



Cochrane
Library

Cochrane Database of Systematic Reviews

Randomisation to protect against selection bias in healthcare trials (Review)

Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD

Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD.
Randomisation to protect against selection bias in healthcare trials.
Cochrane Database of Systematic Reviews 2011, Issue 4. Art. No.: MR000012.
DOI: [10.1002/14651858.MR000012.pub3](https://doi.org/10.1002/14651858.MR000012.pub3).

www.cochranelibrary.com

TABLE OF CONTENTS

| | |
|--|----|
| ABSTRACT | 1 |
| PLAIN LANGUAGE SUMMARY | 2 |
| BACKGROUND | 3 |
| OBJECTIVES | 3 |
| METHODS | 3 |
| RESULTS | 5 |
| Figure 1. | 6 |
| Figure 2. | 7 |
| DISCUSSION | 7 |
| AUTHORS' CONCLUSIONS | 9 |
| ACKNOWLEDGEMENTS | 9 |
| REFERENCES | 11 |
| CHARACTERISTICS OF STUDIES | 17 |
| DATA AND ANALYSES | 38 |
| Analysis 1.1. Comparison 1 Studies of randomised versus non-randomised studies of the same intervention and condition, Outcome 1 Undefined. | 38 |
| Analysis 3.1. Comparison 3 Studies of randomised versus non-randomised studies across different interventions for the same condition, Outcome 1 Undefined. | 39 |
| Analysis 4.1. Comparison 4 Studies of randomised versus non-randomised studies across different interventions and conditions, Outcome 1 Undefined. | 39 |
| Analysis 7.1. Comparison 7 Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions for the same condition, Outcome 1 Undefined. | 41 |
| Analysis 8.1. Comparison 8 Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions and conditions, Outcome 1 Undefined. | 42 |
| ADDITIONAL TABLES | 43 |
| APPENDICES | 44 |
| FEEDBACK | 45 |
| WHAT'S NEW | 47 |
| HISTORY | 47 |
| CONTRIBUTIONS OF AUTHORS | 48 |
| DECLARATIONS OF INTEREST | 48 |
| SOURCES OF SUPPORT | 48 |
| DIFFERENCES BETWEEN PROTOCOL AND REVIEW | 48 |
| INDEX TERMS | 48 |

[Methodology Review]

Randomisation to protect against selection bias in healthcare trials

Jan Odgaard-Jensen¹, Gunn E Vist², Antje Timmer³, Regina Kunz⁴, Elie A Akl⁵, Holger Schünemann⁶, Matthias Briel⁷, Alain J Nordmann⁸, Silvia Pregno⁹, Andrew D Oxman¹

¹Global Health Unit, Norwegian Knowledge Centre for the Health Services, Oslo, Norway. ²Prevention, Health Promotion and Organisation Unit, Norwegian Knowledge Centre for the Health Services, Oslo, Norway. ³Department of Health Services Research, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany. ⁴asim, Swiss Academy of Insurance Medicine, University of Basel Hospital, Basel, Switzerland. ⁵Department of Internal Medicine, American University of Beirut, Beirut, Lebanon. ⁶Departments of Clinical Epidemiology and Biostatistics and of Medicine, McMaster University, Hamilton, Canada. ⁷Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel (USB), Basel, Switzerland. ⁸Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland. ⁹Cattedra di Statistica Medica, University of Modena and Reggio Emilia, 41100 Modena, Italy

Contact: Jan Odgaard-Jensen, Global Health Unit, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs Plass, Oslo, N-0130, Norway. jan@nokc.no.

Editorial group: Cochrane Methodology Review Group.

Publication status and date: Edited (no change to conclusions), comment added to review, published in Issue 4, 2015.

Citation: Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2011, Issue 4. Art. No.: MR000012. DOI: [10.1002/14651858.MR000012.pub3](https://doi.org/10.1002/14651858.MR000012.pub3).

Copyright © 2015 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Randomised trials use the play of chance to assign participants to comparison groups. The unpredictability of the process, if not subverted, should prevent systematic differences between comparison groups (selection bias). Differences due to chance will still occur and these are minimised by randomising a sufficiently large number of people.

Objectives

To assess the effects of randomisation and concealment of allocation on the results of healthcare studies.

Search methods

We searched the Cochrane Methodology Register, MEDLINE, SciSearch and reference lists up to September 2009. In addition, we screened articles citing included studies (ISI Science Citation Index) and papers related to included studies (PubMed).

Selection criteria

Eligible study designs were cohorts of studies, systematic reviews or meta-analyses of healthcare interventions that compared random allocation versus non-random allocation or adequate versus inadequate/unclear concealment of allocation in randomised trials. Outcomes of interest were the magnitude and direction of estimates of effect and imbalances in prognostic factors.

Data collection and analysis

We retrieved and assessed studies that appeared to meet the inclusion criteria independently. At least two review authors independently appraised methodological quality and extracted information. We prepared tabular summaries of the results for each comparison and assessed the results across studies qualitatively to identify common trends or discrepancies.

Main results

A total of 18 studies (systematic reviews or meta-analyses) met our inclusion criteria. Ten compared random allocation versus non-random allocation and nine compared adequate versus inadequate or unclear concealment of allocation within controlled trials. All studies were at high risk of bias.

For the comparison of randomised versus non-randomised studies, four comparisons yielded inconclusive results (differed between outcomes or different modes of analysis); three comparisons showed similar results for random and non-random allocation; two comparisons had larger estimates of effect in non-randomised studies than in randomised trials; and two comparisons had larger estimates of effect in randomised than in non-randomised studies.

Five studies found larger estimates of effect in trials with inadequate concealment of allocation than in trials with adequate concealment. The four other studies did not find statistically significant differences.

Authors' conclusions

The results of randomised and non-randomised studies sometimes differed. In some instances non-randomised studies yielded larger estimates of effect and in other instances randomised trials yielded larger estimates of effect. The results of controlled trials with adequate and inadequate/unclear concealment of allocation sometimes differed. When differences occurred, most often trials with inadequate or unclear allocation concealment yielded larger estimates of effects relative to controlled trials with adequate allocation concealment. However, it is not generally possible to predict the magnitude, or even the direction, of possible selection biases and consequent distortions of treatment effects from studies with non-random allocation or controlled trials with inadequate or unclear allocation concealment.

PLAIN LANGUAGE SUMMARY

Randomised controlled trials as a safeguard against biased estimates of treatment effects

Randomised controlled trials (RCTs) use the play of chance to allocate participants to comparison groups to prevent selection bias. Other means of treatment allocation are more prone to bias because decisions about which treatment to use can be influenced by the preferences of the physician or patient. This review compares random allocation (allocated to treatment using a random method) versus non-random allocation (allocated to treatment using a non-random method, such as alternation or external, uncontrollable factors, with no clinical judgement involved) and controlled trials with adequate versus inadequate/unclear concealment of allocation. Concealed treatment allocation is best described in general terms as the process used to prevent foreknowledge of group assignment in a controlled trial (such as the use of sequentially numbered opaque, sealed envelopes).

The results of randomised and non-randomised studies sometimes differed. Sometimes non-randomised studies yielded larger estimates of effect, and sometimes randomised trials yielded larger estimates of effect. On the other hand, not using concealed random allocation resulted in larger estimates of effect, but sometimes it resulted in similar estimates of effect (from harmful to beneficial or vice versa). It is a paradox that the unpredictability of random allocation is the best protection against the unpredictability of the extent to which non-randomised studies may be biased.

BACKGROUND

Discoveries of dramatically effective healthcare interventions, like epinephrine for anaphylaxis, are not common. The majority of healthcare interventions are at best moderately superior to conventional care or a placebo. Some interventions that are believed to be beneficial are, in fact, no more effective than a placebo and some are even harmful. Well-intentioned clinicians have, for example, treated stroke by applying leeches to the anus (Gubler 1971), treated neurosyphilis by injecting malarial parasites (Austin 1992), treated angina with internal mammary artery ligation (Valenstein 1986), treated symptomatic atherosclerotic disease of the internal carotid artery with extracranial-intracranial bypass surgery (EC/IC Bypass 1985), and treated asymptomatic ventricular arrhythmia after myocardial infarction with class I antiarrhythmic drugs (Echt 1991). It has been estimated that tens of thousands of patients died prematurely from widespread use of class I antiarrhythmic drugs alone (Moore 1995), which caused one death for every 20 patients who were treated (Teo 1993). Failure to evaluate interventions adequately has also delayed the use of effective interventions, such as magnesium sulphate instead of diazepam or phenytoin for the treatment of eclampsia (Eclampsia 1995).

As stated by Archie Cochrane: "Observational evidence is clearly better than opinion, but it is thoroughly unsatisfactory. All research on the effectiveness of therapy was in this unfortunate state until the early 1950s. The only exceptions were the drugs whose effects on immediate mortality were so obvious that no randomised trials were necessary, such as insulin, sulphonamide, and penicillin" (Cochrane 1972). Cochrane, along with many others, credits Austin Bradford Hill with bringing an experimental approach into clinical medicine. The 1948 report of the randomised trial of streptomycin for pulmonary tuberculosis by Hill and his colleagues is widely recognised as a landmark study in this regard (MRC 1948).

"The basic idea, like most good things, is very simple" (Cochrane 1972). The primary reason for random assignment is to remove the potential of bias in the assignment of people to one intervention or another, i.e. to protect against any possible systematic connection between the treatment that people receive and their prognosis. In addition to producing comparable groups of treatment and control patients, which other means of allocation such as alternation can also do, concealed randomisation introduces unpredictability. When alternation or any other pre-set schedule (such as time of admission) is used, a clinician who wants a patient to receive a particular treatment can learn the schedule and might then be able to arrange to enter a patient into the study at an opportune moment. If randomisation has been successfully concealed, however, each patient's treatment will be assigned according to the play of chance. This unpredictability, unless subverted by clinicians who find a way to access the randomisation schedule in advance, should prevent systematic differences in the prognosis of the groups of patients that are being compared. Differences due to chance will still occur and these are minimised by randomising a sufficiently large number of people. Although it is possible to control for differences between comparison groups in other ways, such as statistical adjustment of the analyses, this is only possible for factors that are known and measured. Randomisation is the only means of controlling for unknown and unmeasured factors as well as those that are known and measured.

Despite this simple logic and many anecdotal examples of harm being done because of delays in conducting randomised trials, there are limitations to the use of randomised trials, both real and imagined, and scepticism about the importance of randomisation (Black 1996; Pocock 2000; US Office HTA 1994; Weiss 1998). We believe this scepticism is healthy. It is important to question assumptions about research methods, and to test these assumptions empirically, just as it is important to test assumptions about the effects of health care. Methodological hubris can be just as dangerous as medical hubris. Empirical comparisons of randomised versus non-randomised evaluations of the effects of health care represent important steps away from hubris. This review of such comparisons has been updated from previously published reviews (Kunz 1998, Kunz 2002, see What's new). This review differs from other similar reviews (McKee 1999; Reeves 1998) in the questions that are addressed and the methods that were used, but there is not a major disagreement in the conclusions of these reviews (Britton 1999; Kunz 1999).

Previous versions of this review included a comparison of high and low quality randomised controlled trials (RCTs). However, Jüni and colleagues demonstrated a limited ability of scores to distinguish reliably between high and low quality studies and thus caution against the use of scores (Jüni 1999). This finding has to be taken into account and should lead to a careful interpretation of comparisons between high and low quality studies. Given the results of Jüni et al's analysis and the inability to determine the extent to which any differences in effects can be attributed to randomisation or concealment of allocation, we previously concluded that we would not include the comparison of high and low quality RCTs in this or future updates of this review (Kunz 2002).

Previous versions of this review also included systematic reviews or meta-analyses that incidentally included a comparison of randomised versus non-randomised studies, or randomised trials with and without adequate allocation concealment. In this update we excluded these analyses because they generally lack power and because reporting bias is likely to have occurred (preferentially reported if results were positive). Furthermore, obtaining all such comparisons or an unbiased sample of them would require undertaking a new methodological study, similar to those that are included in this review of methodological studies.

OBJECTIVES

To assess the effects of random allocation and allocation concealment on the results of healthcare studies.

METHODS

Criteria for considering studies for this review

Types of studies

Eligible study designs were cohorts of studies, systematic reviews or meta-analyses that compared random allocation to non-random allocation, or concealment of allocation (the process used to prevent foreknowledge of group assignment in a randomised trial) to non-concealed allocation. We excluded single case studies, systematic reviews or meta-analyses that incidentally reported a comparison of interest, and simulation studies.

Types of data

Studies included were systematic reviews and meta-analyses based on healthcare trials, including trials of clinical interventions ('clinical trials') and non-clinical interventions where the effects of the intervention on one or more health outcomes were measured.

For the comparison of random allocation to non-random allocation the studies considered included:

- randomised trials ('randomised controlled trials' or 'RCTs');
- non-randomised trials with concurrent controls (experimental and prospective studies, in which a non-random but systematic method of allocation, such as alternation, was used to assign participants to the comparison groups; frequently called 'quasi-randomised trials', 'concurrently controlled trials' or 'CCTs');
- non-equivalent control group design (experimental and prospective studies in which a non-random and unsystematic method of allocation was used to assign participants to comparison groups and the allocation was not at the discretion of the person who enrolled the participant, e.g. participants being all patients eligible and referred for the intervention; intervention group being all patients who actually received the intervention; and control group being all patients that did not receive the intervention due to lack of beds, personnel or necessary equipment, but instead received treatment as usual).

For the comparison of random allocation to non-random allocation we excluded cohorts, systematic reviews or meta-analyses that considered the following designs as using non-random allocation:

- studies using historical controls (patients treated earlier than those who received the intervention that is being evaluated, frequently called 'historically controlled trials' or 'HCTs');
- classical observational studies, including cohort studies, cross-sectional studies, case-control studies and 'outcomes studies' (evaluations using large administrative or clinical databases).

For the comparison of adequate concealment of allocation to inadequate or unclear concealment of allocation we included studies considering only controlled trials with some sort of random assignment ('RCTs' and 'quasi-randomised trials'). We excluded studies considering observational study designs.

Types of methods

1. Randomised versus non-randomised studies of the same intervention and condition.
2. Randomised versus non-randomised studies of the same intervention for different conditions.
3. Randomised versus non-randomised studies across different interventions for the same condition.
4. Randomised versus non-randomised studies across different interventions and conditions.
5. Controlled trials with adequate versus inadequate or unclear concealment of allocation of the same intervention and condition.
6. Controlled trials with adequate versus inadequate or unclear concealment of allocation of the same intervention for different conditions.

7. Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions for the same condition.
8. Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions.

Types of outcome measures

The magnitude and direction of estimates of effect (e.g. relative risk reductions, odds ratios, standardised effect sizes) and imbalances in prognostic factors.

Search methods for identification of studies

For this update we conducted our search twice: once in July 2006 and once in September 2009.

July 2006 search

We searched the Cochrane Methodology Register (CMR) (see [Appendix 1](#) for short description) (2006, Issue 3). We retrieved all related articles in PubMed to 31 indexed records of the 32 included studies in [Kunz 2002](#) on 11 July 2005. In addition, we searched the ISI Science Citation Index for articles citing the 32 included studies in [Kunz 2002](#). We also identified studies using bibliographies, handsearching, personal communication with methodologists and the reference lists of relevant articles. Attempts to develop a MEDLINE search strategy were not productive, not least because of the absence of suitable search terms related to methodology of trials/studies. Methodological studies are not easily identifiable in electronic bibliographic databases. An investigation is underway within the Methodology Review Group (CRMG) to assess whether the MESH terms, text words, authors and journals for articles already in the bibliographic database could be used to develop a search strategy for retrospective and prospective searching of MEDLINE ([CRMG Module](#)).

We used the following CMR strategy, using only CMR codes:

CMR: Systematic reviews - comparisons
 CMR: Meta-analysis - heterogeneity
 CMR: Meta-analysis - group allocation
 CMR: Bias in trials - general
 CMR: Bias in trials - random allocation
 CMR: Bias in trials - relationship to trial quality
 CMR: Study design
 CMR: Non-randomised studies - bias

September 2009 search

We searched the CMR (2009, Issue 3) and CENTRAL. We retrieved all related articles in PubMed to eight indexed records of the nine studies identified by the July 2006 search ([Balk 2002](#); [Clifford 2002](#); [Egger 2003](#); [Hedrick 1989](#); [Heinsman 1996](#); [Ioannidis 2001](#); [Kjaergard 2001](#); [Linde 1999](#); [Shang 2005](#)) on 2 September 2009. In addition, we searched the ISI Science Citation Index for articles citing the nine identified studies.

We used the following CMR strategy, which is based on the CMR codes used to index the studies already included in this review:

#1 (CMR):kw and "random allocation":kw
 #2 (CMR):kw and "group allocation":kw
 #3 (CMR):kw and "trial quality":kw

- #4 (CMR):kw and "study design":kw
 #5 (CMR):kw and ("meta-analysis" or "meta analysis" or metaanalysis):kw and ("non-randomized" or "non randomised" or nonrandomized or "non-randomised" or "non randomised" or nonrandomised):kw
 #6 (CMR):kw and ("meta-analysis" or "meta analysis" or metaanalysis):kw and (heterogeneity):kw
 #7 (CMR):kw and ("systematic-review" or "systematic-reviews" or "systematic review" or "systematic reviews"):kw and (comparison or comparisons):kw
 #8 (CMR):kw and (checklist*):kw and (trial*):kw
 #9 (CMR):kw and (checklist*):kw and ("non-trial" or "non trial" or "non-trials" or "non trials" or nontrial*):kw
 #10 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9)

The search strategy for previous versions/updates of this review can be found in [Appendix 2](#).

