How objective are systematic reviews? Differences between reviews on complementary medicine

Klaus Linde MD Stefan N Willich MD PhD[1]

J R Soc Med 2003;**96**:17–22

SUMMARY

Systematic reviews are considered the most reliable tool to summarize existing evidence. To determine whether reviews that address the same questions can produce different answers we examined systematic reviews of herbal medicine, homeopathy, and acupuncture taken from a previously established database. Information on literature searching, inclusion criteria, selection process, quality assessment, data extraction, methods to summarize primary studies, number of included studies, results and conclusions was compared qualitatively.

Seventeen topics (eight on acupuncture, six on herbal medicines, three on homeopathy) had been addressed by 2–5 systematic reviews each. The number of primary studies in the reviews varied greatly within most topics. The most obvious reason for discrepancies between the samples was different inclusion criteria (in thirteen topics). Methods of literature searching may have contributed with some topics but the equivalence of the searches was difficult to assess. Differences were frequently observed in other methodological aspects, in results and in conclusions.

This analysis shows that, at least in the three areas examined, systematic reviews often differ considerably. Readers should be aware that apparently minor decisions in the review process can have major impact.

INTRODUCTION

Systematic reviews and meta-analyses are regarded as the best methods to summarize evidence on the effectiveness of healthcare interventions[1,2]. Systematic methods are designed to avoid biases and make results and conclusions as objective as possible. However, systematic reviews are retrospective and strongly depend on the quality of the primary material. In the review process decisions have to be taken that may influence the findings. Finally, unless the results are very clearcut, reviewers with different prejudices about the hypothesis under investigation may draw different conclusions from the same data. Several articles reporting examples of discordant systematic reviews have been published[3–7] but we have found no empirical studies on how often and why discrepancies occur. Within the framework of a project for collecting and analysing systematic reviews of clinical trials of herbal medicine, homeopathy and acupuncture performed for the Cochrane Collaboration's complementary medicine field[8–10] we compared reviews addressing the same topic.

Centre for Complementary Medicine Research, Department of Internal Medicine II, Technische Universität, Kaiserstrasse 9, 80801 München; [1]Institute for Social Medicine & Epidemiology, Charité, Humboldt-University, 10098 Berlin, Germany

Correspondence to: Dr K Linde

E-mail: Klaus.Linde@lrz.tu-muenchen.de

METHODS

Systematic reviews of clinical trials of herbal medicines, homeopathy and acupuncture published between the years 1989 and 2001 addressing the same topic were identified from the database. To be included, reviews had to explicitly describe inclusion and exclusion criteria, the methods used to search the literature, the methods used to assess study quality and the methods for summarizing results when the review included a meta-analysis. Sets of reviews were judged to address the same topic if they were on the same intervention for the same condition and if they covered the same comparisons. When the focus of one review was broader than in another (for example, back pain in one, low back pain in another) the reviews were included if the subgroup of studies in the broader review could be clearly separated for comparison. Reviews within a review set had to have been published within a period of 4 previous years. One assessor screened all systematic reviews included in the database and selected those which addressed broadly similar questions (for example, all reviews of garlic for cardiovascular risk factors). All reviews identified at the screening step were then checked in detail for whether they addressed the same questions. In case of uncertainty a second assessor was involved. For each review the following details were extracted into a spreadsheet: literature search (databases searched, other search methods used), inclusion criteria (concerning patients, experimental and control

interventions, outcomes, study design, language, other), selection process (whether described or not, number of studies at different selection levels), data extraction, quality assessment methods, methods to summarize primary studies, number of included studies, results and methodological quality of primary studies as assessed by the reviewers, and conclusions drawn. To check whether the same primary studies on a given topic were included and to investigate the influence of the date of publication, all studies included by any of the reviews were entered into a list. The only quantitative outcome criteria were the number of included primary studies and the overlap of included primary studies published at least one year before the oldest review. All other analyses were qualitative.

FINDINGS

Among a total of seventeen review sets consisting of 2–5 overviews addressing the same topics and meeting the inclusion criteria (Table 1), eight were on acupuncture, six on herbal medicines and three on homeopathy. The total number of included reviews was 38[11–48]; three acupuncture reviews[14,18,21] contributed to two review sets, since they covered more than one topic. The sample of primary studies varied by more than 25% in fifteen review sets, and by more than 50% in ten. In just one review set (P6-acupuncture stimulation for morning sickness) the age of the review and the resulting availability of trials explained major differences.

