 *International Journal of Epidemiology* 2001;**30**:526–531

# The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture

Klaus Linde,[a,b] Wayne B Jonas,[c] Dieter Melchart[b] and Stefan Willich[a]


| | |
|---|---|
| **Background** | To investigate the methodological quality of randomized controlled trials in three areas of complementary medicine. |
| **Methods** | The methodological quality of 207 randomized trials collected for five previously published systematic reviews on homeopathy, herbal medicine (Hypericum for depression, Echinacea for common cold), and acupuncture (for asthma and chronic headache) was assessed using a validated scale (the Jadad scale) and single quality items. |
| **Results** | While the methodological quality of the trials was highly variable, the majority had important shortcomings in reporting and/or methodology. Major problems in most trials were the description of allocation concealment and the reporting of drop-outs and withdrawals. There were relevant differences in single quality components between the different complementary therapies: For example, acupuncture trials reported adequate allocation concealment less often (6% versus 32% of homeopathy and 26% of herb trials), and trials on herbal extracts had better summary scores (mean score 3.12 versus 2.33 for homeopathy and 2.19 for acupuncture trials). Larger trials published more recently in journals listed in Medline and in English language scored significantly higher than trials not meeting these criteria. |
| **Conclusion** | Trials of complementary therapies often have relevant methodological weaknesses. The type of weaknesses varies considerably across interventions. |


| | |
|---|---|
| **Keywords** | Alternative medicine, homeopathy, acupuncture, Hypericum, Echinacea, controlled clinical trials, randomized controlled trials, meta-analysis |
| **Accepted** | 21 September 2000 |


Complementary and alternative therapies are widely used but their effectiveness is controversial.[1,2] Therefore, there is a need for systematic overviews summarizing the existing evidence. Assessing the methodological quality of primary studies is an important part of any review process. Methodological quality refers mainly to formal aspects of study design, performance and analysis. Methods for assessment of quality typically focus on the randomization process, blinding and handling of drop-outs and withdrawals. There is increasing evidence from conventional medicine that more rigorous studies yield less positive results.[3,4] In a systematic review of placebo-controlled trials of homeopathy we found similar trends.[5,6]

Many of the currently available systematic reviews criticize the methodological quality of complementary medicine trials (for example [7–11]). However, there are no empirical studies published comparing the methodological quality of trials on different types of complementary therapies or with trials of conventional medicine. We have performed five systematic reviews of trials on homeopathy, herbal extracts and acupuncture.[5,12–15] As a part of these reviews we tried to assess quality using a variety of methods including a validated scale on quality of reporting[16] which has also been used extensively in conventional medicine.[3,17]

We re-analysed our data with the following objectives: (1) to get an overview of the methodological quality of trials in the

[a] Institute for Social Medicine & Epidemiology, Charité, Humboldt-University, Berlin, Germany.

[b] Centre for Complementary Medicine Research, Department of Internal Medicine II, Technische Universität, Kaiserstr. 9, 80801 München, Germany.

[c] Department of Family Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA.

Correspondence: Dr med. Klaus Linde, Centre for Complementary Medicine Research, Department of Internal Medicine II, Technische Universität, Kaiserstr. 9, 80801 München, Germany. E-mail: Klaus.Linde@lrz.tu-muenchen.de

Conflict of interest: None

three areas of complementary medicine reviewed; (2) to find out whether there are differences between those three areas; and (3) to investigate whether methodological quality differs with regard to place, language and year of publication as well as sample size.

## Materials and Methods

The data for the present study was originally collected for systematic reviews on homeopathy,[5] Hypericum extracts for depression,[12] Echinacea extracts for the common cold,[13] acupuncture for asthma[14] and recurrent headaches.[15] The literature for these reviews had been identified through comprehensive searches in Medline, Embase, complementary medicine databases, screening of bibliographies, and by contacting researchers in the field. To be included the trials had to meet the following criteria: (1) subjects treated for curative, palliative or preventative purposes; (2) random or quasi-random allocation (e.g. alternate allocation); trials which did not explicitly describe the method of allocation were included if the methodology applied made random allocation likely (e.g. if the person selecting the patients was blinded); (3) comparison of the test intervention with a placebo condition, no treatment or another treatment. Trials measuring physiological outcomes only were excluded. A total of 207 studies (129 from the homeopathy review, 27 on Hypericum for depression, 15 on Echinacea for the common cold, 15 on acupuncture for asthma, and 21 on acupuncture for recurrent headaches) met these criteria.