Data collection and analysis

Two of RK, GEV, AT, EAA, JOJ, HJS, AN, MB, ADO or Elizabeth Paulsen screened and assessed references for potential relevance independently.

Two of RK, GEV, AT, EAA, JOJ, HJS, AN, MB, ADO or Elizabeth Paulsen retrieved and assessed potentially relevant articles for inclusion independently. Disagreements were resolved by discussion or a third person.

We used the following criteria to appraise the risk of bias of included studies:

- Were explicit criteria used to select the studies?
- Did two or more investigators agree regarding the selection of studies?
- Was there a consecutive or complete sample of studies?
- Were other methodological differences such as double-blinding and complete follow up of included studies controlled for?
- Were clinical differences in the participants and interventions in the included studies controlled for?
- Were similar outcome measures used in the included studies?

We summarised the overall risk of bias of each study as: low risk of bias, unclear risk of bias or high risk of bias.

For each study, two of us independently extracted information about the sample of trials, the comparisons that were made, the type of analysis and the results. We recorded the reported relationship between randomisation or concealment of allocation on one hand, and estimates of effect on the other hand. If possible, we converted the reported relationship to the relative over- or underestimation of the relative risk reduction using the results of randomised trials, and randomised trials with concealed allocation, respectively, as the reference. We prepared tables for each type of comparison to facilitate a qualitative description of the extent to which the included studies yielded similar results. For each type of comparison, we prepared forest plot-like graphs for continuous estimates of effects and ratios of odds ratios if at least two studies reported relevant data. We have not pooled the results of the included studies in a meta-analysis because we expect heterogeneity when it comes to completeness of data, and modes of analysis among the included studies. Furthermore, we expected that the results would vary according to intervention, condition

and outcome, giving a high risk of confounding by intervention and condition.

RESULTS

Description of studies

Results of the search

In July 2006, we screened 5284 citations from the Cochrane Methodology Register plus references found using related articles in PubMed, 4671 citations in SciSearch that cited articles included in the previous version of this review, and reference lists. Seven studies met our inclusion criteria.

In September 2009, we screened a further 2014 citations from the Cochrane Methodology Register plus references found using related articles in PubMed and 475 citations in SciSearch that cited articles included based on the July 2006 search. We classified 44 abstracts and references as possibly eligible and listed these as awaiting classification. The 44 citations in the awaiting classification section of the review may alter the conclusions of the review once assessed.

Included studies

We found 18 studies that met our inclusion criteria, with a total of 1714 healthcare trials plus 74 meta-analyses with an unreported number of included trials. We have included eight new studies ([Balk 2002](#); [Clifford 2002](#); [Egger 2003](#); [Hedrick 1989](#); [Heinsman 1996](#); [Kjaergard 2001](#); [Linde 1999](#); [Shang 2005](#)) since the last version of this review ([Kunz 2002](#)).

The 18 included studies were as follows:

- one study comparing randomised and non-randomised studies of the same intervention and condition (including 31 trials);
- one study of randomised versus non-randomised studies across different interventions for the same condition (including 100 trials);
- eight studies (with nine comparisons) of randomised versus non-randomised studies across different interventions and conditions (including 486 studies and 74 meta-analyses with an unreported number of included studies);
- no study of controlled trials with adequate concealed versus inadequate/unclear concealment of allocation of the same intervention and condition;
- one study of controlled trials with adequate concealed versus inadequate/unclear concealment of allocation across different interventions for the same condition (including 102 trials); and
- eight studies (with 12 comparisons) of controlled trials with adequate versus with inadequate/unclear concealment of allocation across different interventions and conditions (including 1482 trials).

All 18 included studies assessed impact on estimate of effect. Only one study assessed the impact on imbalances in prognostic factors.

Excluded studies

Since the last version of this review ([Kunz 2002](#)) we have now excluded 22 studies previously included in this review. The previously included studies are now excluded due to the following reasons:

- comparison of interest was incidental to main aim of study (the methodological comparison was not part of the objectives of the study): 10 studies (Aronson 1996, Chalmers 1977; Forgie 1998, Mullen 1997; Ottenbacher 1993; Pyorala 1995; Reimold 1992; RMIT Group 1994; Watson 1994; Wortman 1983);
- randomised trials versus historically controlled studies: three studies (Bhansali 1996; Diehl 1986; Sacks 1982);
- randomised trials versus observational studies: two studies (Benson 2000; Guyatt 2000); and
- high versus low quality: seven studies (Emerson 1990; Imperiale 1990; Khan 1996; Nurmohamed 1992; Ortiz 1998; Potter 1998; Stanton 1997).

Risk of bias in included studies

See Table 1 for judgement of risk of bias. We judged all 18 included studies as being at high risk of bias.

Effect of methods

Effects of randomisation

Randomised versus non-randomised studies of the same intervention and condition

One study including a total of 31 randomised and non-randomised studies of the same intervention and condition are summarised in

Analysis 1.1. Carroll 1996 found larger estimates of effect in non-randomised compared to randomised studies. The study did not assess the impact on imbalances in prognostic factors.

Randomised versus non-randomised studies across different interventions for the same condition

We identified no studies that could be included in this comparison.

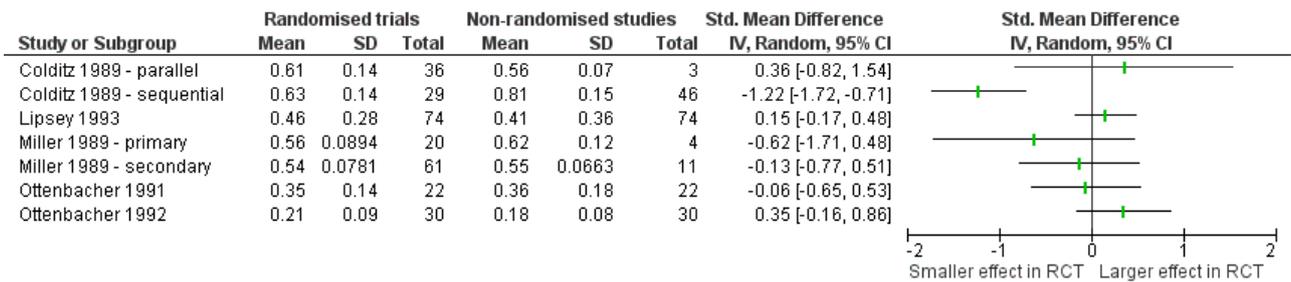
Randomised versus non-randomised studies across different interventions for the same condition

One study including a total of 100 trials compared randomised and non-randomised studies across different interventions for the same condition (Analysis 3.1). Shadish 1996 found that non-randomised trials had on average effect sizes that were 87% smaller than randomised trials of marital and family psychotherapy.

Randomised versus non-randomised studies across different interventions and conditions

Eight studies (with nine comparisons) including a total of 486 trials and 74 meta-analyses with an unreported number of included studies compared randomised and non-randomised studies across different interventions and conditions (Analysis 4.1 and Figure 1).

Figure 1. Studies of randomised trials compared with non-randomised trials across different interventions and conditions - continuous estimates of effect



One study of 89 placebo-controlled trials of homeopathy found that trials without an explicit statement of randomisation had larger effect estimates than trials with an explicit statement of randomisation (Linde 1999). One study found smaller effects in non-randomised studies for one type of intervention (drug use prevention) and similar effects for another (presurgical interventions) (Heinsman 1996). Two other studies found no differences across studies of a variety of surgical and occupational therapy interventions (Miller 1989; Ottenbacher 1992). The four other studies had inconclusive results (conflicting results from different variables or modes of analyses) (Colditz 1989; Hedrick 1989; Lipsey 1993; Ottenbacher 1991).

None of the studies assessed the impact on imbalances in prognostic factors.

Effects of allocation concealment

Adequate versus inadequate/unclear concealment of allocation within controlled trials of the same intervention and condition

We identified no studies that could be included in this comparison.

Adequate versus inadequate/unclear concealment of allocation within controlled trials of the same intervention for different conditions

We identified no studies that could be included in this comparison.

Adequate versus inadequate/unclear concealment of allocation within controlled trials across different interventions for the same condition

One study including a total of 102 trials compared adequate and inadequate concealment of allocation within controlled trials across different interventions for the same condition (Analysis 7.1). Chalmers 1983 found that controlled trials with inadequate concealment of allocation had a larger effect than adequately concealed randomised trials on treatment of acute myocardial infarction.

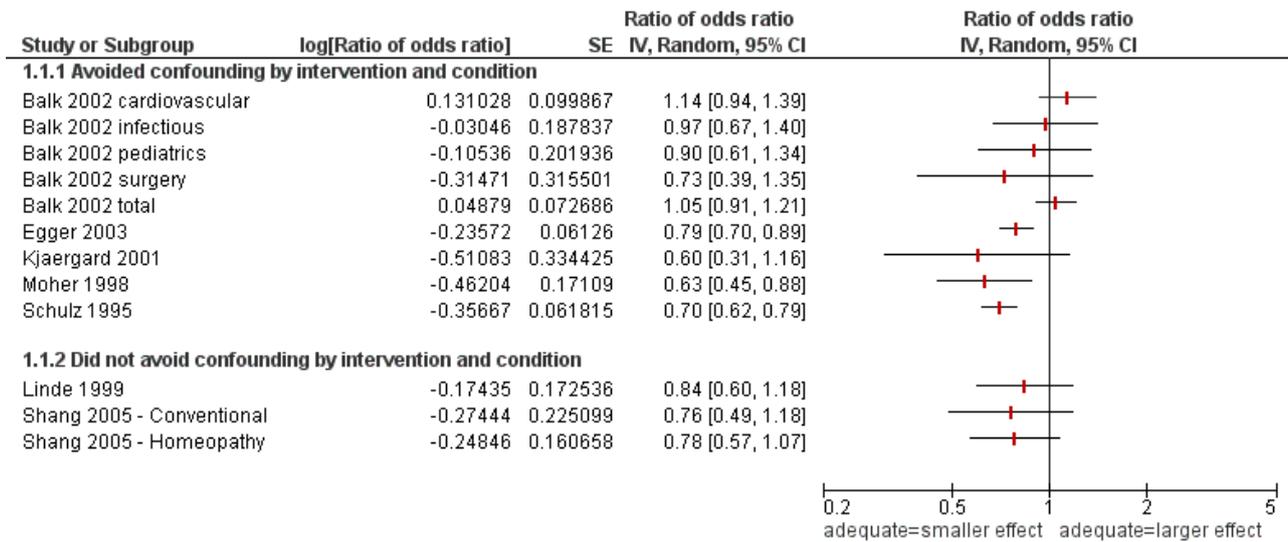
In one study (Chalmers 1983), non-concealed trials showed a larger proportion prognostic factors with imbalance between groups (7%) than randomised trials (3.5%).

Adequate versus inadequate/unclear concealment of allocation within controlled trials across different interventions and conditions

Eight studies (with 12 comparisons) with a total of 1482 trials (74 duplicates included in both [Linde 1999](#) and [Shang 2005](#)

are only counted once) included comparisons between adequate and inadequate concealment of allocation within controlled trials across different interventions and conditions ([Analysis 8.1](#) and [Figure 2](#)). Five of the studies included in this comparison avoided confounding by intervention and condition ([Balk 2002](#); [Kjaergard 2001](#); [Moher 1998](#); [Schulz 1995](#); [Egger 2003](#)).

Figure 2. Studies of controlled trials with adequate concealment of allocation compared with inadequate/unclear concealment of allocation across different interventions and conditions - ratio of odds ratios



In [Schulz 1995](#) it was found that estimates of treatment effect were 41% larger in controlled trials with inadequate allocation concealment compared with controlled trials with adequate allocation concealment and 30% larger in controlled trials with unclear adequacy of allocation concealment. Moher and colleagues ([Moher 1998](#)) found that treatment effects were 37% larger in trials with inadequate concealment compared to trials with adequate concealment for 127 trials in 11 meta-analyses of diverse clinical interventions. Egger and colleagues ([Egger 2003](#)) found that trials with inadequate or unclear allocation concealment had effect estimates that were 21% larger in 39 meta-analyses including 314 trials of various clinical interventions. Kjaergard and colleagues ([Kjaergard 2001](#)) also found that trials with inadequate concealment had effect estimates that were on average 40% larger (not statistically significant) in 14 meta-analyses including 190 trials for eight different interventions.

The other four studies found similar estimates of effect in trials with adequate and inadequate/unclear allocation concealment ([Balk 2002](#); [Clifford 2002](#); [Linde 1999](#); [Shang 2005](#)). Balk found similar treatment effects in four comparisons of 276 trials from 26 meta-analysis of cardiovascular disease (93 trials), infectious disease (56 trials), paediatrics (60 trials) and surgery (67 trials). Clifford also found similar estimates of effect in 100 trials from various clinical areas. Linde found similar estimates of effect in 89 homeopathic trials. Shang found similar treatment effect in two comparisons of 220 trials of homeopathy (110 trials) and conventional medicine (110 trials).

None of the studies assessed the impact on imbalances in prognostic factors.

DISCUSSION

Summary of main results

The results of randomised and non-randomised studies sometimes differed. In some instances non-randomised studies yielded larger estimates of effect and in other instances randomised trials yielded larger estimates of effect. The results of controlled trials with adequate and inadequate/unclear concealment of allocation sometimes differed. When differences occurred, most often trials with inadequate or unclear allocation concealment yielded larger estimates of effects relative to controlled trials with adequate allocation concealment. However, it is not generally possible to predict the magnitude, or even the direction, of possible selection biases and consequent distortions of treatment effects from studies with non-random allocation or controlled trials with inadequate or unclear allocation concealment.

Overall completeness and applicability of evidence

[Schulz 1995](#) provides strong support for the conclusion that clinical trials that lack adequately concealed allocation produce estimates of effect that are, on average, 40% larger than trials with adequately concealed allocation. Moher ([Moher 1998](#)) has replicated those findings in a study based on controlled trials using similar methods. Wood and colleagues ([Wood 2008](#)) based an analysis on a subset of the trials included in [Schulz 1995](#), [Kjaergard 2001](#) and [Egger 2003](#). They found that the average bias associated with lack of adequate allocation concealment was less for trials with objectively assessed outcomes than for trials with subjectively assessed outcomes, and less for trials with all-cause mortality as the outcome than for trials with other outcomes. Wood and colleagues concluded that the bias in part may result from an association with subsequent flaws in the

conduct of the trial rather than from selection biases. They further noted that the effect of allocation concealment remained even after adjustment for blinding, suggesting that it is a marker for other bias-reducing strategies, beyond blinding.

The studies by Schulz (Schulz 1995) and Moher (Moher 1998) also demonstrate the potential contribution that systematic reviews, and particularly the Cochrane Database of Systematic Reviews, can make towards developing an empirical basis for methodological decisions in evaluations of healthcare interventions. Currently this empirical basis is lacking for many methodological decisions, and many methodological debates rely more on logic or rhetoric than evidence. Analyses such as the one undertaken by Schulz 1995 and Moher 1998, in which methodological comparisons are made adjusting for interventions are likely to yield more reliable results than comparisons that are made across different interventions without adjustment for interventions. Comparisons made across different interventions will often have a great deal of clinical and methodological heterogeneity and thus, not surprisingly, tend to have inconclusive results. Nonetheless methodological comparisons would likely yield even more reliable results when made among trials of the same intervention and the same condition. We found only one study of randomised trials versus non-randomised trials of the same intervention and condition. In Carroll 1996, a small study including 31 trials, the results are only analysed using vote counting with no quantitative estimates of effect, making it difficult to generalise from the results.

Assumptions

We have used randomised trials and controlled trials with adequate concealment of allocation as the reference in the comparisons we have made. Implicit in this is an assumption that differences in results are best explained by bias and that the reference randomised trials are less likely to be biased. This assumption is, to a limited extent, supported by findings of larger imbalances in prognostic factors among historical controls compared to randomised controls (Diehl 1986; Sacks 1982) and among non-randomised trials with inadequate/unclear concealment of allocation compared with randomised trials with adequate concealment of allocation (Chalmers 1983). However, it is possible that randomised trials sometimes underestimate the effects of an intervention in routine practice by forcing healthcare professionals and patients to acknowledge their uncertainty and, thereby, increasing the strength of placebo effects (Black 1996; Chalmers 1997; Kleijnen 1997). In addition, the use of intention-to-treat (ITT) analyses in randomised controlled trials will probably also reduce the effect estimate compared to non-randomised studies which will almost always use a 'per-protocol' type analysis. This is because if a treatment is beneficial its benefit will appear less in an ITT analysis of a treatment versus control trial due to some treatment patients not using it, and some control patients using it. It is also possible that publication bias can partly explain some of the differences in results observed in studies such as the one by Sacks and colleagues (Sacks 1982). This would be the case if randomised trials were more likely to be published regardless of the effect size than non-randomised studies with historical controls. We are not aware of any evidence that supports this hypothesis and the available evidence shows consistently that randomised trials, like other research, are also more likely to be published if they have 'significant' results (Dickersin 1997; Hopewell 2001; Hopewell 2006; Song 2000).

