5. PMID: 18316340.

22. Inaba Y, Chen JA, Mehta N, et al. Impact of single or multicentre study design on the results of trials examining the efficacy of adjunctive devices to prevent distal embolisation during acute myocardial infarction. Eurointervention. 2009;5(3):375-83. PMID: 19736164.

23. Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum. 2009;61(12):1633-41. PMID: 19950329.

24. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339:b3244. PMID: 19736281.

25. van Tulder MW, Suttorp M, Morton S, et al. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine (Phila Pa 1976). 2009 Jul 15;34(16):1685-92. PMID: 19770609.

26. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials [Electronic Resource]. 2010;11:52. PMID: 20462436.

27. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365. PMID: 20156912.

28. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. [Review]. PLoS ONE [Electronic Resource]. 2011;6(2):e17242. PMID: 21390219.

29. Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. J Clin Epidemiol. 2011;64(10):1070-5. PMID: 21474279.

30. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. [Review] [60 refs]. J Clin Epidemiol. 2006;59(12):1249-56. PMID: 17098567.

31. Liu CJ, LaValley M, Latham NK. Do unblinded assessors bias muscle strength outcomes in randomized controlled trials of progressive resistance strength training in older adults? Am J Phys Med Rehabil. 2011 Mar;90(3):190-6. PMID: 21173683.

32. Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012 Sep 18;157(6):429-38. PMID: 22945832.

33. Hartling L, Hamm M, Milne A, et al. Validity and inter-rater reliability testing of quality assessment instruments. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2007-10021-I.) AHRQ Publication No. 12-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

34. Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012;344:e1119. PMID: 22371859.

35. Mhaskar R, Djulbegovic B, Magazin A, et al. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012 Jun;65(6):602-9. PMID: 22424985