For the re-analysis the following information was taken from the databases established for the five reviews: Type of publication (Medline-listed journal, other journal, other form of publication, unpublished report), year of first publication, country of the first author, availability of a report published in English language (whether or not this was the principal publication), sample size, type of control intervention (no treatment, placebo, other therapy), data on study quality.

Methodological quality was assessed in all five reviews using a scale developed and validated by Jadad *et al.*[16] This scale assesses the completeness of reporting using three items with a maximum score of five points. If the allocation to groups is explicitly randomized item 1 is scored. A bonus point is given if an adequate method to generate the random sequence is described. If there is an explicit statement that the study is double-blind (blinding of patients and evaluators, not necessarily therapists) item 2 is scored. A bonus point is given if the method of blinding is described and adequate. Item 3 is scored if there is either an explicit statement that all patients included were also analysed or if the number and reasons for drop-outs in all groups are given separately. For being classified as adequately reported a trial should score at least three of five points, a cut-off point recommended by the authors of the scale.[18] In addition, a self-developed instrument was used from which we report here only the data on allocation concealment. This item could be scored as unclear (if there was no information), inadequate (methods in which allocation is not blinded, such as alternation, date of birth, etc.), possibly inadequate (methods which, in principle, are adequate, but for which cases of unblinding have been reported, such as sealed envelopes) or adequate (methods in which concealment is generally considered as adequately blinded, such as central randomization or consecutively numbered, identically packaged drug containers).

All extraction and quality assessments were performed by at least two independent reviewers using standard forms developed for each review. Disagreements were documented and discussed with final decisions made by the principal reviewer. A total of nine reviewers participated in scoring; one reviewer (KL) assessed all trials. Agreement before discussion (calculated for the four reviewers assessing more than ten trials) proved to be good to very good with intraclass correlation coefficients between 0.65 and 0.96.

The data from the five reviews were combined in an SPSS database file (SPSS Corp., Chicago, Ill.). The Kruskal-Wallis test and the $\chi^2$ test were used to analyse differences between homeopathy, herbal medicine, and acupuncture trials regarding Jadad scores and single quality criteria. To test whether Jadad quality scores were correlated with publication place (Medline-listed or not), language (English or other language), date (until or after 1989) and sample size (less than 100 and more patients) crude and adjusted mean quality scores were calculated using the GLM univariate procedure (general linear regression model) in SPSS. To test whether these factors were also correlated with single aspects of methodological quality a logistic regression was performed. When tests were performed across therapies only a randomly chosen third of the homeopathy trials were included in the analysis to prevent the disproportionately large number of these trials from influencing the results. $P$-values $< 0.05$ were considered statistically significant; no adjustments were made for multiple comparisons.

## Results

The trials on homeopathy, herbal extracts and acupuncture differed considerably in various aspects. Trials on acupuncture were generally smaller (median sample size was 15 patients in asthma trials and 33 in the headache trials) than trials on homeopathy or herbal remedies (Table 1). Trials on herbal remedies were almost exclusively performed in Germany and mostly published in the German language. Apart from acupuncture less than half of the trials were available in English or referenced in Medline. In all three areas the great majority of trials were placebo-controlled.

While the quality of the trials was highly variable, the majority had important shortcomings in reporting and/or methodology (Table 2). Most trials did not describe the generation of the random sequence, an adequate method to conceal allocation, and the number and reasons for drop-outs and withdrawals. There were considerable differences regarding single components of methodological quality between homeopathy, herb and acupuncture trials. Homeopathy trials were less often explicitly randomized while successful blinding was more often questionable in trials of herbs and acupuncture. Herb trials (and particularly Hypericum trials) more frequently described drop-outs and withdrawals in sufficient detail, but intent-to-treat analysis was very rare in all areas. The summary scores of trials on herbal extracts were on average higher than of those on homeopathy and acupuncture. This was mainly due to the better quality of the Hypericum trials (mean Jadad score 3.56) while the Echinacea trials scored in a similar range (mean score 2.33) to the acupuncture and homeopathy trials.