Possible explanations for discrepancies

There are a number of other possible explanations for discrepancies between estimates of effect derived from randomised and non-randomised trials. For example, it can be argued that estimates of treatment effect might be larger in randomised trials if the care provided in the context of randomised trials is better than that in routine practice, assuming this is the case for the treatment group and not the control group. Similarly, strict eligibility criteria might select people with a higher potential to benefit from a treatment, resulting in larger estimates of effect in randomised trials than non-randomised trials with less strict eligibility criteria. If patients with a poorer prognosis were more likely to be allocated to the treatment group in non-randomised trials for some reason, this would also result in larger estimates of effect in randomised trials. Conversely, if patients with a poorer prognosis are more likely to be allocated to the control group in non-randomised trials, as one study in this review showed, this would result in larger estimates of effect in non-randomised studies.

Heterogeneity

One could argue that heterogeneity is bound to be present in this review, as one would not expect to find exactly the same estimates of effect for different but somewhat similar questions in a diverse range of clinical areas. A visual inspection of Figure 2 suggest that results from methodology studies that did not avoid confounding by intervention and condition are more homogeneous than results from methodology studies that avoided confounding by intervention and condition. This suggests that a meaningful grand mean effect does not exist, but that the effect of allocation concealment varies with clinical area, intervention and outcomes examined.

The primary included reviews span several decades and thus the primary studies looked at will span an even longer time period. Given the advances in the methodology of conducting trials and technology, and also in analytic approaches, one would expect that this considerable time span would introduce some sort of heterogeneity in the data. It was, however, outside the scope of this review to investigate this further, as it would constitute a methodology review in itself.

In a systematic review or overview of reviews like this we heavily depend on the studies included in the original reviews when arriving at our conclusion. Given that several of the included reviews span a variety of clinical areas, and that overlap in clinical areas examined exists between reviews included in the same comparison, there is a risk that some primary studies are included in more than one review. Thus there might be some degree of multiple counting of effects.

The scope of this review

In this version of the review we have not included comparisons between randomised trials and cohort studies (Guyatt 2000), case-control studies (Stieb 1990) or 'outcomes studies' (evaluations of effectiveness using large administrative or clinical databases) (US Office HTA 1994). Observational studies often provide valuable information complementary to the results of controlled trials. For example, case-control studies may be the best available study design for evaluating rare adverse effects, and large database studies may provide important information about the extent

to which effects that are expected based on randomised trials are achieved in routine practice. It is an important issue also to consider the possible discrepancies between results from randomised controlled trials and results from observational studies (such as cohort studies, case-control studies etc.). However, in our opinion these comparisons should be treated separately from the comparison between different types of randomised trials, as the biases that would be addressed are different between the two comparisons. As far as we know a methodology review comparing randomised controlled studies and observational studies is not yet available.

A concluding remark

As Cochrane stated: "The RCT is a very beautiful technique, of wide applicability, but as with everything else there are snags" (Cochrane 1972). Those making decisions on the basis of randomised trials need to be cautious of small trials, even when participants are properly randomised, and systematic reviews of small randomised trials, both because of chance effects and the risk of biased reporting (Counsell 1994; Egger 1997). It is also, of course, possible to introduce bias into a randomised trial despite adequate allocation concealment (Guyatt 2002; Schulz 1995). Finally, even when the risk of error due to either bias or chance is small, judgements must be made about the applicability of the results to individual patients (Dans 2002; Rothwell 2005) and about the relative value of the probable benefits, harms and costs.

Quality of the evidence

The quality of the data is limited by the quality of the studies that we have reviewed. All of the 18 included studies have one or more methodological limitations. All studies were judged as being at high risk of bias. In most of the studies (13 of 18) no control for clinical differences in participants and interventions was made as part of the statistical analyses. Thus results from these studies could be at risk of bias from confounding by intervention and condition.

The included studies are inconsistent when it comes to the bias associated with lack of randomisation and adequate allocation concealment. In four comparisons comparing randomised and non-randomised trials inconclusive results were found; in three comparisons similar results were found in randomised and non-randomised trials; in two comparisons larger estimates of effect were found in non-randomised trials than in randomised trials; and in two comparisons smaller estimates of effect were found in non-randomised trials than in randomised trials.

For the comparison of adequate versus inadequate/unclear allocation concealment, four comparisons found similar estimates of effect, five comparisons yielded larger estimates of effect, and no comparisons yielded smaller estimates of effect in trials with inadequate/unclear compared with trials with adequate allocation concealment.

Potential biases in the review process

It has been difficult to develop efficient search strategies for locating empirical methodological studies eligible for inclusion in this review. However, we believe it is unlikely that there are many published eligible methodological studies that we have not identified.

It is possible that there is publication bias or that we have identified a non-representative sample of published studies. This is due to the inefficiency of the search strategies that we can use and a possible bias amongst the people we have contacted. However, two other systematic reviews using different search strategies and methods (McKee 1999; Reeves 1998) did not identify any studies that we have not included. We have also used citation searching, checked related articles in PubMed, checked the reference lists of relevant articles and used personal contacts to help ensure that we have included as many relevant studies as possible. We have not received any comments on previous versions of this review identifying studies that we had not identified. However, we did, due to a better search, identify two studies with this update that were published prior to the previous version (Heinsman 1996; Linde 1999).

The main caveat regarding the results of this review concerns the layers of filters and processing between the primary evidence and the conclusion. Our body of evidence consists of cohorts of trials, systematic reviews and meta-analyses conducted by others, who potentially had other objectives. The results of the studies we have included are all at risk of bias from searches, inclusion criteria, analyses and reporting. On top of this comes the biases potentially introduced through our search, inclusion and data extraction. Twelve of the 18 studies included in this review are based on convenience samples and may not necessarily constitute the complete picture of the body of evidence. This may introduce bias and heterogeneity. Likewise, differences when it comes to statistical analysis (methods and outcomes analysed) in the included studies can also introduce heterogeneity and bias.

AUTHORS' CONCLUSIONS

Implication for methodological research

Additional well-designed studies comparing randomised and non-randomised trials, in particular, and adequately and inadequately concealed allocation in controlled trials of the same intervention and condition could help strengthen inferences about the importance of randomisation and allocation concealment or potentially modify the above inferences. Further comparisons across different interventions and conditions, and comparisons of trials based on quality scales are of questionable value (Jüni 1999; Jüni 2001). A methodology review of comparisons between randomised trials and observational studies, including cohort studies, case-control studies and 'outcomes research' (evaluations using large administrative or clinical databases) is needed. The scope of such a methodology review is so large that it is best done in a separate review, and not combined with the review at hand. To investigate the role of varying baseline risk, heterogeneity or study quality adequately, individual trial analysis might be required.

ACKNOWLEDGEMENTS

We are grateful to Marit Johansen for conducting the searches for this update; to Dave Sackett and Iain Chalmers for encouragement and advice; to Kay Dickersin, Annie Britton and other colleagues who previously generously provided us with their bibliographies on research methodology; to Elizabeth Paulsen (EP) who helped us with screening of references and assessment for inclusion; and to the investigators who have conducted the studies we have reviewed. We are also grateful to the referees and editors who have commented on earlier drafts of this review and helped to improve it.

However, we cannot hold them responsible for our errors. We thank the Department of Health, England for support for this update.

REFERENCES

References to studies included in this review

Balk 2002 {published data only}

Balk EM, Bonis PAL, Moskowitz H, Schmid CH, Ioannidis JPA, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analysis of randomized controlled trials. *JAMA* 2002;**287**:2973-82.

Carroll 1996 {published data only}

Carroll D, Tramer M, McQuay H, Nye B, Moore A. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain.. *British Journal of Anaesthesia* 1996;**77**:798-803.

Chalmers 1983 {published data only}

Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials.. *New England Journal of Medicine* 1983;**309**:1358-61.

Clifford 2002 {published data only}

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**:18.

Colditz 1989 {published data only}

Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Statistics in Medicine* 1989;**8**:441-54.

Egger 2003 {published data only}

Egger M, Jüni P, Bartlett C, Hohenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical Study. *Health Technology Assessment* 2003;**7**(1):1-76.

Hedrick 1989 {published data only}

Hedrick SC, Koepsell TD, Inui T. Meta-analysis on home-care effects on mortality and nursing-home placement. *Medical Care* 1989;**27**:1015-26.

Heinsman 1996 {published data only}

Heinsman DT, Shadish WR. Assignment methods in experimentation: when do non-randomised experiments approximate answers from randomised experiments?. *Psychological Methods* 1996;**1**:154-69.

Kjaergard 2001 {published data only}

Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analysis. *Annals of Internal Medicine* 2001;**135**:982-9.

Linde 1999 {published data only}

Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, et al. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo controlled trials. *Lancet* 1997;**350**:834-43.

* Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *Journal of Clinical Epidemiology* 1999;**52**:631-6.

Lipsey 1993 {published data only}

Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist* 1993;**48**(12):1181-209.

Miller 1989 {published data only}

Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II Surgical. *Statistics in Medicine* 1989;**8**:455-66.

Moher 1998 {published data only}

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

Ottensbacher 1991 {published data only}

Ottensbacher KJ. Epistemology and experimentation: an examination of quality factors in research design. *American Journal of Occupational Therapy* 1991;**45**(10):917-23.

Ottensbacher 1992 {published data only}

Ottensbacher K. Impact of random assignment on study outcome: an empirical examination. *Controlled Clinical Trials* 1992;**13**:50-61.

Schulz 1995 {published data only}

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408-12.

Shadish 1996 {published data only}

Shadish WR, Ragsdale K. Random versus nonrandom assignment in controlled experiments: do you get the same answer?. *Journal of Consulting and Clinical Psychology* 1996;**64**(6):1290-305.

Shang 2005 {published data only}

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

References to studies excluded from this review

Aronson 1996 {published data only}

Aronson R, Offman HJ, Joffe RT, Naylor CD. Triiodothyronine augmentation in the treatment of refractory depression. A meta-analysis. *Archives of General Psychiatry* 1996;**53**(9):842-8.

Benson 2000 {published data only}

Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine* 2000;**342**:1878-86.

Bhansali 1996 {published data only}

Bhansali MS, Vaidya JS, Bhatt RG, Patil PK, Badwe RA, Desai PB. Chemotherapy for carcinoma of the esophagus: a comparison of evidence from meta-analyses of randomized trials and of historical control studies. *Annals of Oncology* 1996;**7**(4):355-9.

Chalmers 1977 {published data only}

Chalmers TC, Matta RJ, Smith H Jr, Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *New England Journal of Medicine* 1997;**297**:1091-6.

Diehl 1986 {published data only}

Diehl LF, Perry DJ. A comparison of randomized concurrent control groups with matched historical control groups: are historical controls valid?. *Journal of Clinical Oncology* 1986;**4**:1114-20.

Emerson 1990 {published data only}

Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clinical Trials* 1990;**11**:339-52.

Forge 1998 {published data only}

Forge MA, Wells PS, Laupacis A, Fergusson D. Preoperative autologous donation decreases allogeneic transfusion but increases exposure to all red blood cell transfusion: results of a meta-analysis. International Study of Perioperative Transfusion (ISPOT) Investigators. *Archives of Internal Medicine* 1998;**158**(6):610-6.

Gilbert 1977 {published data only}

Gilbert JP, McPeck B, Mosteller F. Statistics and ethics in surgery and anesthesia. *Science* 1977;**198**(4318):684-9.

Guyatt 2000 {published data only}

Guyatt GH, DiCenso A, Farewell V, Willan A, Griffith L. Randomized trials versus observational studies in adolescent pregnancy prevention. *Journal of Clinical Epidemiology* 2000;**53**(2):167-74.

Hovell 1982 {published data only}

Hovell MF. The experimental evidence for weight-loss treatment of essential hypertension: a critical review. *American Journal of Public Health* 1982;**72**(4):359-68.

Hutchinson 1999 {published data only}

Hutchinson BG, Oxman AD, Shannon HS, Lloyd S, Altmayer CA, Thomas K. Clinical effectiveness of pneumococcal vaccine. Meta-analysis. *Canadian Family Physician* 1999;**45**:2381-93.

Imperiale 1990 {published data only}

Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta analysis of the randomized trials. *Annals of Internal Medicine* 1990;**113**:299-307.

Ioannidis 2001 {published data only}

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;**286**:821-30.

Kasiske 1993 {published data only}

Kasiske BL, Heim-Duthoy K, Ma JZ. Elective cyclosporine withdrawal after renal transplantation. A meta-analysis. *JAMA* 1993;**269**:395-400.

Kasiske 1998 {published data only}

Kasiske BL, Lakatua JD, Ma JZ, Louis TA. A meta-analysis of the effects of dietary protein restriction on the rate of decline in renal function. *American Journal of Kidney Disease* 1998;**31**:954-61.

Kerlikowske 1995 {published data only}

Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA* 1995;**273**(2):149-54.

Khan 1996 {published data only}

Khan KS, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. *Archives of Internal Medicine* 1996;**156**:661-6.

Koes 1994 {published data only}

Koes BW, van Tulder MW, van der Windt WM, Bouter LM. The efficacy of back schools: a review of randomized clinical trials. *Journal of Clinical Epidemiology* 1994;**47**(8):851-62.

Kownacki 1999 {published data only}

Kownacki RJ, Shadish WR. Does Alcoholics Anonymous work? The results from a meta-analysis of controlled experiments. *Substance Use & Misuse* 1999;**34**:1897-916.

MacArthur 1995 {published data only}

Macarthur C, Foran PJ, Bailar JC 3rd. Qualitative assessment of studies included in a meta-analysis: DES and the risk of pregnancy loss. *Journal of Clinical Epidemiology* 1995;**48**(6):739-47.

Mehta 1999 {published data only}

Mehta RH, Bates ER. Coronary stent implantation in acute myocardial infarction. *American Heart Journal* 1999;**137**:603-11.

Moher 1999 {published data only}

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analysis. *Health Technology Assessment* 1999;**3**(12):100p.

Morrison 1997 {published data only}

Morrison B, Lilford RJ, Earnst E. Methodological rigor and results of clinical trials of homeopathic remedies. *Perfusion* 2000;**13**:132-8.

Mullen 1997 {published data only}

Mullen PD, Simons-Morton DG, Ramirez G, Frankowski RF, Green LW, Mains DA. A meta-analysis of trials evaluating patient

education and counseling for three groups of preventive health behaviors. *Patient Education and Counseling* 1997;**32**(3):157-73.

Nurmohamed 1992 {published data only}

Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low molecular weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;**340**:152-6.

Ortiz 1998 {published data only}

Ortiz Z, Shea B, Suarez Almazor ME, Moher D, Wells GA, Tugwell P. The efficacy of folic acid and folinic acid in reducing methotrexate gastrointestinal toxicity in rheumatoid arthritis. A meta-analysis of randomized controlled trials. *Journal of Rheumatology* 1998;**25**:36-43.

Ottenbacher 1993 {published data only}

Ottenbacher KJ, Jannell S. The results of clinical trials in stroke rehabilitation research. *Archives of Neurology* 1993;**50**(1):37-44.

Pagnin 2004 {published data only}

Pagnin D, Queiroz V, Pini S, Cassano GB. Efficacy of ECT in depression: a meta-analytic review. *Journal of ECT* 2004;**20**:13-20.

Potter 1998 {published data only}

Potter J, Langhorne P, Roberts M. Routine protein energy supplementation in adults: systematic review. *BMJ* 1998;**317**:495-501.

Pyorala 1995 {published data only}

Pyorala S, Huttunen NP, Uhari M. A review and meta-analysis of hormonal treatment of cryptorchidism. *Journal of Clinical Endocrinology & Metabolism* 1995;**80**:2795-9.

RMIT Group 1994 {published data only}

Recurrent Miscarriage Immunotherapy Trialists Group. Worldwide collaborative observational study and meta analysis on allogenic leukocyte immunotherapy for recurrent spontaneous abortion. *American Journal of Reproductive Immunology* 1994;**32**:55-72.

Rozenberg 1999 {published data only}

Rozenberg S, Dubourg G, Khalifa P, Paolozzi L, Maheu E, Ravaud P. Efficacy of epidural steroids in low back pain and sciatica. A critical appraisal by a French Task Force of randomized trials. Critical Analysis Group of the French Society for Rheumatology. *Revue de Rhumatisme (English edition)* 1999;**66**(2):79-85.

Sacks 1982 {published data only}

Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *American Journal of Medicine* 1982;**72**:233-40.

Shadish 1997 {published data only}

Shadish WR, Heinsman DT. Experiments versus quasi-experiments: do they yield the same answer?. In: Bukoski WJ editor(s). Meta-analysis of Drug Abuse Prevention Programs. Washington DC: US Department of Health and Human Services, National Institutes of Health, 1997:147-64.

Shadish 2001 {published data only}

Shadish WR, Glaser RR. Differences between randomised and nonrandomised experiments using meta-analysis: a methodological note. 2001.

Stanton 1997 {published data only}

Stanton MD, Shadish WR. Outcome, attrition and family-couples treatment for drug abuse. *Psychological Bulletin* 1997;**122**:170-91.

Stieb 1990 {published data only}

Stieb D, Frayha HH, Oxman AD, Shannon HS, Hutchison BG, Crombie F. The effectiveness and usefulness of Haemophilus influenzae type B vaccines: a systematic overview (meta-analysis). *Canadian Medical Association Journal* 1990;**142**:719-32.

Watson 1994 {published data only}

Watson A, Vandekerckhove P, Lilford R, Vail A, Brosens I, Hughes E. A meta-analysis of the therapeutic role of oil soluble contrast media at hysterosalpingography: a surprising result?. *Fertility and Sterility* 1994;**61**:470-7.

Weisburd 2001 {published data only}

Weisburd D, Lum CM, Petrosino A. Does research design affect study outcomes in criminal justice. *Annals of the American Academy of Political and Social Science* 2001;**578**:50-70.