The most common reason for discrepancies regarding the sample of included studies was differences in inclusion criteria. This is exemplified by the reviews on hypericum extracts for depression. All these reviews aimed to assess whether hypericum extracts are more effective than placebo or similar in efficacy to standard antidepressants. The number of primary studies varied between 2 and 17 for placebo-controlled trials (with older reviews including more studies) and between 3 and 10 for trials against

Table 1 **Overview on the review sets included in the analysis**

Intervention and condition	*Control group interventions*	*Number of studies included (total included in any of the reviews)* —review publication years*	*Main reasons for discrepant numbers*
Acupuncture			
Acupuncture for chronic pain	Sham, other treatment	14 [Ref. 11], 46 [Ref. 12] (48)—1989–1990	S,IC
Acupuncture for neck pain	Sham, other treatment	11 [Ref. 13], 5 [Ref. 14] (11)—1999–2000	S (IC)
Acupuncture for low back pain	Sham, no/other treatment	10 [Ref. 15], 11 [Ref. 16], 8 [Ref. 14] (11) —1998–2000	IC, S
Acupuncture for tension-type headache	Sham, other treatment	7 [Ref. 17], 8 [Ref. 18], 6 [Ref. 19] (9)— 1999–2000	IC
Acupuncture for migraine prophylaxis	Sham, no/other treatment	6 [Ref. 20], 16 [Ref. 18] (16)—1999	IC
P6-stimulation for postoperative nausea	Sham, no/other treatment	18 [Ref. 21], 19 [Ref. 22] (24)—1996–1999	IC, PD
P6-stimulation for morning sickness	Sham, no treatment	7 [Ref. 21], 4 [Ref. 23], 7 [Ref. 24] (8)— 1996–1999	IC
Acupuncture for tinnitus	Sham, other treatment	6 [Ref. 25], 6 [Ref. 26] (7)—1999–2000	—
Herbal medicine			
Ginkgo for intermittent claudication	Placebo	5 [Ref. 27], 8 [Ref. 28] (9)—2000	IC, S
Hypericum for depression	Placebo, standard	14 [Ref. 29, 27 [Ref. 30], 6 [Ref. 31], 14 [Ref. 32], 8 [Ref. 33] (29)—1997–2000	IC (PD)
Garlic for cholesterol lowering			
new review set	Placebo	13 [Ref. 34], 37 [Ref. 35] (37)—2000	IC, S
old review set	Placebo	5 [Ref. 36], 13 [Ref. 37] (13)—1993–1994	IC, S
Echinacea for colds	Placebo	13 [Ref. 38], 13 [Ref. 39] (15)—1999	—
Peppermint oil for irritable bowel syndrome	Placebo	7 [Ref. 40], 3 [Ref. 41] (8)—1998–2000	S, SE
Homeopathy			
All homeopathy for any disease	Placebo	89 [Ref. 42], 36 [Ref. 43], 16 [Ref. 44] (97)— 1997–2000	IC
All homeopathy for any disease	Placebo, other treatment	40 [Ref. 45], 105 [Ref. 46] (405)—1990–1991	IC
Arnica for trauma	Placebo	35 [Ref. 47], 8 [Ref. 48] (35)—1998	IC, SE

*Reviews ordered according to ascending publication year
S=Search; IC=inclusion and exclusion criteria; PD=publication date; SE=application of inclusion criteria in the selection process

standard antidepressants. Table 2 shows variations in the inclusion criteria between the five reviews, and for the general reader it is almost impossible to know which differences are relevant. For example, the restriction to trials published in English in the review by Gaster[33] explains the exclusion of 12 of the 17 placebo-controlled trials included in the older review by Linde *et al.*[30]; the restriction to mono-preparations explains only one exclusion; and the restriction to double-blind trials had no consequences at all. The main reasons for exclusion of available randomized trials of hypericum in depressed patients are printed in italics in Table 2.