**Table 1** Overview of the trials included

| | Homeopathy | Herbs | Acupuncture |
|---|---|---|---|
| **No. of trials** | 129 | 42 | 36 |
| **Interventions included** | All homeopathy | Echinacea | All acupuncture |
| | | Hypericum | |
| **Conditions included** | All | Common cold | Asthma |
| | | Depression | Idiopathic headache |
| **Total sample size (median, range)** | 60 (5–1306) | 100 (28–646) | 28 (10–150) |
| **Type of control**[a] | | | |
| Placebo | 114 (88%) | 29 (69%) | 29 (81%) |
| Other therapy | 19 (15%) | 10 (24%) | 9 (25%) |
| No treatment | 5 (4%) | 3 (7%) | 2 (11%) |
| **Countries of origin** | | | |
| North America | 4 (3%) | – | 2 (6%) |
| UK (incl. Hong Kong) | 26 (20%) | 1 (2%) | 8 (22%) |
| Germany | 43 (33%) | 39 (93%) | 8 (22%) |
| France | 33 (26%) | – | 2 (6%) |
| Remaining Europe | 17 (13%) | 2 (5%) | 9 (25%) |
| Other countries | 6 (5%) | – | 7 (19%) |
| **Available in English** | 49 (38%) | 12 (29%) | 25 (69%) |
| **Listed in Medline** | 30 (23%) | 7 (17%) | 25 (69%) |

[a] Some trials with more than one control group.

**Table 2** Results of the quality assessments. Values are number of trials (percentages) unless indicated otherwise

| Criterion (score points) | Homeopathy | Herbs | Acupuncture | *P*-value |
|---|---|---|---|---|
| **Jadad score** | | | | |
| Randomization | | | | |
| Not randomized/unclear (0) | 45 (35%) | 5 (12%) | 4 (11%) | |
| Randomization stated (1) | 63 (49%) | 24 (57%) | 29 (81%) | |
| Rand. stated + sequence generation described (2) | 21 (16%) | 13 (31%) | 3 (8%) | <0.001 |
| Double-blinding | | | | |
| Not double-blind/unclear (0) | 24 (19%) | 7 (17%) | 17 (47%) | |
| Stated (1) | 51 (39%) | 18 (43%) | 8 (22%) | |
| Described and adequate (2) | 54 (42%) | 17 (40%) | 11 (31%) | 0.006 |
| Drop-outs | | | | |
| Not/insufficiently described (0) | 96 (74%) | 17 (41%) | 22 (61%) | |
| Described (1) | 33 (26%) | 25 (59%) | 14 (39%) | <0.001 |
| Mean score (standard deviation) | 2.33 (1.36) | 3.12 (1.33) | 2.19 (1.17) | 0.002 |
| Trials scoring <3 points | 78 (60%) | 12 (29%) | 22 (61%) | |
| Trials scoring ⩾3 points | 51 (40%) | 30 (71%) | 14 (39%) | 0.001 |
| **Allocation concealment** | | | | |
| Inadequate (e.g. alternation)/unclear | 88 (68%) | 31 (74%) | 31 (88%) | |
| Possibly inadequate (e.g. sealed envelope) | – | – | 2 (6%) | |
| Adequate (e.g. consecutively numbered drugs) | 41 (32%) | 11 (26%) | 2 (6%) | 0.001 |

*P*-values from $\chi^2$-test.

Larger trials published more recently in journals listed in Medline and in English language scored significantly higher than trials not meeting these criteria. (Table 3). Double-blinding, reporting of an adequate method of concealment and reporting of drop-outs and withdrawals were all significantly more frequent in more recent studies. Other factors were less consistently associated with indicators of better methodological quality (Table 4).

## Discussion

The results of our analysis (1) confirm that many randomized controlled trials of complementary medicine interventions have relevant methodological flaws; (2) show that trial characteristics and methodological shortcomings can vary considerably between different types of complementary therapies; and that