Wortman 1983 {published data only}

Wortman P, Yeaton WH. Synthesis of results in trials of coronary artery bypass graft surgery. In: Light R editor(s). Evaluation Studies Review Annual. Vol. 8, Beverley Hills, CA: Sage Publications, 1983:536-57.

References to studies awaiting assessment

Abraham 2010 {published data only}

Abraham NS, Byrne CJ, Young JM, Solomon MJ. Meta-analysis of well-designed nonrandomized comparative studies of surgical procedures is as good as randomized controlled trials. *Journal of Clinical Epidemiology* 2010;**63**(3):238-45.

Bellavite 2006 {published data only}

Bellavite P, Ortolani R, Pontarollo F, Piasere V, Benato G, Conforti A. Immunology and homeopathy. 4. Clinical studies - Part 1. *Evidence-Based Complementary and Alternative Medicine* 2006;**3**:293-301.

Bhandari 2004 {published data only}

Bhandari M, Tornetta P, Ellis T, Audige L, Sprague S, Kuo JC, et al. Hierarchy of evidence: differences in results between non-randomized studies and randomized trials in patients with femoral neck fractures. *Archives of Orthopaedic and Trauma Surgery* 2004;**124**:10-6.

Bhogal 2005 {published data only}

Bhogal SK, Foley N, Teasell R, Speechley M. Methodological differences influencing treatment effect [abstract]. *American Journal of Epidemiology* 2005;**161**:S67.

Britton 1998 {published data only}

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technology Assessments* 1998;**2**:No.13.

Cipriani 2006 {published data only}

Cipriani A, Malvini L, Furukawa T, Barbui C. Does the quality of antidepressant trials affect outcome estimates? [abstract]. XIV Cochrane Colloquium; 2006 October 23-26; Dublin, Ireland 2006; Vol. 169.

Cipriani 2007 {published data only}

Cipriani A, Malvini L, Furukawa TA, Barbui C. Relationship between quality of reports of antidepressant randomized controlled trials and treatment estimates: systematic review, meta-analysis, and meta-regression analysis. *Journal of Clinical Psychopharmacology* 2007;**27**:352-6.

Concato 2000 {published data only}

Concato J, Shah N, Horwitz RJ. Randomized controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine* 2000;**342**:1887-92.

Fenwick 2008 {published data only}

Fenwick J, Needleman I, Moles D. Effect of allocation concealment and examiner masking on magnitude of clinical outcomes. Poster presentation at the 16th Cochrane Colloquium: Evidence in the era of globalisation; 2008 Oct 3-7; Freiburg, Germany [abstract]. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen* 2008;**102**:59.

Fenwick 2008a {published data only}

Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. *Journal of Clinical Periodontology* 2008;**35**:775-82.

Fernandez-de-las-Penas 06 {published data only}

Fernandez-de-las-Penas C, Alonso-Blanco C, San Roman J, Miangolarra-Page JC. Methodological quality of randomized controlled trials of spinal manipulation and mobilization in tension-type headache, migraine, and cervicogenic headache. *Journal of Orthopaedic & Sports Physical Therapy* 2006;**36**:160-9.

Ferriter 2005 {published data only}

Ferriter M, Huband N. Does the non-randomized controlled study have a place in the systematic review? A pilot study. *Criminal Behaviour and Mental Health* 2005;**15**:111-20.

Foley 2005 {published data only}

Foley N, Speechley M, Salter K, Bhogal S, Jutai J, Teasell R. Concealed allocation: an under-reported and misunderstood component of trial methodology in stroke rehabilitation [abstract]. *American Journal of Epidemiology* 2005;**161**:S67.

Furlan 2008a {published data only}

Furlan AD, Tomlinson G, Jadad AA, Bombardier C. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. *Spine* 2008;**33**:339-48.

Furlan 2008b {published data only}

Furlan AD, Tomlinson G, Jadad AA, Bombardier C. Methodological quality and homogeneity influenced agreement between randomized trials and nonrandomized studies of the same intervention for back pain. *Journal of Clinical Epidemiology* 2008;**61**:209-31.

Gluud 2008 {published data only}

Gluud LL, Thorlund K, Gluud C, Woods L, Harris R, Sterne JA. Correction: reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine* 2008;**149**:219.

Grossarth-Maticek 2006 {published data only}

Grossarth-Maticek R, Ziegler R. Randomised and non-randomised prospective controlled cohort studies in matched-pair design for the long-term therapy of breast cancer patients with a mistletoe preparation (Iscador): a re-analysis. *European Journal of Medical Research* 2006;**11**:485-95.

Hartz 2005 {published data only}

Hartz A, Bentler S, Charlton M, Lanska D, Butani Y, Soomro GM, et al. Assessing observational studies of medical treatments. *Emerging Themes in Epidemiology* 2005;**2**:8.

Henry 2001 {published data only}

Henry DM. Agreement between randomized and non-randomized studies - the effects of bias and confounding. 9th Annual Cochrane Colloquium Abstracts, October 2001 in Lyon 2001.

Jin 2003 {published data only}

Jin SL, Wang YP, Zhao PC, Guo Z, Yang JL. Analysis of randomized controlled trials/clinical controlled trials on chronic gastritis in China. *Chinese Journal of Evidence Based Medicine* 2003;**3**:22-5.

Kelly 2001 {published data only}

Kelly AK. The last word in trial quality? The impact of selection and performance bias within a series of integrated systematic reviews. 9th Annual Cochrane Colloquium Abstracts, October 2001 in Lyon 2001.

Khan 1996b {published data only}

Khan KS, Daya SD, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertility and Sterility* 1996;**65**:939-45.

Kitsios 2009 {published data only}

Kitsios G, Zintzaras E. ACE (I/D) polymorphism and response to treatment in coronary artery disease: a comprehensive database and meta-analysis involving study quality evaluation. *BMC Medical Genetics* 2009;**10**:50.

Kjaergaard 2008 {published data only}

Kjaergaard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses (vol 135, pg982, 2001). *Annals of Internal Medicine* 2008;**149**:219.

Kyriakidi 2002 {published data only}

Kyriakidi M, Ioannidis JPA. Design and quality considerations for randomized controlled trials in systemic sclerosis. *Arthritis Care & Research* 2002;**47**:73-81.

Linde 1999b {published data only}

Linde K, Melchart D, Scholz M. Should systematic reviews include non-randomized trials? Quality and results of various study types investigating the effect or acupuncture on idiopathic headaches. 7th Annual Cochrane Colloquium Abstracts, October 1999 in Rome 1999.

Manzoli 2007 {published data only}

Manzoli L, Schioppa F, Boccia A, Villari P. The efficacy of influenza vaccine for healthy children: a meta-analysis evaluating potential sources of variation in efficacy estimates including study quality. *Pediatric Infectious Disease Journal* 2007;**26**:97-106.

Papanikolaou 2006 {published data only}

Papanikolaou PN, Christidi GD, Ioannidis JPA. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *Canadian Medical Association Journal* 2006;**174**(5):635-41.

Pildal 2007 {published data only}

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

Reimold 1992 {published data only}

Reimold SC, Chalmers TC, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic atrial fibrillation: observations on the role of trial design and implications of drug related mortality. *American Heart Journal* 1992;**124**:924-32.

Rodgers 2008 {published data only}

Rodgers M, Chambers D, Woolacott N. To randomise or not to randomise: a matter of perspective? [abstract]. Fifth Annual Meeting HTAi; 2008 July 6-9; Montreal, Canada. 2008; Vol. 41.

Sacks 1983 {published data only}

Sacks HS, Chalmers TC, Smith H. Sensitivity and specificity of clinical trials. Randomized v historical controls. *Archives of Internal Medicine* 1983;**143**:753-5.

Savovic 2008 {published data only}

Savovic J, Harris RJ, The BC. The association of three bias domains with the treatment effect estimates in randomised control trials: combined analysis of meta-epidemiological studies. Oral presentation at the 16th Cochrane Colloquium: Evidence in the era of globalisation; 2008 Oct 3-7; Freiburg, Germany [abstract]. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen* 2008;**102**:29-30.

Shadish 2008 {published data only}

Shadish WR, Clark MH, Steiner PM. Can nonrandomized experiments yield accurate answers? A randomized experiment

comparing random and nonrandom assignments.. *Journal of the American Statistical Association* 2008;**103**:1334-43.

Shikata 2008 {published data only}

Shikata S, Nakayama T, Yamagishi H. Quality of surgical randomized controlled trials for acute cholecystitis: assessment based on CONSORT and additional check items. *Journal of Hepato-Biliary-Pancreatic Surgery* 2008;**15**:297-303.

Siersma 2007 {published data only}

Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Statistics in Medicine* 2007;**26**:2745-58.

Singh 2009 {published data only}

Singh JA, Murphy S, Bhandari M. Trial sample size, but not trial quality, is associated with positive study outcome. *Journal of Clinical Epidemiology* 2010;**63**(2):154-62.

Verhagen 2002 {published data only}

Verhagen AP, de Vet HCW, Vermer F, Widdershoven JWMG, de Bie RA, Kessels AG, et al. The influence of methodologic quality on the conclusion of a landmark meta-analysis on thrombolytic therapy. *International Journal of Technology Assessment in Health Care* 2002;**18**:11-23.

Verhagen 2008 {published data only}

Verhagen AP, De Vet HCW, Willemsen S, Stijnen T. A meta-regression analysis shows no impact of design characteristics on outcome in trials on tension-type headaches. *Journal of Clinical Epidemiology* 2008;**61**:813-8.

Villari 2004 {published data only}

Villari P, Manzoli L, Boccia A. Methodological quality of studies and patient age as major sources of variation in efficacy estimates of influenza vaccination in healthy adults: a meta-analysis. *Vaccine* 2004;**22**:3475-86.

Wahlbeck 2000 {published data only}

Wahlbeck K, Tuunainen A, Gilbody S, Adams CE. Influence of methodology on outcomes of randomised Clozapine trials. *Pharmacopsychiatry* 2000;**33**:54-9.

Wang 2009 {published data only}

Wang L, Li YM, Li L. Meta-analysis of randomized and controlled treatment trials for achalasia. *Digestive Diseases and Sciences* 2009;**54**(11):2303-11.

Wood 2006 {published data only}

Wood L, Egger M, Gluud LL, Schulz K, Altman D, Jüni P, et al. The association of allocation concealment and blinding with estimated treatment effect varies according to type of outcome: a combined analysis of meta-epidemiological studies [abstract]. XIV Cochrane Colloquium; 2006 October 23-26; Dublin, Ireland 2006; Vol. 52.

Wood 2008 {published data only}

Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in

controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;**336**:601-5.

Additional references

Austin 1992

Austin SC, Stolley PD, Lasky T. The history of malariotherapy for neurosyphilis. *JAMA* 1992;**268**:516-9.

Black 1996

Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215-8.

Britton 1999

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Three systematic reviews - not so different answers. *BMJ Electronic letters* 29 September 1999.

Chalmers 1997

Chalmers I. Assembling comparison groups to assess the effects of health care. *Journal of the Royal Society of Medicine* 1997;**90**:379-86.

CMRG Module

The Editorial Team. Cochrane Methodology Review Group. About The Cochrane Collaboration (Cochrane Review Groups (CRGs)) 2005, Issue 2:Art. No.: METHOD.

Cochrane 1972

Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972:20-5.

Counsell 1994

Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis?. *BMJ* 1994;**309**:1677-81.

Dans 2002

Dans A, McAlister F, Dans L, Richardson WS, Straus S, Guyatt G. Applying the results to individual patients. In: Guyatt G, Rennie D editor(s). *Users' Guides to the Medical Literature*. Chicago: AMA Press, 2002:369-84.

Dickersin 1993

Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials* 1993;**Doc No 50**:4967.

Dickersin 1997

Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Education and Prevention* 1997;**9**(1 Suppl):15-21.

EC/IC Bypass 1985

The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *New England Journal of Medicine* 1985;**313**:1191-200.

Echt 1991

Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *New England Journal of Medicine* 1991;**324**:781-8.

Eclampsia 1995

Eclampsia Trial Collaborative Group. Which anticonvulsant for women with eclampsia?. *Lancet* 1995;**345**:1455-63.

Egger 1997

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

Gubler 1971

Gubler A. Alternating hemiplegia, a sign of pontine lesion, and documentation of the proof of the facial decussation. In: Wolf JK editor(s). *The Classical Brainstem Syndromes*. Springfield: Charles C Thomas, 1971:9-13.

Guyatt 2002

Guyatt GH, Cook D, Devereaux PJ, Meade M, Straus S. *Therapy. Users' guides to the medical literature*. Chicago: AMA Press, 2002:55-79.

Hopewell 2001

Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database of Methodology Reviews* 2001, Issue 3. [DOI: [10.1002/14651858.MR000011](https://doi.org/10.1002/14651858.MR000011)]

Hopewell 2006

Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *The Cochrane Database of Methodology Reviews* 2006, Issue 2. [DOI: [10.1002/14651858.MR000010.pub2](https://doi.org/10.1002/14651858.MR000010.pub2)]

Jüni 1999

Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**(11):1054-60.

Jüni 2001

Jüni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. In: Egger M, Smith GD, Altman D editor(s). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001.

Kleijnen 1997

Kleijnen J, Gøtzsche P, Kunz RH, Oxman AD, Chalmers I. So what's so special about randomisation?. In: Maynard A, Chalmers I editor(s). *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:93-106.

Kunz 1999

Kunz R, Oxman AD. Two systematic reviews - two different answers?. *BMJ Electronic Letters* 28 August 1999.

McKee 1999

McKee M, Britton A, Black N, Mcpherson K, Sanderson C, Bain C. *Methods in health services research. Interpreting the evidence:*

choosing between randomised and non-randomised studies. *BMJ* 1999;**319**:312-5.

Moore 1995

Moore TJ. *Deadly Medicine*. New York: Simon and Schuster, 1995.

MRC 1948

Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;**ii**:769-82.

Pockock 2000

Pockock SJ, Elbourne DR. Randomised trials or observational tribulations?. *New England Journal of Medicine* 2000;**324**:1907-9.

Reeves 1998

Reeves BC, MacLehose RR, Harvey IM, Sheldon TA, Russell IT, Black AMS. Comparison of effect size estimates derived from randomised and non-randomised studies. In: Black N, Brazier J, Fitzpatrick R, Reeves B editor(s). *Health Services Research Methods: A Guide to Best Practice*. London: BMJ Publishing Group, 1998:73-85.

Rothwell 2005

Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?". *Lancet* 2005;**365**:82-93.

Song 2000

Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000;**4**(10):1-115.

Teo 1993

Teo KK, Yusuf S, Furberg CD. Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction.

An overview of results from randomized trials. *JAMA* 1993;**270**:1589-95.

US Office HTA 1994

US Congress, Office of Technology Assessment. *Identifying Health Technologies that Work: Searching for Evidence*. OTA-H-608. Washington DC: US Government Printing Office, 1994:41-51.

Valenstein 1986

Valenstein ES. *Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Radical Treatments for Mental Illness*. New York: Basic Books, 1986.

Weiss 1998

Weiss CH. *Evaluation. Methods for Studying Programs and Policies*. 2nd Edition. Upper Saddle River: Prentice Hall, 1998:229-33.