The comprehensiveness of the literature searches was very difficult to assess. Searches in the database Medline were sometimes described in sufficient detail to allow a comparison. However, Medline covers only a small minority of complementary medicine journals and almost all reviewers searched additional sources. In a published paper, to describe these searches in a manner that will allow replication is almost impossible. The comprehensiveness of literature searches could therefore be evaluated only indirectly, by comparing the sample of included studies in a single review with the total sample of studies in any of the reviews, with exclusions taken into account. Obvious relevant differences in comprehensiveness existed in seven review sets (see Table 1). However, there were examples of reviews with quite different search strategies coming up with almost identical study samples (for example, the Echinacea reviews[38,39]).

Although the methods for quality assessment of primary studies in the reviews differed considerably (a wide variety of scores and checklists), major disagreements about overall quality were rare. A striking exception is the three reviews including trials of acupuncture for low back pain. Only one of these reviews is explicitly restricted to low back pain[16], one is on back pain[15] and one on back and neck pain[14]. However, most of the primary studies in the latter two are also on low back pain. Ernst and White[15] described the methodological quality of the back pain studies reviewed as 'good in the majority of studies'; van Tulder *et al.*[16] concluded for the low back pain trials that 'methodological quality was . . . extremely poor' and Smith *et al.*[14] judged that the 'majority of trials were of poor quality'.

Because of the heterogeneity of the primary studies, the variability of outcome measures and insufficient reporting, only 20 reviews included a quantitative meta-analysis. In six review sets more than one review included a meta-analysis. While the reported effect sizes differed to some extent, this was mainly because of differences in the study samples. Only in the 3 reviews addressing the question whether homeopathy is any different from placebo did the meta-analytic methods differ fundamentally and this, together with differences in the study samples, led to discrepant conclusions (Table 3).

Instead of or in addition to meta-analysis, results of primary studies were summarized descriptively or in vote counts. As the vote-counting systems often differed slightly, formal analysis of agreement proved difficult. In the case of trials of acupuncture for low back pain the discrepancies were large (Table 4).

There was good agreement in almost all review sets that further research on the respective topic is needed; only one review explicitly states that new studies on homeopathy would be unlikely to end the controversy on this therapy[45]. Strong disagreements about the available evidence were seen in reviews of acupuncture for low back pain (as we have noted earlier) and of homeopathy versus placebo; more subtle differences in conclusions were common, and seemed to depend more on the prior beliefs of the reviewers than on the data.

Table 2 **Inclusion criteria in six systematic reviews of clinical trials of hypericum extracts versus placebo or standard antidepressants for depression**

Review	*Patients/condition*	*Hypericum preparations*	*Design*	*Other*	*Versus placebo/ versus standard*
Volz 1997 (Ref. 29)	Depressed patients	Mono-preparations	Double-blind clinical trials	—	11/3
Linde 1998 (Ref. 30)	Depressive disorders	Mono-preparations and combinations*	RCTs	—	17/10
Kim 1999 (Ref. 31)	*Depressive disorder acc. to ICD10/DSM*	Mono-preparations	Blinded controlled trials	Outcome Hamilton Rating Scale	2/4
Williams 2000 (Ref. 32)	Depressive disorders	Mono-preparations and combinations*	RCTs	*At least 6 weeks' duration*	8/6
Gaster 2000 (Ref. 33)	Depressive disorders	Mono-preparations	Double-blind RCTs	*Available in English*	4/4

Italics indicate the criteria most likely to have reduced number of trials included
*Combinations with other herbal extracts
RCT=Randomized clinical trial

Table 3 **Inclusion criteria, number of included trials, methods for summarizing study results and main result in three meta-analyses of placebo-controlled trials of homeopathy**

	Linde* et al. *(Ref. 42)	***Walach (Ref. 43)***	***Cucherat* et al. *(Ref. 44)***
Main differences regarding inclusion criteria	All explicitly randomized and/or double-blind trials with data suitable for meta-analysis	Only randomized trials published in non-homeopathic peer-reviewed journals listed in Medline or Embase	Only randomized trials with a predefined main outcome measure
Number of trials checked for eligibility	119	Unclear	118
Number of trials included in meta-analysis	89	36	16
Approach to summarizing effect sizes	Combined odds ratio	Standardized effect size	Combination of *P* values
Meta-analytic model chosen	Random effects	Random effects	Weighted sum of Zs
Main result	Significant: OR=2.45 (95% CI 2.05 to 2.93)	Not significant: g=0.259 (95% CI −0.319 to 0.837)	Significant: combined *P*=0.000036