**Table 3** Mean quality scores stratified for publication source, time, language, and sample size (adjusted scores and *P*-values calculated from linear regression)

| | **All therapies** | | | **Homeopathy** | | | **Herbs** | | | **Acupuncture** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean score | | | Mean score | | | Mean score | | | Mean score | | |
| | n[b] | adj.[c] | crude | n | adj. | crude | n | adj. | crude | n | adj. | crude |
| **Publication in** | | | | | | | | | | | | |
| Medline-listed journal | 46 | 2.95 | 3.04 | 30 | 3.10 | 3.27 | 9 | 3.98 | 4.56 | 25 | 2.42 | 2.40 |
| Other source/unpublished | 74 | 2.42 | 2.36 | 99 | 2.10 | 2.05 | 32 | 2.88 | 2.72 | 11 | 1.68 | 1.73 |
| *P*-value | | 0.031 | | | 0.000 | | | 0.030 | | | 0.115 | |
| **Publication** | | | | | | | | | | | | |
| Before 1990 | 68 | 2.22 | 2.21 | 85[d] | 2.14 | 2.07 | 14 | 2.66 | 0.21 | 24 | 2.00 | 2.08 |
| 1990 and later | 52 | 3.15 | 3.17 | 42 | 2.74 | 2.88 | 27 | 3.63 | 3.59 | 12 | 2.58 | 2.42 |
| *P*-value | | 0.000 | | | 0.009 | | | 0.084 | | | 0.176 | |
| **Report available** | | | | | | | | | | | | |
| In English language | 61 | 2.88 | 2.97 | 49 | 2.70 | 2.84 | 12 | 3.61 | 4.25 | 25 | 2.33 | 2.40 |
| Only in another language | 59 | 2.36 | 2.27 | 80 | 2.11 | 2.03 | 29 | 2.92 | 2.66 | 11 | 1.89 | 1.73 |
| *P*-value | | 0.027 | | | 0.009 | | | 0.148 | | | 0.353 | |
| **Sample size** | | | | | | | | | | | | |
| <100 | 87 | 2.47 | 2.48 | 91 | 2.19 | 2.14 | 19 | 3.13 | 3.16 | 35 | no analysis | |
| ⩾100 | 33 | 3.03 | 3.00 | 38 | 2.67 | 2.79 | 22 | 3.11 | 3.09 | 1 | | |
| *P*-value | | 0.023 | | | 0.041 | | | 0.947 | | | | |

[a] Only 42 randomly selected homeopathy trials were included to prevent the large number of homeopathy trials bias the results.

[b] No. of trials.

[c] Adjusted mean scores.

[d] Two unpublished studies without publication date.

**Table 4** Influence of language, source, year of publication and sample size on reporting of key methodological issues. Values are odds ratios (OR) (95% CI) of trials meeting the condition compared to those not meeting the condition calculated by logistic regression

| | **Double-blinding** | | **Adequate concealment** | | **Reporting of drop-outs** | |
|---|---|---|---|---|---|---|
| **Factor** | OR (95%CI) | *P* | OR (95%CI) | *P* | OR (95%CI) | *P* |
| Available in English | 1.69 (0.64–4.47) | 0.290 | 4.54 (1.47–14.02) | 0.009 | 0.92 (0.40–2.31) | 0.921 |
| Medline-listed publication | 0.77 (0.28–2.06) | 0.600 | 0.70 (0.24–2.03) | 0.511 | 3.50 (1.41–8.68) | 0.007 |
| Published 1990 or later | 3.72 (1.37–10.10) | 0.010 | 2.57 (1.02–6.46) | 0.046 | 3.39 (1.51–7.59) | 0.003 |
| sample size ⩾100 | 2.09 (0.67–6.26) | 0.208 | 4.64 (1.70–12.65) | 0.003 | 1.32 (0.54–3.31) | 0.523 |

(3) date, place and language of publication as well as sample size correlate with methodological quality in the study sample reviewed.

When interpreting our data it has to be kept in mind that we focused only on three major complementary therapies. Design features and quality problems might be quite different in other areas of complementary medicine. Furthermore, while we have collected all available randomized controlled trials on homeopathy up to 1996 our reviews on acupuncture and herbal medicine were refined to specific questions and, therefore, the trials are not necessarily representative for all acupuncture and all herbal medicine trials.

The usefulness of scores to assess methodological quality is controversial. While some consider quality scores a useful tool for a pragmatic overall assessment of the quality of a study,[19] others argue that scores can be misleading and prefer to focus on single quality components and their impact on study outcome.[20,21] Therefore, our analysis relies both on summary scores and single quality aspects whose relevance have been shown empirically.[3,4,6] The Jadad scale was systematically developed and underwent an empirical validation process. As the scale seemed problematic to us on a conceptual level (lack of items for randomization concealment, baseline comparability, and focus on reporting of drop-outs rather than risk of bias introduced by drop-outs) we also used a scale which we had developed at the same time. The scales produced similar results but the Jadad scale had better discriminative power.