References to other published versions of this review

Kunz 1998

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

Kunz 2002

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Methodology Reviews* 2002, Issue 4. [DOI: [10.1002/14651858.MR000012](https://doi.org/10.1002/14651858.MR000012)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Balk 2002

| | |
|-------------|--|
| Methods | The authors selected cardiovascular meta-analyses from among those used in a previous analysis by their group. For three other areas meta-analyses were found by searching MEDLINE and Cochrane Database of Systematic Reviews For each pre-defined quality measure a relative OR for treatment effect was calculated |
| Data | RCTs from 4 different areas of medicine: 93 RCTs from cardiovascular disease 56 RCTs from infectious disease 60 RCTs from paediatrics 67 RCTs from surgery |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |

Balk 2002 (Continued)

| | |
|----------|--|
| Outcomes | Mortality in studies from the area of cardiovascular disease; from the other 3 areas it varies between studies |
| Notes | — |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | No; they selected cardiovascular meta-analysis from their own group However, meta-analyses from infectious disease, paediatrics and surgery were found by searching MEDLINE and the Cochrane Database of Systematic Reviews. |
| Control for methodological differences, i.e. double-blinding? | No | YES for mortality, but NO for subjective outcomes; assessed and analysed for, but not controlled/adjusted for |
| Control for methodological differences, i.e. completeness of follow up? | No | Assessed and analysed for, but not controlled/adjusted for |
| Control for clinical differences in the participants? | No | Assessed and analysed for, but not controlled/adjusted for |
| Control for clinical differences in the interventions? | Yes | Used a Bayesian hierarchical model with random-effects that accounted for the nesting of trials within meta-analyses as well as the variability across meta-analyses |
| Explicit inclusion criteria? | Yes | Included meta-analysis with 6 or more RCTs and dichotomous outcomes, and significant between-study heterogeneity |
| Two or more agreed on inclusion? | Unclear | Not described |
| Two or more assessed quality and conducted data extraction? | Yes | "Data from each trial were extracted by 2 investigators" |
| Similar outcomes measured? | No | For cardiovascular studies: yes, mortality. Otherwise, the outcome used varied across meta-analysis |

Carroll 1996

| | |
|-------------|---|
| Methods | MEDLINE (1966 to 1995): Knowledge Server version 3.23: January 1996) and the Oxford Pain Relief Database (1950 to 1992) + reference lists |
| Data | 17 RCTs and 19 non-RCTs on transcutaneous electrical nerve stimulation (TENS) and postoperative pain |
| Comparisons | Randomised versus non-randomised trials for the same intervention and condition |
| Outcomes | Pain outcomes |

Carroll 1996 (Continued)

Notes 4 retrospective studies and 1 matched case-control study in the non-RCT group left out from our synthesis; leaving 14 non-RCTs

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | MEDLINE (1966 to 1995): Knowledge Server version 3.23: January 1996) and the Oxford Pain Relief Database (1950 to 1992) + reference lists |
| Control for methodological differences, i.e. double-blinding? | No | Analysed using vote-counting |
| Control for methodological differences, i.e. completeness of follow up? | No | Analysed using vote-counting |
| Control for clinical differences in the participants? | No | Analysed using vote-counting |
| Control for clinical differences in the interventions? | No | Analysed using vote-counting |
| Explicit inclusion criteria? | Yes | Inclusion criteria were full journal publication, TENS and postoperative pain with pain outcomes. Reports of TENS for the relief of other acute pain conditions, such as labour pain, acute infections and procedures, or those where the number of patients per treatment group was fewer than 10 were excluded |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Yes | Each report which could possibly meet the inclusion criteria was read by each author independently and scored for inclusion and quality using a 3-item scale |
| Similar outcomes measured? | No | Pain outcomes included, but reported by vote counting |

Chalmers 1983

| | |
|-------------|--|
| Methods | Therapeutic trials of treatment for acute myocardial infarction were identified through a MEDLINE search, Current Contents and a review of the references listed in the more recently published studies. Only studies that included a control group were included. |
| Data | 145 papers on the treatment of acute myocardial infarction; 102 randomised or quasi-randomised controlled trials and 43 non-random controlled trials |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions for the same condition |
| Outcomes | Case-fatality |
| Notes | The included trials were classified in 3 groups according to assignment to treatment groups: |

Chalmers 1983 (Continued)

Blinded randomisation (57 trials): assignment prearranged at random and communicated to the investigator only after the patient had been accepted for the study and informed consent had been obtained. **(In the context of this review corresponds to randomised controlled trials with adequate concealed randomisation)**

Unblinded randomisation (45 trials): assignment from an open table of random numbers, according to date of birth or chart number, or by some other variably random system in which the patient could present for study in chance order but be selected or rejected after physician knew the treatment assignment. **(In the context of this review corresponds to inadequately concealed trials (a mix of randomised and non-randomised trials))**

Non-random assignment (43 trials): included the use of both simultaneous selected controls and historical controls. Assignment to treatment was made by a method more susceptible to clinical judgement than chance. An example of the use of non-randomly assigned simultaneous controls was an evaluation of coronary-care units in which patients who were admitted were compared with those who were not admitted because the unit was full. **(In the context of this review corresponds to non-randomised controlled studies)**

In the context of this review the category non-random assignment will not be considered, as this group of trials included historical control and assignment to treatment by a method more susceptible to clinical judgement than change. For this reason the group 'non-random assignment' meets our prespecified criteria for exclusion. Thus this study will contribute data for the comparison 'Randomised controlled trials versus non-randomised controlled studies across different interventions' (57 adequately concealed randomised controlled trials and 43 non-randomised studies)

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | Trials were identified through a MEDLINE search, Current Contents and a review of references listed in the more recently published studies |
| Control for methodological differences, i.e. double-blinding? | Yes | They did not, but for case-fatality it will not introduce bias |
| Control for methodological differences, i.e. completeness of follow up? | No | No analyses performed that adjust for confounding factors |
| Control for clinical differences in the participants? | No | No analyses performed that adjust for confounding factors |
| Control for clinical differences in the interventions? | No | Data for subgroups presented, but not controlled or adjusted for in analysis |
| Explicit inclusion criteria? | Yes | Therapeutic trials of treatment of acute myocardial infarction were included. Only studies that used a control group were included. |
| Two or more agreed on inclusion? | Unclear | Not mentioned |
| Two or more assessed quality and conducted data extraction? | Unclear | Not mentioned |
| Similar outcomes measured? | Yes | Case-fatality |

Clifford 2002

| | |
|-------------|--|
| Methods | A convenience sample of 100 randomised controlled trials (RCTs) was identified by handsearching recent issues of 5 peer-reviewed, high impact factor general medical journals (<i>Annals of Internal Medicine</i> , <i>BMJ</i> , <i>JAMA</i> , <i>Lancet</i> , the <i>New England Medical Journal</i>). Issues published between January 1999 and October 2000 were searched until 20 RCTs/journal were identified |
| Data | 100 RCTs from any clinical area |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |
| Outcomes | Main outcome |
| Notes | RRs calculated from information in paper |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | Issues from 5 top journals were handsearched from January 1999 to October 2002 until 20 RCTs/journal were identified |
| Control for methodological differences, i.e. double-blinding? | No | Analysed as a outcome in its own right, but not controlled for in the relevant comparison |
| Control for methodological differences, i.e. completeness of follow up? | No | Analysed as a outcome in its own right, but not controlled for in the relevant comparison |
| Control for clinical differences in the participants? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the interventions? | No | No adjusted or controlled analyses performed |
| Explicit inclusion criteria? | Yes | The RCTs needed to be published as a full report. Interventions were restricted to pharmaceuticals (medical devices, surgical procedures and methods of medical management were excluded). No attempt was made to limit the selection of any particular RCT design, number of treatment arms, comparator, study population or disease category. |
| Two or more agreed on inclusion? | Unclear | Not mentioned |
| Two or more assessed quality and conducted data extraction? | Yes | "Reporting quality was evaluated by two independent, experienced reviewers" |
| Similar outcomes measured? | Unclear | The primary outcome was defined as the one stated as such by the authors, if there was no such statement, the one that was most clinically relevant. Highly likely different from different trials |

Colditz 1989

| | |
|-------------|---|
| Methods | After ranking journals listed under these disciplines (cardiology, neurology, psychiatry and respiratory medicine) in the Index Medicus in 1980 by their impact factor, they drew a stratified random sample of journals within each discipline |
| Data | 36 randomised controlled trials compared with 3 non-randomised parallel studies 29 randomised controlled cross-over trials, 46 non-randomised sequential comparisons, 5 externally-controlled studies and 9 observational studies |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions |
| Outcomes | Gain, as defined by the Mann-Whitney statistics and a rating of the authors' conclusion. Based on a response to therapy outcome measure - the basis for gain and rating varies across studies. |
| Notes | In this review only the 36 randomised controlled trials, 3 non-randomised parallel studies, 29 randomised controlled cross-over trials and 46 non-randomised sequential comparisons were considered |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | After ranking journals listed under these disciplines (cardiology, neurology, psychiatry and respiratory medicine) in the Index Medicus in 1980 by their impact factor, they drew a stratified random sample of journals within each discipline |
| Control for methodological differences, i.e. double-blinding? | No | No adjusted or controlled analyses performed |
| Control for methodological differences, i.e. completeness of follow up? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the participants? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the interventions? | No | No adjusted or controlled analyses performed |
| Explicit inclusion criteria? | Yes | An evaluation of medical therapy with the response to therapy as the outcome measure, at least 10 subjects, and outcome reported for both comparators |
| Two or more agreed on inclusion? | Unclear | They subjected the articles to a second level of reading to determine final eligibility |
| Two or more assessed quality and conducted data extraction? | Yes | 2 readers with training in statistical methods independently read each article. Pairs of readers recorded the study design and completed a checklist |
| Similar outcomes measured? | No | Gain, as defined by the Mann-Whitney statistics and a rating of the authors' conclusion. Based on a response to therapy outcome measure - the basis for gain and rating varies across studies. |

Egger 2003

| | |
|-------------|---|
| Methods | For the comparisons regarding methodological quality every systematic review published in Issue 1 1998 of the Cochrane Database of Systematic Reviews were searched for relevant meta-analyses. Only meta-analyses where information on quality was available for at least 80% of included trials and which contained both trials with and without the quality characteristic were included in the analyses. |
| Data | <p>39 meta-analysis including 118 trials with adequate allocation concealment and 186 trials with inadequate or unclear allocation concealment. The trials were within the following medical specialities:</p> <p>Infectious diseases: 30 trials with adequate concealment and 25 with inadequate/unclear concealment</p> <p>Neurology: 18 trials with adequate concealment and 16 with inadequate/unclear concealment</p> <p>Obstetrics and gynaecology: 46 trials with adequate concealment and 76 with inadequate/unclear concealment</p> <p>Other: 24 trials with adequate concealment and 69 with inadequate/unclear concealment</p> |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |
| Outcomes | Unclear |
| Notes | The authors note: "Trials that reported adequate concealment of allocation were published more recently and enrolled more participants than trials with inadequate or unclear concealment of allocation. Interestingly, there was no difference in the distribution of p-values, despite the clear difference in sample size." |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | For our comparison of adequate versus inadequate concealment of allocation, inclusion was restricted to meta-analysis published in Cochrane (1998) |
| Control for methodological differences, i.e. double-blinding? | No | No controlled analyses performed |
| Control for methodological differences, i.e. completeness of follow up? | No | No controlled analyses performed |
| Control for clinical differences in the participants? | Yes | Ratios of pooled estimates from adequate trials to pooled estimates from inadequate or unclear trials within each meta-analysis were pooled using random-effects meta-analysis |
| Control for clinical differences in the interventions? | Yes | Ratios of pooled estimates from adequate trials to pooled estimates from inadequate or unclear trials within each meta-analysis were pooled using random-effects meta-analysis. Subgroup analysis for clinical areas |
| Explicit inclusion criteria? | Yes | Meta-analysis of therapeutic or preventive interventions that were based on comprehensive literature searches, which combined the binary outcome of at least 5 controlled trials. Method of analysis well described and information about the quality available for at least 80% of the trials. |
| Two or more agreed on inclusion? | Unclear | Not mentioned |

Egger 2003 (Continued)

| | | |
|---|-----|--|
| Two or more assessed quality and conducted data extraction? | Yes | 2 of the reviewers independently classified all component trials from the eligible meta-analysis |
| Similar outcomes measured? | No | They analysed by disease area, not by outcome |

Hedrick 1989

| | | |
|-------------|--|--|
| Methods | Study reports were located through a combination of searches of article files in gerontologic and health-services libraries, searches of computerised databases (MEDLARS and Health Planning and Administration), and personal contacts with researchers in the field. Contacted the investigators of many of the studies cited in previous reviews or in their own work on related topics to ask about other relevant research. | |
| Data | 11 RCTs compared with 2 quasi-experimental studies. Home care. | |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions | |
| Outcomes | Mortality | |
| Notes | — | |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | Computerised databases (MEDLARS, Health Planning and Administration) and personal contacts with researchers in the field |
| Control for methodological differences, i.e. double-blinding? | Yes | They did not, but reported mortality and nursing-home placement |
| Control for methodological differences, i.e. completeness of follow up? | No | No controlled analyses performed |
| Control for clinical differences in the participants? | No | No controlled analyses performed |
| Control for clinical differences in the interventions? | No | Performed an analysis based on all included to studies to see if effect varied according to 2 different characteristics of the interventions (team approach and physician involved in intervention) |
| Explicit inclusion criteria? | Yes | RCTs or quasi-experimental studies of home care. The experimental group receiving home care services. Studies with composite intervention where other community services are included were excluded. |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Unclear | Not reported |

Randomisation to protect against selection bias in healthcare trials (Review)

Hedrick 1989 (Continued)

| | | |
|----------------------------|-----|--------------------------------------|
| Similar outcomes measured? | Yes | Mortality and nursing home placement |
|----------------------------|-----|--------------------------------------|

Heinsman 1996

| | |
|-------------|---|
| Methods | For 2 two healthcare outcomes the present study drew from 2 past meta-analyses that contained both random and non-randomised experiments on psychosocial interventions for postsurgery outcomes and juvenile drug use prevention programmes |
| Data | Drug use prevention 12 RCTs compared with 17 non-random experiments Presurgical intervention 27 RCTs versus 14 non-randomised experiments |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions |
| Outcomes | Standardised mean differences from a variety of different outcomes |
| Notes | — |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | No | Convenience sample of 4 past meta-analyses. Juvenile drug use prevention, psychosocial interventions for postsurgery outcomes, coaching for Scholastic Aptitude Test performance and ability grouping in secondary school classes. Not all are healthcare outcomes |
| Control for methodological differences, i.e. double-blinding? | Unclear | Regression analysis conducted, but adjustments for blinding not mentioned |
| Control for methodological differences, i.e. completeness of follow up? | Yes | Adjusted for total attrition rate and percentage differential attrition in regression model |
| Control for clinical differences in the participants? | No | Not adjusted for in regression model |
| Control for clinical differences in the interventions? | No | Not adjusted for in regression model |
| Explicit inclusion criteria? | Yes | Studies that compared treatments with control conditions rather than with other treatments and did so at the post-test rather than at follow up. Excluded studies that did not report the statistics required. Excluded effect sizes reported only as significant or non-significant. Excluded dichotomous outcomes. Excluded unclear subject assignment or haphazard assignment |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Unclear | 2 authors trained to meet reliability from each of 30 studies. Unclear if all were collected in duplicate |

Heinsman 1996 (Continued)

| | | |
|----------------------------|----|---------------|
| Similar outcomes measured? | No | They used SMD |
|----------------------------|----|---------------|

Kjaergard 2001

| | |
|-------------|--|
| Methods | <i>The Cochrane Library</i> , MEDLINE on PubMed and reference lists of relevant articles were searched to identify potentially eligible meta-analyses than included at least 1 large trial (at least 1000 participants) |
| Data | 14 meta-analyses with a total of 190 randomised controlled trials |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |
| Outcomes | Primary binary outcome measure described by the largest number of trials in each meta-analysis |
| Notes | Correction published: Kjaergard LL, Villumsen J, Gluud C. Correction: Reported methodologic quality and discrepancies between large and small randomised trials in meta-analysis. <i>Annals of Internal Medicine</i> 2008;149:219 |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | They searched <i>The Cochrane Library</i> , MEDLINE on PubMed (using meta-analysis, review as text words) + reference lists |
| Control for methodological differences, i.e. double-blinding? | Yes | 7 of 11 outcomes were objective: mortality or caesarean section. Subjective outcomes: the effect of blinding was assessed but not controlled for |
| Control for methodological differences, i.e. completeness of follow up? | No | Not adjusted for in analyses |
| Control for clinical differences in the participants? | Yes | Among the independent factors that were adjusted for was the interaction between treatment group and trials nested within meta-analysis (difference in treatment effect between different trials/participants) |
| Control for clinical differences in the interventions? | Yes | Among the independent factors that were adjusted for was the interaction between treatment group and meta-analysis (difference in treatment effect between different meta-analyses/interventions) |
| Explicit inclusion criteria? | Yes | Meta-analysis that included at least 1 large trial (1000 or more participants). Excluded meta-analysis that had excluded studies of low quality. Excluded trials that were unpublished trials, quasi-randomised trials |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Yes | "Data were extracted independently by two reviewers" |
| Similar outcomes measured? | No | The primary outcome measure described by the largest number of trials in each meta-analysis (5 mortality, 2 caesarean section, deep venous thrombosis, |

Randomisation to protect against selection bias in healthcare trials (Review)

Kjaergard 2001 (Continued)

dropouts, endocervical cells, resumed smoking). They do re-express the outcomes as unwanted endpoints and so analysed on the same scale, but we consider this similar to SMD calculations.