OR=Odds ratio; CI=confidence interval

Table 4 **Vote counts for low back pain trials included both by van Tulder *et al.* (Ref. 16) and by Smith *et al.* (Ref. 14)**

	Author conclusion		***Reviewer conclusion***	
Trial	***van Tulder* et al.**	***Smith* et al.**	***van Tulder* et al.**	***Smith* et al.**
Coan *et al*.	Positive	Positive	Unclear	Positive
Duplan *et al*.	Positive	Positive	Positive	Negative
Edelist *et al*.	Neutral	Negative	Neutral	Negative
Garvey *et al*.	Positive	Negative	Neutral	Negative
Gunn *et al*.	Positive	Positive	Neutral	Positive
Lehmann *et al*.	Positive	Negative	Neutral	Negative
Lopacz and Gralewski	Neutral	Negative	Neutral	Positive
Mendelson *et al*.	Neutral	Negative	Unclear	Negative

van Tulder *et al*. used a 3-step vote count (positive, neutral, negative) with an unclear option while Smith *et al*. voted trials as either positive or negative. Both reviewers tried to categorize the conclusions of the authors of the primary studies and presented their own conclusion (reviewer conclusion)

DISCUSSION

This qualitative analysis indicates that systematic review of clinical trials of herbal medicine, homeopathy and acupuncture can greatly differ in their conclusions. We were surprised by the number and scale of the discrepancies. In large part, we believe, they are traceable to the multiple decisions taken during the planning, performance and interpretation.

A limitation of our study is that the extractions and assessments were done mainly by a single investigator. A crucial issue is also whether a set of reviews is considered to address the same topic. Researchers doing systematic reviews and general readers probably have different ideas about this. For researchers it will be clear that subtle differences in inclusion criteria mean that slightly different questions are answered. The general reader, however, reads a systematic review to learn whether there is evidence that, for example, 'hypericum works for depression'. This reader will not know that the words 'attempting to retrieve all relevant English-language articles' will exclude most of the relevant work.

There is evidence that well-conducted clinical trials yield the least promising results[49]. Could it be that differences in quality explain the discrepancies between systematic reviews. Jadad and McQuay did find that less rigorous reviews more often had positive conclusions[6], but Katerndahl and Lawler[4] and Assendelft *et al*.[50] reached the opposite conclusion. Jadad *et al*., looking at asthma reviews[51], found no differences related to quality. Nor, in our review samples, do differences in the quality of reviews seem to contribute to the discrepancies. Undoubtedly, readers should check whether systematic reviews fulfil

common quality criteria, but often there is no right or wrong answer on what should be included. With hypericum for depression, for example, there are good arguments for all three strategies that were used—to include all trials[30], only those that comply with up-to-date diagnostic criteria[31] or those with observation periods of at least 6 weeks. Jadad *et al*.[52] provide some guidance on how to cope with discordant quantitative meta-analyses, but the reader must be in possession of all the discordant reviews, as well as the time and specialized knowledge to decide which methods were most appropriate. We have looked only at reviews in complementary medicine but we suspect that the problem applies also to conventional medicine[3–7].

What are the implications of our findings? They must not be misinterpreted as an argument for returning to unsystematic reviews, in which the discrepancies tend to be greater[50,53]. In the past ten years the methodology of systematic reviews has developed considerably, and recent guidelines[54] should improve the reporting in future years. Even so, caution will still be needed in their interpretation. Discrepancies between high-quality reviews will always be possible.

Acknowledgment The work of KL was partly funded by a grant from the Karl and Veronica Carstens Foundation, Essen, Germany.