Linde 1999

| | |
|-------------|---|
| Methods | Eligible trials were identified through multiple sources including MEDLINE, EMBASE, complementary medicine databases, contacts with researchers, and checking bibliographies of identified articles. Eligible trials had to be double-blinded and/or randomised placebo-controlled clinical trials |
| Data | 89 placebo-controlled clinical trials of homeopathy |
| Comparisons | Explicit random allocation (64) versus not explicit randomisation (25) - qualifies as studies of 'Randomised versus non-randomised trials across different interventions and conditions' Controlled trials with adequate (34) versus inadequate or unclear concealment (55) of allocation across different interventions and conditions |
| Outcomes | Varied across included trials. Selected according to preference list: 1) Pre-defined main outcome measure (i.e.. outcome on which sample size calculations was based) 2) Patients' global assessment of improvement, if measured 3) Physicians' global assessment of improvement 4) Outcome measures that, in the judgement of the reviewers, were the most important 5) Else randomly selected from reported outcomes |
| Notes | Re-analysis of data from Linde K, Clausius N, Ramirez G, Melcart D, Eitel F, Hedges LV, et al. Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. <i>Lancet</i> 1997;350:834-43. 74 of the trials included in this study are also included in Shang 2005 |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | "Eligible trials were identified through multiple sources including MEDLINE, EMBASE, complementary medicine databases, contacts with researchers, and checking bibliographies of identified articles." |
| Control for methodological differences, i.e. double-blinding? | Yes | Multivariate component analysis which adjusted for: explicitly randomised, adequate concealment, double-blinding and complete follow up |
| Control for methodological differences, i.e. completeness of follow up? | Yes | Multivariate component analysis which adjusted for: explicitly randomised, adequate concealment, double-blinding and complete follow up |
| Control for clinical differences in the participants? | No | Analyses did only adjust for methodological differences |
| Control for clinical differences in the interventions? | No | Analyses did only adjust for methodological differences |

Linde 1999 (Continued)

| | | |
|---|---------|--|
| Explicit inclusion criteria? | Yes | “We included all available double-blind and/or randomised clinical trials in which a homeopathic intervention and a placebo had been compared for preventive or therapeutic purposes.” |
| Two or more agreed on inclusion? | Unclear | Not reported in paper |
| Two or more assessed quality and conducted data extraction? | Yes | “Study characteristics and results were extracted by two independent reviewers using a pretested form” |
| Similar outcomes measured? | No | Outcomes varied across trials. Preferred pre-defined main outcome measure. |

Lipsey 1993

| | | |
|-------------|---|--|
| Methods | A series of computer and manual searches of bibliographies of articles dealing with meta-analyses, various standard social science abstracts (Psychological abstracts, Sociological abstracts, etc) and listings of unpublished materials (ERIC). Of interest here was meta-analysis of research on the effects of treatments that are based on manipulation of psychological variables and are intended to induce psychological change, whether emotional, attitudinal, cognitive or behavioral (referred to as psychological treatments. Attention is restricted to those treatments that are directed at practical individual and social problems. | |
| Data | 302 meta-analyses included: 137 within mental health 11 within work setting or organisational studies 154 within education | |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions | |
| Outcomes | Different across meta-analyses. The effect size metric used is the standardised difference between the mean of the treatment group and the mean of the control group for a given outcome in a given study. | |
| Notes | For our comparison the analysis is based on 74 meta-analyses, from which effect sizes for randomised trials and non-randomised studies could be extracted separately | |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | Yes | A series of computer and manual searches of bibliographies of articles dealing with meta-analyses, various standard social science abstracts (Psychological abstracts, Sociological abstracts, etc) and listings of unpublished materials (ERIC) |
| Control for methodological differences, i.e. double-blinding? | No | No adjusted or controlled analyses performed |
| Control for methodological differences, i.e. completeness of follow up? | No | No adjusted or controlled analyses performed |

Lipsey 1993 (Continued)

| | | |
|---|---------|---|
| Control for clinical differences in the participants? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the interventions? | No | No adjusted or controlled analyses performed |
| Explicit inclusion criteria? | Unclear | Not reported |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Unclear | Not reported |
| Similar outcomes measured? | No | Typically, a mean effect size over all studies and outcome measures is shown for each study |

Miller 1989

| | |
|-------------|--|
| Methods | All articles published during 1983 in 6 surgery journals were reviewed for inclusion: <i>American Journal of Surgery, Annals of Surgery, Archives of Surgery, British Journal of Surgery, Surgery and Surgery, Gynecology and Obstetrics</i> |
| Data | 81 Randomised controlled trials, 15 non-randomised controlled studies, 27 externally controlled trials, 91 observational studies and 7 pre/post comparisons |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions |
| Outcomes | Gain, as defined by the Mann-Whitney statistic and a rating of the authors' conclusion. Based on a response to therapy outcome measure - the basis for gain and rating varies across studies. |
| Notes | In this review only the 81 randomised controlled trials and 15 non-randomised studies were considered |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | No | Articles published in 1983 in 6 leading surgery journals |
| Control for methodological differences, i.e. double-blinding? | No | No adjusted or controlled analyses performed |
| Control for methodological differences, i.e. completeness of follow up? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the participants? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the interventions? | No | No adjusted or controlled analyses performed |

Miller 1989 (Continued)

| | | |
|---|---------|---|
| Explicit inclusion criteria? | Yes | An evaluation of medical therapy with the response to therapy as the outcome measure, at least 10 subjects, and outcome reported for both comparators |
| Two or more agreed on inclusion? | Yes | 2 readers independently read each article to decide whether it qualified for inclusion |
| Two or more assessed quality and conducted data extraction? | Unclear | No mention |
| Similar outcomes measured? | No | Gain, as defined by the Mann-Whitney statistic and a rating of the authors' conclusion. Based on a response to therapy outcome measure - the basis for gain and rating varies across studies. |

Moher 1998

| | |
|-------------|---|
| Methods | 12 meta-analyses were randomly (random numbers table) selected from the investigators database of 491 meta-analyses of RCTs; 3 each on digestive diseases, circulatory diseases and mental health, and further 3 randomly chosen from the Cochrane Database of Systematic Reviews - 1 on stroke and 2 on pregnancy and childbirth. One meta-analysis excluded was provided to the principal investigator solely for the purpose of his meta-analysis. |
| Data | 127 RCTs from 11 meta-analyses |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |
| Outcomes | Main outcome |
| Notes | — |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | No | "We randomly (random number table) selected 12 meta-analysis from our large database of 491 meta-analysis of RCTs." |
| Control for methodological differences, i.e. double-blinding? | Unclear | Logistic-regression models were used to explore the relation between the binary outcome and individual component (e.g. double-blinding). But included in our results? The majority of outcomes (15 of 22) can be defined as objective (histological remission, major amputation, overall mortality, conception rate (still too many that is not objective)) |
| Control for methodological differences, i.e. completeness of follow up? | No | Not adjusted for in analyses |
| Control for clinical differences in the participants? | Yes | Among the independent factors that were adjusted for were trial indicators to allow for variation among the trials (differences in participants) |

Moher 1998 (Continued)

| | | |
|---|-----|--|
| Control for clinical differences in the interventions? | Yes | Among the independent factors that were adjusted for were modified treatment effects to capture variation among the meta-analyses (differences in interventions) |
| Explicit inclusion criteria? | Yes | 3 inclusion criteria: English; no formal incorporation of quality scores in the quantitative analysis; that the outcomes were binary data; and that summary results were available |
| Two or more agreed on inclusion? | Yes | "Each meta-analysis was reviewed by two of the investigators to agree on the reported principal outcome or outcomes" |
| Two or more assessed quality and conducted data extraction? | Yes | "The quality of the reporting of each of the resulting 254 RCTs was assessed by all of the investigators..." "The data were extracted independently by two investigators." |
| Similar outcomes measured? | No | Primary outcome or the outcome with the most trials included in the meta-analysis |

Ottenbacher 1991

| | |
|-------------|---|
| Methods | A selection of 44 articles from the <i>American Journal of Occupational Therapy</i> (AJOT) and the <i>Occupational Therapy Journal of Research</i> (OTJR). Review of individual studies beginning with the last issue in the 1990 volume year and working backward through previous issues. The range of issues reviewed was from 1981 through 1990 for OTJR and from 1980 through 1990 for AJOT. |
| Data | 22 pretest-post randomised controlled trials versus 22 pretest-post non-randomised controlled studies |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions |
| Outcomes | Varied across studies. The d-index was computed for each of the primary hypothesis under evaluation in the 44 trials |
| Notes | — |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | 44 articles from two journals (AJOT and OTJR) during 1990 back to 1981 (OTJR) and 1980 (AJOT) |
| Control for methodological differences, i.e. double-blinding? | No | Unadjusted analyses |
| Control for methodological differences, i.e. completeness of follow up? | No | Unadjusted analyses |
| Control for clinical differences in the participants? | No | Unadjusted analyses |
| Control for clinical differences in the interventions? | No | Unadjusted analyses |

Ottenbacher 1991 (Continued)

| | | |
|---|-----|--|
| Explicit inclusion criteria? | Yes | Study with pretest - post-test control group design, involving a comparative research question related to the effectiveness of some therapeutic intervention |
| Two or more agreed on inclusion? | No | Ottenbacher only |
| Two or more assessed quality and conducted data extraction? | Yes | "The 44 articles were coded by two raters ... " (p. 920) |
| Similar outcomes measured? | No | All outcomes represented by effect sizes |

Ottenbacher 1992

| | | |
|-------------|---|--|
| Methods | The articles included in the analysis were identified by reviewing individual issues of <i>JAMA</i> and the <i>New England Journal of Medicine</i> beginning with the last issue of 1989 volume year and backward through previous issues. From each journal 15 trials using a parallel-group design with random assignment and 15 trials using a similar design without random assignment based on recentness of publication | |
| Data | 30 random trials and 30 non-random studies | |
| Comparisons | Randomised versus non-randomised trials across different interventions and conditions | |
| Outcomes | Varied across studies. The d-index was computed for each of the primary hypothesis under evaluation in the 60 trials. | |
| Notes | — | |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|--|
| Complete sample of trials? | No | 30 articles from JAMA and 30 articles from NEJM |
| Control for methodological differences, i.e. double-blinding? | No | No adjusted or controlled analyses performed |
| Control for methodological differences, i.e. completeness of follow up? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the participants? | No | No adjusted or controlled analyses performed |
| Control for clinical differences in the interventions? | No | No adjusted or controlled analyses performed |
| Explicit inclusion criteria? | Yes | An examination of therapeutic effectiveness using a parallel-group design in which 1 group received the intervention and the other group did not. Contained sufficient information to compute an effect size measure, vital information about design and analysis characteristics. |

Ottensbacher 1992 (Continued)

| | | |
|---|-----|---|
| Two or more agreed on inclusion? | Yes | 2 examiners with research doctorates and clinical research experience independently reviewed each issue |
| Two or more assessed quality and conducted data extraction? | Yes | The 60 articles were coded by two Ph.D. trained researchers |
| Similar outcomes measured? | No | The d-index was computed for each of the primary hypotheses |

Schulz 1995

| | | |
|-------------|---|--|
| Methods | The systematic review of controlled trials used in this methodology study have all been published by the Pregnancy and Childbirth Group of The Cochrane Collaboration. Published and unpublished primary trials potentially relevant for the review were entered into a register. The database contained more than 500 systematic reviews. The authors derived a defined universe from all the reviews in 3 steps. First, they identified 82 meta-analyses, which included at least 5 trials with a total of at least 25 outcome events among the control group. Second, all meta-analyses to which component trials had contributed and retained only the meta-analysis with the most homogeneous grouping of interventions for inclusion. Third, the meta-analyses had to comprise at least 1 component trial with adequate concealment of the allocation schedule and 1 trial without. | |
| Data | 250 controlled trials from 33 meta-analyses; 36 trials with adequate concealment of allocation, 21 trials with inadequate concealment of allocation, and 150 trials with unclearly concealed allocation | |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions | |
| Outcomes | Each meta-analysis investigated similar comparison groups with the same binary outcome measure. However, the outcome measures might vary across meta-analyses. | |
| Notes | — | |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | Only included trials that were included in systematic reviews from the register of systematic reviews maintained by the Cochrane Pregnancy and Childbirth Group |
| Control for methodological differences, i.e. double-blinding? | Yes | Adjusting for double-blinding and adequate sequence generation in multiple logistic regression model |
| Control for methodological differences, i.e. completeness of follow up? | Yes | Adjusting for "exclusion of randomised participants" in multiple logistic regression model |
| Control for clinical differences in the participants? | Yes | "Indicator variables to Control for the effects in each of the 250 trials" (cited from footnotes to Table 1 and Table 2 in the paper) |
| Control for clinical differences in the interventions? | Yes | "Terms for the "Meta-analysis by treatment group" interaction to control for the different summary odds ratios for the treatment effects in the 33 meta-analyses" |

Schulz 1995 (Continued)

| | | |
|---|---------|---|
| Explicit inclusion criteria? | No | For some trials that was included in more than one of the included 33 meta-analyses only one listing was included in this study. The choice of meta-analysis was decided using a random-number table. |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | No | One of the authors assessed the methodological quality of the included trials. 10 randomly chosen trials were reassessed by a second author. |
| Similar outcomes measured? | Unclear | "Each (<i>meta-analysis</i>) investigated similar comparison groups with the same binary outcome measure" |

Shadish 1996

| | | |
|-------------|---|--|
| Methods | Most of the studies had already been gathered in the process of doing a previous meta-analysis. How the remaining studies were found is unclear. | |
| Data | 100 studies of marital or family psychotherapy (n = 84) or enrichment (n = 16). The sample consists of: 34 published randomised experiments 17 published non-equivalent control group designs 30 unpublished randomised experiments 19 unpublished non-equivalent control group designs | |
| Comparisons | Randomised versus non-randomised trials across different interventions for the same condition | |
| Outcomes | Varied across studies. The standardised mean difference was calculated for continuous outcomes from each study. Effect sizes within studies were averaged to the study level | |
| Notes | — | |

Risk of bias

| Item | Authors' judgement | Description |
|---|--------------------|---|
| Complete sample of trials? | No | Convenience sample of a subset of studies from a previous meta-analysis. "Most studies has already been gathered in the process of doing a previous meta-analysis". |
| Control for methodological differences, i.e. double-blinding? | No | No adjustment done |
| Control for methodological differences, i.e. completeness of follow up? | Yes | Adjusted for effects of level of attrition |
| Control for clinical differences in the participants? | No | No adjustment done |

Shadish 1996 (Continued)

| | | |
|---|---------|--|
| Control for clinical differences in the interventions? | No | No adjustment done |
| Explicit inclusion criteria? | Yes | Studies that compared treatments with control conditions rather than with other treatments and did so at the post-test rather than at follow up. Excluded studies that did not report the statistics required. Excluded effect sizes reported only as significant or non-significant. Excluded dichotomous outcomes. Excluded unclear subject assignment or haphazard assignment |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Unclear | 2 authors trained to meet reliability from each of 30 studies. Unclear if all were collected in duplicate |
| Similar outcomes measured? | No | They used standardised mean difference |

Shang 2005

| | |
|-------------|--|
| Methods | <p>The authors updated a previous comprehensive search for placebo-controlled trials of homeopathy which covered publication up to August 1995 (Linde 1999). They searched 19 electronic databases covering the period from 1995 to January 2003: MEDLINE, Pre-MEDLINE, EMBASE, DARE, CCTR, CDSR, CINAHL, AMED, MANTIS, Toxline, PASCAL, BIOL, Science Citation Indx, CISCOM, British Homeopathic Library, the Homeopathy Abstract page, HomInform Homeopathic Library, NCCAM and SIGLE. They also checked the reference lists of relevant papers, including reviews and meta-analyses of homeopathic interventions, and contacted experts in the speciality.</p> <p>The authors searched the Cochrane Controlled Trials Register to identify placebo-controlled trials of conventional medicine</p> <p>For each homeopathy trial, the authors identified matching trials of conventional medicine that enrolled patients with similar disorders and assessed similar outcomes. They used computer-generated random numbers to select one of several eligible trials of conventional medicine. Outcomes were selected and trials matched without knowledge of trial results.</p> |
| Data | <p>110 controlled trials of homeopathy</p> <p>110 controlled trials of conventional medicine</p> |
| Comparisons | Controlled trials with adequate versus inadequate or unclear concealment of allocation across different interventions and conditions |
| Outcomes | Varied across studies. Eligible trials should be available with sufficient data to allow the calculation of odds ratios. |
| Notes | 74 of the homeopathy trials included in this study are also included in Linde 1999 |

Risk of bias

| Item | Authors' judgement | Description |
|----------------------------|--------------------|---|
| Complete sample of trials? | No | Comprehensive search including 19 electronic databases for trials of homeopathy |

Shang 2005 (Continued)

Matched trials of conventional medicine were found in the Cochrane Controlled Trials Register; that is a convenience sample of controlled trials of conventional medicine

| | | |
|---|---------|---|
| Control for methodological differences, i.e. double-blinding? | Yes | Results from univariate meta-regression analysis of treatment effects are reported. The authors have performed multivariate meta-regression analysis which was adjusted for trial quality (masking, generation of allocation sequence, and intention-to-treat analysis); but these results are only reported as "In multivariable analyses, the SE of the log odds ratio was the dominant variable in both groups. Coefficients of other variables, including study quality was attenuated and became non-significant". |
| Control for methodological differences, i.e. completeness of follow up? | No | The analysis did not adjust for completeness of follow up |
| Control for clinical differences in the participants? | No | Only for homoeopathy trials, the authors examined whether effects varied between types of indications (acute, chronic, primary prevention or prophylaxis) |
| Control for clinical differences in the interventions? | No | Only for homoeopathy trials, the authors examined whether effects varied between types of homoeopathy |
| Explicit inclusion criteria? | Yes | <p>Inclusion criteria: controlled trials of treatments of preventive measures with clinical outcomes; parallel-group design with placebo-control; random- or quasi-random assignment to treatment and placebo groups; a written report was available with sufficient data to allow the calculation of odds ratio.</p> <p>Exclusion criteria: homoeopathic "provings" in which remedies are given to healthy individuals to assess their effect, cross-over trials and N-of-1 trials</p> |
| Two or more agreed on inclusion? | Unclear | Not reported |
| Two or more assessed quality and conducted data extraction? | Yes | Data were extracted independently by two observers, and discrepancies were resolved by consensus. |
| Similar outcomes measured? | No | Matched between trials of homeopathy and trials of conventional medicine, but varied between trials within each group |

AJOT: American Journal of Occupational Therapy
 JAMA: Journal of the American Medical Association
 NEJM: New England Journal of Medicine
 OR: odds ratio
 OTJR: Occupational Therapy Journal of Research
 RCT: randomised controlled trial
 RR: risk ratio
 SMD: standardised mean difference