REFERENCES

1 Chalmers I, Altman DG, eds. *Systematic Reviews*. London: BMJ Publishing Group, 1995

2 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;**126**:376–80

3 Hopayian K. The need for caution in interpreting high quality systematic reviews. *BMJ* 2001;**323**:681–4

4 Katerndahl DA, Lawler WR. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: a meta-analysis. *Am J Epidemiol* 1999;**129**:429–41

5 Pettigrew M, Kennedy SC. Detecting the effects of thromboprophylaxis: the case of the rogue reviews. *BMJ* 1997;**315**: 665–8

6 Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol* 1996;**49**:235–43

7 Cook DJ, Reeve BK, Guyatt GH, Griffith LF, Heyland DK, Tryba N. Stress ulcer prophylaxis in the critically ill: resolving discordant meta-analyses. *JAMA* 1996;**275**:308–14

8 Linde K, Vickers A, Hondras M, *et al*. Systematic reviews of complementary therapies—an annotated bibliography. Part 1: Acupuncture. *BMC Complement Alt Med* 2001;**1**:3

9 Linde K, ter Riet G, Hondras M, Vickers A, Saller R, Melchart D. Systematic review of complementary therapies—an annotated bibliography. Part 2: Herbal medicine. *BMC Complement Alt Med* 2001;**1**:5

10 Linde K, Hondras M, Vickers A, ter Riet G, Melchart D. Systematic review of complementary therapies—an annotated bibliography. Part 3: Homeopathy. *BMC Complement Alt Med* 2001;**1**:4

11 Patel MS, Gutzwiller F, Paccaud F, Marazzi A. A meta-analysis of acupuncture for chronic pain. *Int J Epidemiol* 1989;**18**:900–6

12 ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990;**43**:1191–9

13 White AR, Ernst E. A systematic review of randomized controlled trials of acupuncture for neck pain. *Rheumatology* 1999;**38**:143–7

14 Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. *Pain* 2000;**86**:119–32

15 Ernst E, White AR. Acupuncture for back pain. A meta-analysis of randomized controlled trials. *Arch Intern Med* 1998;**158**:2235–41

16 Tulder MW van, Cherkin DC, Berman B, Lau L, Koes BW. Acupuncture for low back pain (Cochrane Review). In: *The Cochrane Library, Issue 1*, 2000. Oxford: Update Software

17 Vernon H, McDermaid CS, Hagino C. Systematic review of randomized clinical trials of complementary/alternative therapies in the treatment of tension-type and cervicogenic headache. *Complement Ther Med* 1999;**7**:142–55

18 Melchart D, Linde K, Fischer P, *et al*. Acupuncture for recurrent headaches: a systematic review of randomized controlled trials. *Cephalalgia* 1999;**19**:776–86

19 McCrory D, Penzien DB, Gray RN, Hasselblad V. Behavioral and physical treatments for tension-type and cervicogenic headache. Prepared for the Foundation for Chiropractic Education and Research, 2000. [www.fcer.org]

20 Goslin RE, Gray RN, McCrory DC, Penzien D, Rains J, Hasselblad V. Behavioral and physical treatments for migraine headache. Technical review 2.2. Prepared for the Agency for Health Care Policy and Research, 1999. [www.clinpol.mc.duke.edu]

21 Vickers AJ. Can acupuncture have specific effects on health? A systematic review of acupuncture antiemesis trials. *J Roy Soc Med* 1996; **89**:303–11

22 Lee A, Done ML. The use of nonpharmacologic techniques to prevent postoperative nausea and vomiting: a meta-analysis. *Anesth Analg* 1999; **88**:1362–9

23 Jewell D, Young G. Interventions for nausea and vomiting in early pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 4*, 1998. Oxford, Update Software

24 Aikins Murphy P. Alternative therapies for nausea and vomiting of pregnancy. *Obstet Gynecol* 1998;**91**:149–55

25 Dobie RA. A review of randomized clinical trials in tinnitus. *Laryngoscope* 1999;**109**:1202–11

26 Park J, White AR, Ernst E. Efficacy of acupuncture as a treatment of tinnitus. *Arch Otolaryngol Head Neck Surg* 2000;**126**:489–2

27 Moher D, Pham B, Ausejo M, Saenz A, Hood S, Barber GG. Pharmacological management of intermittent claudication: a meta-analysis of randomised trials. *Drugs* 2000;**59**:1057–70

28 Pittler MH, Ernst E. Ginkgo biloba extract for the treatment of intermittent claudication: a meta-analysis of randomized trials. *Am J Med* 2000;**108**:276–81

29 Volz HP. Controlled clinical trials of hypericum extract in depressed patients—an overview. *Pharmacopsychiatry* 1997;**30**(suppl):72–5