Characteristics of excluded studies [ordered by study ID]

| Study | Reason for exclusion |
|------------------------------|---|
| Aronson 1996 | Comparison of interest was incidental to main aim of study |
| Benson 2000 | Comparison of randomised trials versus observational studies (non-experimental studies) |

| Study | Reason for exclusion |
|----------------------------------|--|
| Bhansali 1996 | Comparison of randomised trials versus historical controlled trials |
| Chalmers 1977 | Comparison of interest was incidental to main aim of study |
| Diehl 1986 | Comparison of randomised trials versus historical controlled trials |
| Emerson 1990 | Comparison of high versus low quality randomised controlled trials |
| Forge 1998 | Comparison of interest was incidental to main aim of study |
| Gilbert 1977 | Narrative assessment only |
| Guyatt 2000 | Comparison of randomised trials versus cohort studies |
| Hovell 1982 | No systematic review: no search strategy; no explicit inclusion or exclusion criteria; no explicit re-search question; narrative assessment |
| Hutchinson 1999 | Comparison of interest was incidental to main aim of study |
| Imperiale 1990 | Comparison of high versus low quality randomised controlled trials |
| Ioannidis 2001 | Non-randomised studies included retrospective cohort studies, case-control studies, studies with historical controls (observational studies). It was therefore not a clearcut comparison between randomised trials and non-randomised studies, as per our definition. |
| Kasiske 1993 | Comparison of interest was incidental to main aim of study |
| Kasiske 1998 | Comparison of interest was incidental to main aim of study |
| Kerlikowske 1995 | Comparison of randomised trials versus case-control studies |
| Khan 1996 | Comparison of high versus low quality randomised controlled trials |
| Koes 1994 | Descriptive assessment of individual studies, no summary comparison provided |
| Kownacki 1999 | Comparison of RCT versus non-RCTs on the effect of alcoholics anonymous. Coerced participation in group sessions in RCTs and voluntary participation in non-RCTs together with contradicting results raise high suspicion of confounder, as acknowledged by the authors. |
| MacArthur 1995 | Narrative assessment only |
| Mehta 1999 | Descriptive assessment of individual studies, no summary comparison provided |
| Moher 1999 | Studies and comparisons already included in Moher 1998 |
| Morrison 1997 | Based on a selection of trials from a study where the comparison of interest was incidental |
| Mullen 1997 | Comparison of interest was incidental to main aim of study |
| Nurmohamed 1992 | Comparison of high versus low quality randomised controlled trials |
| Ortiz 1998 | Comparison of high versus low quality randomised controlled trials |
| Ottenbacher 1993 | Comparison of interest was incidental to main aim of study |

| Study | Reason for exclusion |
|---------------------------------|---|
| Pagnin 2004 | Comparison of interest was incidental to main aim of study |
| Potter 1998 | Comparison of high versus low quality randomised controlled trials |
| Pyorala 1995 | Comparison of interest was incidental to main aim of study |
| RMIT Group 1994 | Comparison of interest was incidental to main aim of study |
| Rozenberg 1999 | Narrative assessment only |
| Sacks 1982 | Comparison of randomised trials versus historical controlled trials |
| Shadish 1997 | Review of Heinsman 1996 which is already included |
| Shadish 2001 | Reanalysis of some of the same data that are already presented in Lipsey 1993 |
| Stanton 1997 | Comparison of high versus low quality randomised controlled trials |
| Stieb 1990 | Comparison of randomised trials versus case-control studies |
| Watson 1994 | Comparison of interest was incidental to main aim of study |
| Weisburd 2001 | Non-healthcare interventions |
| Wortman 1983 | Comparison of interest was incidental to main aim of study |

RCT: randomised controlled trial

DATA AND ANALYSES

Comparison 1. Studies of randomised versus non-randomised studies of the same intervention and condition

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|-----------------------------|----------------|---------------------|--------------------|-----------------|
| 1 Undefined | | | Other data | No numeric data |

Analysis 1.1. Comparison 1 Studies of randomised versus non-randomised studies of the same intervention and condition, Outcome 1 Undefined.

| Study | SAMPLE | Undefined COMPARISONS | RESULTS | DIRECTION OF BIAS |
|------------------------------|---|--|--|---|
| Carroll 1996 | 17 RCTs versus 14 non-RCTs (no information on the design) on transcutaneous electrical nerve stimulation (TENS) in acute postoperative pain | Comparison of RCTs and non-RCTs on analgesic effectiveness | Reported using vote-counting. 12 of 14 non-RCTs were judged positive for TENS. 15 of the 17 RCTs were judged to show no effect | Larger effects in non-RCTs than in RCTs (High risk of bias) |

Comparison 3. Studies of randomised versus non-randomised studies across different interventions for the same condition

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---------------------------|----------------|---------------------|--------------------|-----------------|
| 1 Undefined | | | Other data | No numeric data |

Analysis 3.1. Comparison 3 Studies of randomised versus non-randomised studies across different interventions for the same condition, Outcome 1 Undefined.

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|--------------|--|---|--|--|
| Shadish 1996 | 100 comparative studies (34 published and 30 non-published RCTs; 17 published and 19 non-published non-RCTs) of marital and family psychotherapy identified through a systematic search. All non-RCTs were non-equivalent control group designs. | Comparison of the effect size of all RCTs versus non-RCTs; effect sizes present at pre-test, publication status, level of attrition, matching and stratification as well as regression analysis including all important independent variables | The overall effect observed in non-RCTs was 87% smaller than the one observed in RCTs ($P < 0.05$). This difference was weaker but was maintained after control for other methodological features. Correlation between pre- and post-test effect size was significant in both designs, but much stronger in non-RCTs (0.84) than in RCTs (0.39). | Smaller effects in non-RCTs than in RCTs (High risk of bias) |

Comparison 4. Studies of randomised versus non-randomised studies across different interventions and conditions

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---------------------------|----------------|---------------------|--------------------|-----------------|
| 1 Undefined | | | Other data | No numeric data |

Analysis 4.1. Comparison 4 Studies of randomised versus non-randomised studies across different interventions and conditions, Outcome 1 Undefined.

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|--------------|---|--|---|----------------------------------|
| Colditz 1989 | 114 studies published in 1980 comparing new interventions with old, identified in leading cardiology, neurology, psychiatry and respiratory journals by a systematic search | 36 parallel randomised controlled trials, 29 randomised and 46 non-randomised sequential comparisons and 3 non-randomised parallel comparisons were compared on 'treatment gain' (Mann-Whitney statistic) and the relation between quality score and 'treatment gain' was assessed | All but one design yielded similar estimates of 'treatment gains': parallel randomised controlled trials 0.61, randomised controlled cross-over trials 0.63, non-randomised parallel comparisons 0.56 and non-randomised sequential comparisons 0.81. Only non-randomised sequential comparisons detected a significantly higher 'treatment gain' from the new treatment compared to randomised controlled parallel trials. | Inconclusive (High risk of bias) |
| Hedrick 1989 | 11 RCTs and 2 non-RCTs (unclear design) of home care | Comparison of RCT and non-RCT on mortality and nursing-home placement | The estimated mortality ORs for RCTs were 0.89 (95% CI 0.76 to 1.04) and for non-RCTs 1.16 (95% CI 0.76 to -1.56); that is similar effects. The estimated nursing-home placement OR for RCTs were | Inconclusive (High risk of bias) |

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|-------------------|--|---|---|--|
| | | | 0.84 (95% CI 0.67 to 1.04) and for non-RCTs 0.41 (95% CI 0.22 to 0.76); that is larger effect for non-RCTs | |
| Heinsman 1996 | 27 RCTs and 14 non-RCTs on the effect of presurgical interventions, and 12 RCTs and 17 non-RCTs on drug use prevention. Both identified in convenience sample. All non-RCTs were non-equivalent control group designs | Comparison of RCTs and non-RCTs on standardised effect size | RCTs had significantly more positive effect sizes for drug use prevention studies ($P = 0.00008$) than non-RCTs. RCTs and non-RCTs had similar effect sizes for presurgical interventions. In a regression model adjusting for potential confounders (all 98 studies, including studies within coaching for Scholastic Aptitude Test performance, ability grouping, presurgical interventions and drug-use prevention) the main effect hovered around the 0.05 significance level, with an un-standardised regression weight that suggested that random assignment adds between 0.05 and 0.10 to the standardized mean difference statistics that would occur in a non-randomised experiment; effect estimate 0.082 (95% CI -0.016 to 0.176) | Larger effect in RCTs than in non-RCTs for one comparison and similar effects for one comparison (High risk of bias) |
| Linde 1999 | 89 placebo controlled trials of homeopathy, identified by systematic search | Comparison of 64 trials that were explicitly randomised with 25 that were not explicitly randomised | In the studies without an explicit statement of randomisation the effect was larger compared to the studies with explicit statements about randomisation (adjusted ratio of odds ratio: 0.64; 95% CI 0.43 to 0.94) | Larger effect in non-RCTs than in RCTs (High risk of bias) |
| Lipsey 1993 | 302 meta-analyses on mental health, work place/organisational and educational interventions identified on a systematic search - all referred to as psychological interventions This review is based upon a subgroup consisting of 74 meta-analyses allowing breakdown of results according to random and non-random allocation | After conversion of the results to mean treatment effect sizes, the following comparisons were included: overall effectiveness of psychological interventions; random versus non-random treatment allocation | Overall effectiveness of psychological interventions showed a mean effect size (MES) of 0.5 ± 0.29 ($N = 302$). 74 meta-analyses allowed further breakdown of results according to random and non-random allocation. No difference in MES was detected (0.46 ± 0.28 , respectively 0.41 ± 0.36). However, in 28% (21 of 74 meta-analyses), the difference in MES between RCTs and non-RCTs within an individual meta-analysis (MES (RCT) - MES (non-RCT)) was larger than 0.2 in both directions. | Inconclusive (High risk of bias) |
| Miller 1989 | 96 studies comparing new surgical interventions with old, published in 1983 and identified in leading surgical journals by a systematic search | 81 randomised controlled trials and 15 non-randomised controlled trials compared on 'treatment gain' (Mann-Whitney statistic). The association between treatment success and study design and the relation between quality score and treatment gains were assessed. | For new therapies on the principal disease no difference in 'treatment gain' was found in non-randomised controlled trials (0.62) compared with RCTs (0.56). For therapies defined as the secondary treatments the 'treatment gain' was similar across non-randomised trials (0.54) and RCTs (0.55). Within RCTs, there was no correlation between quality scores and treatment gains ($P = 0.7$). | Similar effects (High risk of bias) |
| Ottensbacher 1991 | 22 RCTs and 22 non-RCTs on the effectiveness of occupational therapy-interventions, identified by 2 occupational therapy journals | Crude comparison of RCTs versus non-RCTs on the number of statistically significant results, and comparison of effect size estimates (d-index) weighted by sample size. The | Non-RCTs had more often significant test results compared to RCTs ($P < 0.01$). However, the d-index revealed similar effect sizes for both designs | Inconclusive (High risk of bias) |

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|------------------|---|---|---|--|
| | | d-index is used to estimate the difference between 2 groups in terms of their common (average) standard deviation. If $d = 0.30$, then 3/10 of a standard deviation separates the average subject in the 2 groups. | (non-RCTs: 0.36 +/- 0.18; RCTs: 0.35 +/- 0.14). | |
| Ottenbacher 1992 | Sample of 30 RCTs and 30 non-RCTs from a systematic search in NEJM and JAMA across a variety of medical specialties | RCTs were compared with non-RCTs on treatment effects as measured by standardised mean differences | No difference in treatment effect was found between non-RCTs (mean effect size 0.21; mean effect size weighted by sample size 0.18 (95% CI 0.16 to 0.23) and RCTs (mean effect size 0.23; mean effect size weighted by sample size 0.21 (95% CI 0.15 to 0.25)). | Similar effects (High risk of bias) |

Comparison 7. Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions for the same condition

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---------------------------|----------------|---------------------|--------------------|-----------------|
| 1 Undefined | | | Other data | No numeric data |

Analysis 7.1. Comparison 7 Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions for the same condition, Outcome 1 Undefined.

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|---------------|--|--|---|---|
| Chalmers 1983 | 102 controlled trials of the treatment of acute myocardial infarction, identified by a systematic search | Comparison of studies with different allocation schemes (non-random trials, non-concealed controlled trials, and concealed randomised controlled trials) on misdistribution of prognostic variables, frequency of significant outcomes and case-fatality rates | In non-randomised studies, non-concealed controlled trials, and RCTs with concealed allocation, the maldistribution of prognostic factors was 34%, 7% and 3.5% respectively, the frequency of significant outcomes was 58%, 24% and 9% respectively. The case-fatality rate for the control groups was 32%, 23% and 16% and for the treatment groups was 21%, 18% and 16% respectively. | Larger effects in trials with inadequate concealment of allocation (High risk of bias) |

Comparison 8. Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions and conditions

| Outcome or subgroup title | No. of studies | No. of participants | Statistical method | Effect size |
|---------------------------|----------------|---------------------|--------------------|-----------------|
| 1 Undefined | | | Other data | No numeric data |

Analysis 8.1. Comparison 8 Studies of controlled trials with adequate versus inadequate/unclear concealment of allocation across different interventions and conditions, Outcome 1 Undefined.

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|----------------|--|---|---|--|
| Balk 2002 | 276 RCTs from 26 meta-analyses from 4 medical areas, identified by convenience sample | Comparison between 47 RCTs with adequate allocation concealment versus 46 RCTs with inadequate concealment in cardiovascular disease. Comparison between 19 RCTs with adequate concealment versus 37 with inadequate concealment in infectious disease. Comparison between 21 RCTs with adequate concealment versus 39 RCTs with inadequate concealment in paediatrics. Comparison between 21 RCTs with adequate concealment versus 46 RCTs with inadequate concealment | The treatment effect was similar (measured using ratio of odds ratios (ROR)) for RCTs with adequate and inadequate allocation concealment for cardiovascular disease (ROR 1.14 (0.96 to 1.42)), infectious diseases (ROR 0.97 (0.68 to 1.42)), paediatrics (ROR 0.90 (0.58 to 1.28)), surgery (ROR 0.73 (0.36 to 1.24)) For all 4 medical areas combined the treatment effect was similar for RCTs with adequate and inadequate allocation concealment (ROR 1.05 (0.91 to 1.21)) | Similar effect in 4 comparisons (High risk of bias) |
| Clifford 2002 | 100 controlled trials, 20 each from top 5 peer reviewed journals, identified by convenience sample | Comparison of controlled trials with adequate and inadequate/unclear concealment of allocation for outcome measure generated by authors based on direction of the main outcome in the included studies | The generated outcome measure was similar from trials with adequate and inadequate/unclear allocation concealment (RR of favouring new treatment versus other results: 0.81 (95% CI 0.60 to 1.11)) | Similar effects (High risk of bias) |
| Egger 2003 | 314 RCTs from 39 meta-analyses identified by a search in Issue 1 1998 of the Cochrane Database of Systematic Reviews | Comparison of RCTs with adequate and inadequate or unclear concealment of allocation on the ratio of the pooled effect estimate from adequate concealed trials to the pooled effect estimate from the inadequate or unclear concealed trials | Treatment effect estimates were on average 21% more beneficial in the trials with inadequate or unclear allocation concealment (95% CI 11% to 30% more beneficial; $P < 0.001$) | Larger effects in trials with inadequate or unclear allocation concealment (High risk of bias) |
| Kjaergard 2001 | 190 randomised controlled trials from 14 meta-analyses from 8 therapeutic areas, identified by systematic search | Comparison between 68 RCTs with adequate allocation concealment versus 122 RCTs with inadequate allocation concealment on the intervention effect | All trials with inadequate allocation concealment yielded (statistically insignificant) larger treatment effects compared with all trials reporting adequate allocation concealment (ratio of odds ratios: 0.60 95% CI 0.31 to 1.15) | Similar effects or inconclusive (High risk of bias) |
| Linde 1999 | 89 placebo-controlled trials of homeopathy, identified by systematic search | Comparison between 34 controlled trials with adequate allocation concealment versus 55 controlled trials with inadequate allocation concealment on the treatment effect | Similar treatment effect was reported from studies with adequate allocation concealment and inadequate allocation concealment | Similar effects (High risk of bias) |
| Moher 1998 | 127 controlled trials from a randomly selected set of 11 meta-analyses on digestive, circulatory, mental diseases, stroke and infertility from a convenience database of meta-analyses, resp. from the Cochrane Database of Systematic Reviews | Comparison of the impact of controlled trials with unclear/no reporting of allocation concealment, of double-blinding, of random generation versus clear reporting of these features (measured by odds ratio) | The treatment effect was significantly overestimated by 37% in studies lacking allocation concealment (ratio of odds ratios: 0.63 95% CI 0.45 to 0.88). Lack of reporting how double-blinding or randomisation generation was achieved had no influence on treatment effect | Larger effects in trials with inadequate or unclear allocation concealment (High risk of bias) |
| Schulz 1995 | 250 controlled trials from 33 meta-analyses from the Cochrane Pregnancy and Childbirth Group database | Association between methodological features of controlled trials (allocation concealment, double-blinding and follow up) and the treatment effect (measured by the odds ratio) | The treatment effect was 41% larger in controlled trials with inadequate allocation concealment and 30% larger in controlled trials with unclear adequacy of allocation concealment compared with those with adequate allocation concealment ($P < 0.001$) after adjustment for other methodological features. Studies with no double-blinding had | Larger effects in trials with inadequate or unclear allocation concealment (High risk of bias) |

| Study | SAMPLE | Undefined COMPARISON | MAIN RESULTS | DIRECTION OF BIAS |
|------------|---|--|---|--|
| Shang 2005 | 110 controlled trials in homeopathy and 110 controlled trials in conventional medicine, identified by systematic search | Comparison between 49 controlled trials with adequate allocation concealment versus 61 controlled trials with inadequate allocation concealment for homeopathy trials. Comparison between 21 controlled trials with adequate allocation concealment versus 89 controlled trials with inadequate allocation concealment of conventional medicine trials | a treatment effect 17% larger than double-blinded studies ($P = 0.01$). Lack of complete follow up had no influence on the treatment effect (7% smaller, $P = 0.32$). Using univariate meta-regression the treatment effects were similar for controlled trials with adequate and inadequate allocation concealment for both homeopathy trials (ROR 0.78 (0.57 to 1.07)) and for conventional medicine (ROR 0.76 (0.48 to 1.16)) In multivariable analyses, the SE of the log odds ratio was the dominant variable in both groups. Coefficients of other variables, including study quality, were attenuated and became non-significant. | Similar effects in both comparisons (High risk of bias) |