30 Linde K, Mulrow CD. St John's wort for depression (Cochrane Review). In: *The Cochrane Library, Issue 4*, 1998. Oxford: Update Software

31 Kim HL, Streltzer J, Goebert D. St John's wort for depression: A meta-analysis of well-defined clinical trials. *J Nerv Ment Dis* 1999;**187**: 532–9

32 Williams JW Jr, Mulrow CD, Chiquette E, Hitchcock Noel P, Aguilar C, Cornell J. A systematic review of newer pharmacotherapies for depression in adults: Evidence report summary. *Ann Intern Med* 2000; **132**:743–56

33 Gaster B. St John's wort for depression. A systematic review. *Arch Intern Med* 2000;**160**:152–6

34 Stevinson C, Pittler MH, Ernst E. Garlic for treating hypercholesterinemia—a meta-analysis of randomized clinical trials. *Ann Intern Med* 2000;**133**:420–9

35 Lawrence V, Mulrow C, Ackerman R, *et al*. Garlic: effects on cardiovascular risks and disease, protective effects against cancer, and clinical adverse effects. *Evidence Report/Technology Assessment: Number 20*, 2000 [www.ahcpr.gov./clinic/garlicsum.htm]

36 Warshafsky S, Kamer RS, Sivak SL. Effect of garlic on total serum cholesterol. *Ann Intern Med* 1993;**119**:599–605

37 Neil HAW, Silagy CA, Lancaster *et al*. Garlic powder in the treatment of moderate hyperlipidaemia: a controlled trial and meta-analysis. *J R Coll Gen Pract* 1996;**30**:329–34

38 Melchart D, Linde K, Fischer P, Kaesmayr J. Echinacea for prevention and treatment of the common cold (Cochrane Review). In: *The Cochrane Library, Issue 1*, 1999. Oxford: Update Software

39 Barratt B, Vohmann M, Calabrese C. Echinacea for upper respiratory tract infection. *J Fam Pract* 1999;**48**:628–35

40 Pittler MH, Ernst E. Peppermint oil for irritable bowel syndrome: a critical review and meta-analysis. *Am J Gastroenterol* 1998;**93**: 1131–5

41 Jailwala J, Imperiale TF, Kroenke K. Pharmacologic treatment of the irritable bowel syndrome: a systematic review of randomized, controlled trials. *Ann Intern Med* 2000;**133**:136–47

42 Linde K, Clausius N, Ramirez G, *et al*. Are the clinical effects of homeopathy placebo effects? A meta-analysis of randomized placebo-controlled trials. *Lancet* 1997;**350**:834–43

43 Walach H. Unspezifische Therapie-Effekte. Das Beispiel Homöopathie. Habilitationsschrift, Psychologisches Institut, Albert-Ludwigs-Universität, Freiburg, 1997

44 Cucherat M, Haugh MC, Gooch M, Voissel JP. Evidence of clinical efficacy of homeopathy. A meta-analysis of clinical trials. *Eur J Clin Pharmacol* 2000;**56**:27–33

45 Hill C, Doyon F. Review of randomized trials of homoeopathy. *Rev Epidémiol Santé Publique* 1990;**38**:139–47

46 Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homoeopathy. *BMJ* 1991;**302**:316–23

47 Ernst E, Pittler MH. Efficacy of homoeopathic Arnica. A systematic review of placebo-controlled clinical trials. *Arch Surg* 1998;**133**:1187–90

48 Lüdtke R, Wilkens J. Klinische Wirksamkeitsstudien zu Arnica in homöopathischen Zubereitungen. In: Albrecht H, Frühwald M, eds. *Karl und Veronica Carstens-Stiftung, Jahrbuch Band 5*. Essen: KVC Verlag, 1999:97–112

49 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA* 1995;**273**:408–12

50 Assendelft WJJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA* 1995;**274**:1942–8

51 Jadad AR, Moher M, Browman GP, *et al*. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ* 2000; **320**:537–40

52 Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;**156**:1411–16

53 Linde K, Melchart D, Brandmaier R, Eitel F. Critical evaluation of papers reviewing controlled clinical trials in homoeopathy. *Br Hom J* 1994;**83**:167–73

54 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement. *Lancet* 1999;**354**:1896–900