ADDITIONAL TABLES

Table 1. Critical appraisal of the risk of bias

| Study | Sample | Confounding control? | Reproducibility | Outcome measure | Overall judgement |
|----------------|---------------------------|--|---|---------------------------|-------------------|
| | Complete sample of trials | 1. Double-blinding 2. Complete follow up 3. Clinical differences in participants 4. Clinical differences in interventions | 1. Explicit criteria used to select the trials 2. Agreement regarding selection of trials 3. Two or more assessed quality and conducted data extraction | Similar outcomes measured | |
| Balk 2002 | No | 1 No, 2 No, 3 Yes, 4 Yes | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |
| Carroll 1996 | Yes | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |
| Chalmers 1983 | Yes | 1 Yes, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Unclear | Yes | High risk of bias |
| Clifford 2002 | No | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Yes | Unclear | High risk of bias |
| Colditz 1989 | No | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |
| Egger 2003 | No | 1 No, 2 No, 3 Yes, 4 Yes | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |
| Hedrick 1989 | Yes | 1 Yes, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Unclear | Yes | High risk of bias |
| Heinsman 1996 | No | 1 Unclear, 2 Yes, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Unclear | No | High risk of bias |
| Kjaergard 2001 | Yes | 1 Yes, 2 No, 3 Yes, 4 Yes | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |

Table 1. Critical appraisal of the risk of bias (Continued)

| | | | | | |
|----------------------------------|-----|-------------------------------|---------------------------------|---------|-------------------|
| Linde 1999 | Yes | 1 Yes, 2 Yes, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |
| Lipsey 1993 | Yes | 1 No, 2 No, 3 No, 4 No | 1 Unclear, 2 Unclear, 3 Unclear | No | High risk of bias |
| Miller 1989 | No | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 Yes, 3 Unclear | No | High risk of bias |
| Moher 1998 | No | 1 Unclear, 2 No, 3 Yes, 4 Yes | 1 Yes, 2 Yes, 3 Yes | No | High risk of bias |
| Ottenbacher 1991 | No | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 No, 3 Yes | No | High risk of bias |
| Ottenbacher 1992 | No | 1 No, 2 No, 3 No, 4 No | 1 Yes, 2 Yes, 3 Yes | No | High risk of bias |
| Schulz 1995 | No | 1 Yes, 1 Yes, 2 Yes, 3 Yes | 1 No, 2 Unclear, 3 No | Unclear | High risk of bias |
| Shadish 1996 | No | 1 No, 2 Yes, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Unclear | No | High risk of bias |
| Shang 2005 | No | 1 Yes, 2 No, 3 No, 4 No | 1 Yes, 2 Unclear, 3 Yes | No | High risk of bias |

APPENDICES

Appendix 1. The Cochrane Methodology Register

In the [CMRG Module](#) CMR is described in the following way: "The broad intention of CMR is to include all published reports of empirical studies of methods used in reviews, as well as methodological studies that are directly relevant to doing a review, such as empirical studies of the association between research methods and bias in randomised controlled trials. Details of ongoing methodological research are also included. Books, conference proceedings and special journal issues devoted to the topic of systematic reviews and meta-analysis have been included, but in general their constituent chapters and articles have not been listed separately. Articles introducing systematic reviews and meta-analysis to a wide audience have been included, as well as others addressing specific issues of relevance; but a number of general articles directed at specialist audiences have not been listed. The content of the Cochrane Methodology Register (CMR) is being constantly expanded upon as a direct result of an extensive handsearching programme and the development of a series of search strategies in MEDLINE and EMBASE to identify relevant reports."

Appendix 2. Search strategy for the previous versions of this review

[Kunz 2002](#):

Studies were identified using the Cochrane Methodology Register, bibliographies, MEDLINE, SciSearch, handsearching, personal communication with methodologists and the reference lists of relevant articles up to August 2000. Exploratory handsearching of methodological journals (*Controlled Clinical Trials*, *Statistics in Medicine*, *Journal of Clinical Epidemiology*) for four volumes (1970, 1980, 1990 and 1995) was not productive. These journals, and many others, have now been handsearched in full for the Cochrane Methodology Register. Repeated efforts have been undertaken to develop an efficient electronic search strategy using MEDLINE since 1994. Early attempts were not efficient due to poor indexing of methodological studies. Since 1999 MEDLINE searches have been more successful, particularly by searching for "Related Articles" in PubMed using seven key articles ([Chalmers 1977](#); [Colditz 1989](#); [Emerson 1990](#); [Kunz 1998](#); [Ottenbacher 1992](#); [Sacks 1982](#); [Schulz 1995](#)). This was supplemented with a search strategy using the following combinations of MeSH-terms:

- [Random Allocation OR Randomised Controlled Trial (exp)] AND Bias (epidemiology)
- [Random Allocation OR Randomised Controlled Trial (exp)] AND research /cl,mt,sn,st,td
- [Random Allocation OR Randomised Controlled Trial (exp)] /cl,mt,sn,st,td,ut AND Double Blind Method
- [Random Allocation OR Randomised Controlled Trial (exp)] /cl,mt,sn,st,td,ut AND Clinical Trials /cl,mt,sn,st,td,ut
- Randomised Controlled Trial (exp) AND Selection Bias
- Randomised Controlled Trial (exp) AND Follow-Up Studies
- Randomised Controlled Trial /mt,sn,st,ut AND Follow-Up Studies

Randomisation to protect against selection bias in healthcare trials (Review)

Copyright © 2015 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

SciSearch was searched for articles that cited the following articles: [Chalmers 1977](#); [Colditz 1989](#); [Emerson 1990](#); [Kunz 1998](#); [Miller 1989](#); [Ottenbacher 1992](#); [Sacks 1982](#); [Schulz 1995](#).

A large proportion of studies were assembled through personal contacts with methodologists and from bibliographies and reference lists.

FEEDBACK

Justification of the unpredictability paradox, 11 December 2014

Summary

Submitted by Jeremy Howick

Thank you for this great review. I believe the so-called "unpredictability paradox" requires further justification for five reasons:

1. Invoking the unpredictability paradox makes the conclusions of your review unfalsifiable and perhaps unscientific (from a Popperian perspective). If it turned out that randomized trials had average significantly different average results from nonrandomized studies, you could have pooled the results and concluded that adequately randomized trials were better. In fact, adequate randomization did not yield statistically significant different average results, so you drew the very same conclusion that they could have had the data indicated differences between adequately and inadequately randomized trials. Drawing the same conclusion from conflicting evidence allows us to make assertions that do not take empirical evidence into account, which is unscientific in the absence of further justification.
2. Appeal to the unpredictability paradox reveals an inconsistent approach with regards to pooling data in Cochrane Review methodology. When we pooled the results from your review we found no statistically significant difference between randomized and non-randomized trials (standardized mean difference = -0.17, 95% CI = -0.64 to 0.29; P = 0.47). The decision to pool appears to justify the inference to the conclusion that adequate randomization was not a methodological benefit easy to draw. (As an aside, the problem is not whether to pool itself, but rather the inference from the unpooled result to the conclusion of a difference in an unpredictable direction.) The Cochrane Handbook recommends not pooling highly heterogeneous results, yet the results of your review were remarkably consistent in terms of effect direction, with all but one included study revealing no statistically significant difference. Moreover Cochrane Reviews conducted by the same review group (see reference 8, below) have pooled results with substantially higher heterogeneity ($I^2 = 87%$) (see reference 8, below). The inconsistency in Cochrane methodology was further highlighted in a recent similar systematic review (reference 9, below) of randomized versus observational studies. The authors of the latter review found similarly heterogeneous results, but decided to pool and concluded that randomized and non-randomized studies were not qualitatively different. Had they adopted the same strategy as you had, they could have chosen not to pool, postulated the "unpredictability paradox" and concluded that randomized trials have different results from observational studies, but in an unpredictable direction.
3. The unpredictability paradox has not been used or replicated independently. If proposing that the unpredictability paradox is justified, one would expect independent research to use and validate it. This has not been done.
4. Invoking the unpredictability paradox discourages researchers from investigating the conditions under which randomization over- and under-exaggerates apparent treatment benefits. If, indeed, adequate randomization makes a difference, it would be interesting to know what made adequate randomization increase effect size and what made it decrease effect size. Proposing the unpredictability paradox as an explanation for the effect of adequate randomization suggests that there is nothing more fundamental to be learned about the conditions under which adequate randomization makes a difference, precisely because it is unpredictable. This approach therefore arguably stifles future research in the area.
5. If it turns out that adequate randomization is not a powerful protection against bias, it could obscure the relative importance of allocation concealment and blinding which may be more important.

I am not implying that inadequate randomization is acceptable. It is self-evident that inadequate randomization is a sign of sloppy research, and also makes allocation concealment and blinding more difficult. Allocation concealment and blinding, in turn, have been shown empirically to reduce bias in many cases. It follows that, when results from adequately randomized studies and inadequately randomized studies (or observational studies) differ, the results of the adequately randomized trial is likely to be closer to the truth (all other things being equal).

However when investigating the potential benefits of randomization, adequate methods must be used.

References

1. Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995, 273:408–412.
2. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP: Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998, 352:609–613.

3. Kjaergard LL, Villumsen J, Gluud C: Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Inter Med* 2001, 135:982–989.
4. Jüni P, Altman DG, Egger M: Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001, 323:42–46.
5. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD: Randomization to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011, 4:MR000012.
6. Popper KR: *The Logic of Scientific Discovery*. London: Hutchinson; 1968.
7. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
8. Hróbjartsson A, Gøtzsche PC: Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2010, 1:CD003974.
9. Anglemeyer A, Horvath HT, Bero L: Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014, 4:MR000034.
10. Kunz R, Oxman AD: The unpredictability paradox: review of empirical comparisons of randomized and non-randomised clinical trials. *BMJ* 1998, 317:1185–1190.
11. Howick J: *The Philosophy of Evidence-Based Medicine*. Chichester: Wiley Blackwell & BMJ Books; 2011.
12. Savović J, Jones HE, Altman DG, Harris R, Jüni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J: Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Inter Med* 2012, 157:429–438.
13. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, GRADE Working Group: GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008, 336:924–926.
14. OCEBM Levels of Evidence Working Group: Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence. [<http://www.cebm.net/index.aspx?o=5653>].

Reply

Thank you for your interest in our review.

First, it is important to note that we did not use the “unpredictability paradox” as a way of defending our decision not to pool or a way to explain the observed results, but rather as a reflection on the variation in direction and size of the observed impact of adequate randomization on effect sizes. The reference to the “unpredictability paradox” is not essential to the key messages of this review and could be removed or reworded without changing the conclusions. Moreover, the “unpredictability paradox” is falsifiable. Finding a way of predicting the extent to which non-randomised trials are biased would falsify it. For example, finding no important differences between the findings of studies with and without adequate randomization would be a falsification, as would finding a consistent direction of bias across studies.

We disagree with you when it comes to the interpretation of the pooled standardized mean difference (SMD) you found based on some of the results presented in the review. Based on an SMD of -0.17 with a 95% confidence interval from -0.64 to 0.29 we would not draw the conclusion that no differences exist between randomized and non-randomized trials when it comes to effect sizes. It is correct that the point estimate suggests a small effect size, but on the other hand, a small to medium effect size in either direction cannot be ruled out. Lack of statistically significant differences does not imply that a difference does not in fact exist. We can only conclude that a difference most likely does not exist if both limits of the confidence interval are within a pre-specified difference from the point of no difference. Our conclusion would be “based on the available evidence it is not possible to draw any firm conclusions when it comes to the direction and magnitude of the impact of adequate randomization on effect sizes.” Thus, the conclusions of the review would not be changed by pooling the results from the included methodological studies in a meta-analysis. In addition, we did not conclude “that randomized trials have different results from observational studies.” We concluded that “most often non-randomised and randomised trials appeared to have similar results.”

In our opinion the choice whether to pool data across included studies should not be based on the observed statistical heterogeneity, but rather on an assessment of the degree to which an pooled estimate of the (average) effect would be meaningful. Basing this choice on the observed statistical heterogeneity would make the analysis strategy data driven. This would be unfortunate in light of the low power of both the Chi squared test and the I-squared statistic (with associated uncertainty interval) when based on few studies (1). Consequently, we feel that that choice whether to meta-analyse or not should be guided by the same recommendations as for the choice between fixed and random effects meta-analysis. The recommendation on this from the Cochrane Handbook (2) section 9.5.4 is “The choice between a fixed-effects and a random-effects meta-analysis should never be made on the basis of a statistical test for heterogeneity”.

We based the choice not to perform any meta-analysis on the risk of confounding by intervention and condition (in addition to outcome measure). We believe that the difference in effect sizes between randomized controlled trials and non-randomized controlled trials might depend on the intervention, condition, and outcome (as also suggested by peer reviewers for previous versions of this review). We therefore introduced a risk of bias item to assess the extent to which the included methodological studies took the possibility of confounding by intervention or condition into account in the statistical analysis. We scored methodological studies including primary trials with the same intervention and condition as having a low risk of bias for this item. We scored methodological studies including primary trials with different interventions and conditions as having a low risk of bias if differences in interventions and conditions were taken into account in the analysis, and as having a high risk of bias otherwise (provided sufficient information to make a proper assessment). As we rated failure to take account for differences in interventions and conditions as a high risk of bias, we did not find it appropriate to perform any pooling of results across methodological studies with different interventions and conditions (as specified in the Methods section of the review (3)).

We agree that there is a need for further research regarding the role of randomisation in protecting against selection bias, as we noted under “Implications for methodological research” and we specifically noted the need to investigate the role of varying baseline risk, heterogeneity and study quality using individual trial analysis (analogous to individual patient data meta-analysis). This would build on the principles used for statistical analysis in the methodological study by Savović and colleagues (4) and other studies using this approach.

References

- Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J: Assessing heterogeneity in Meta-analysis: Q statistic or I² index?. Center for Health, Intervention, and Prevention (CHIP); 2006. Available from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1019&context=chip_docs
- Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database of Systematic Reviews 2011, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.pub3.
- Savović J, Jones HE, Altman DG, Harris R, Jüni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J: Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Inter Med* 2012, 157:429–438.

Contributors

This response was prepared by the authors of the review (21 January 2015).

WHAT'S NEW

| Date | Event | Description |
|--------------|--------------------------------|--|
| 1 April 2015 | Feedback has been incorporated | Feedback and the authors' response have been incorporated. |

HISTORY

Protocol first published: Issue 3, 2002
 Review first published: Issue 3, 2002

| Date | Event | Description |
|------------------|--|--|
| 1 March 2011 | New citation required and conclusions have changed | Review updated. |
| 21 December 2009 | New search has been performed | New search performed. We have elaborated the comparisons in a way that results in eight comparisons rather than three. The comparisons should be more intuitive and highlight where more research is needed. We have also elaborated and refined the inclusion criteria in order to exclude comparisons of observational studies with randomised trials. |

| Date | Event | Description |
|------------------|--|--|
| | | <p>We identified seven new studies and excluded 22 previously included studies (10 studies had a comparison of interest incidental to the main aim of the study; three studies looked at the comparison randomised trials versus historically controlled studies; two studied the comparison randomised trials versus observational studies and seven studies compared high versus low quality).</p> <p>We have incorporated feedback on the previous draft (August 2006).</p> |
| 27 December 2007 | Amended | Converted to new review format. |
| 20 February 2007 | New citation required and conclusions have changed | Substantive amendment. |
| 23 August 2006 | New search has been performed | This review has been updated from a previously published Cochrane Review (Kunz 2002). |

CONTRIBUTIONS OF AUTHORS

GEV prepared the first draft of this update. JOJ prepared the revised draft and finalised this manuscript. GEV, AT, RK, EAA, JOJ, HJS, MB, AN and EP screened references, assessed the relevance of retrieved studies, assessed the methodological quality of included studies and extracted data. GEV, AT, RK, EAA, JOJ, HJS, MB and AN contributed to the manuscript for this update. GEV, SP and JOJ prepared the 'Risk of bias' tables. ADO contributed to the manuscript. RK prepared the first draft of the protocol and earlier versions of this review and collected data from included studies; ADO contributed to the preparation of the protocol and the final manuscript and helped assess the relevance and methodological quality of retrieved reports. GEV and ADO checked the collected data against the original reports and contributed to the manuscript.

DECLARATIONS OF INTEREST

JOJ is statistician with the Methodology Review Group.

SOURCES OF SUPPORT

Internal sources

- Norwegian Knowledge Centre for the Health Services, Norway.
- Italian National Cancer Centre Regina Elena, Rome, Italy.
- Department of Medicine, State University of New York at Buffalo, USA.
- Swiss National Science Foundation, Switzerland.

External sources

- Department of Health, UK.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We have elaborated the comparisons in a way that results in eight comparisons rather than three. The comparisons should be more intuitive and highlight where more research is needed. We have also elaborated and refined the inclusion criteria in order to exclude comparisons of observational studies with randomised trials.

INDEX TERMS

Medical Subject Headings (MeSH)

*Clinical Trials as Topic [methods] [standards] [statistics & numerical data]; *Random Allocation; *Selection Bias; Controlled Clinical Trials as Topic [methods] [standards] [statistics & numerical data]; Randomized Controlled Trials as Topic [methods] [standards] [statistics & numerical data]; Treatment Outcome