From our experience a major problem with all formal assessments of quality—be it by scales or by single quality components—is that they are very crude and mainly an assessment of reporting quality. A few words (or the lack thereof) can determine whether a trial is considered 'high' or 'low' quality. While better reporting generally tends to be associated with better methodological quality the assessment can be highly misleading for single studies.

Our finding that many of the complementary medicine trials reviewed have relevant methodological shortcomings is in accordance with other analyses (e.g. [7–11,22,23]). The differences regarding single quality items indicate that certain methodological problems are more prominent in some areas of complementary medicine than in others which is probably due to the different nature of the interventions. For example, double-blinding is

more difficult for physical interventions like acupuncture than for drug interventions like homeopathy or herbs. Also, while adequate allocation concealment in drug trials can be done at very low cost using coded drug containers, the cheapest concealment method for physical interventions (sealed envelopes) is generally not considered as foolproof.[24] The reasons for other quality differences are less clear: We cannot explain why homeopathy trials are less frequently randomized and report less details on drop-outs and withdrawals. The almost complete absence of intent-to-treat analysis was a particular feature in all complementary medicine areas we reviewed.

Some of the discrepancies regarding other issues have implications for performing systematic reviews: While reviews on acupuncture restricted to trials published in English language in Medline-listed journals might gather most of the relevant information such restrictions would result in the exclusion of more than two-thirds of homeopathy and herb trials (at least in case of the two examples studied). On the other hand we found that publication in English and in a Medline-listed journal was associated with higher quality scores. Similar analyses in conventional medicine did not find a correlation between language and quality scores.[17,25]

We had planned to compare the reporting quality of complementary medicine trials with a set of 'conventional medicine trials' from the studies by Moher *et al*. from 1996 and 1998.[3,17] For the 1996 study 229 randomized trials published between 1989 and 1994 were collected from seven English language journals and from six journals published in other languages to investigate differences in trials published in English or another language. For the 1998 study Moher *et al*. had collected 127 randomized trials from 11 meta-analyses on a variety of interventions for different conditions to investigate the impact of quality aspects on trial outcome. Both studies used the Jadad scale and assessed adequacy of allocation concealment in addition. One of us (KL) was involved in the 1996 study.[17] The instructions for scoring were identical to those used in our reviews. However, we were unable to obtain data on the characteristics and results of the conventional medicine trials beyond that published. Consequently, a detailed comparison was not possible. The mean quality scores of the trials reviewed in the studies by Moher *et al*. were very similar to those in the complementary medicine trials reviewed (2.55 and 2.74 versus 2.61) but the latter were more often double-blind and more often reported an adequate method of allocation concealment. Instead, the complementary medicine trials less frequently reported on drop-outs and withdrawals. Without further information on the characteristics of the conventional medicine trials these results are difficult to interpret. However, they seem to suggest that the average methodological quality of the complementary medicine trials reviewed is not necessarily worse than that of trials in conventional medicine.

In conclusion, there is a clear need for improved methods and reporting of methods in clinical research on homeopathy, herbal medicine and acupuncture. According to our results (in a sample which might not be representative for all herbal medicine and acupuncture trials), reporting and handling of drop-outs and withdrawals seem to be a major problem in all therapies reviewed. Homeopathic researchers should take advantage of random allocation wherever possible. People undertaking randomized trials on acupuncture should make sure that they use adequate methods for concealment of allocation. The adoption of the CONSORT guidelines[26] for reporting would allow much more reliable assessment of methodological quality in future systematic reviews.

## Acknowledgement

This study was in part supported by the Karl und Veronica Carstens Foundation and the NIH Grant No 5 U24-AR-43346–02. We would like to thank Wolfgang Weidenhammer for his help with the analysis of the data.

KEY MESSAGES

- Aspects of the methodological quality of 207 randomized trials of homeopathy, herbal medicines, and acupuncture were reviewed.
- The majority of trials had important shortcomings; major problems were the reporting of allocation concealment and the reporting and handling of drop-outs and withdrawals.
- The extent of specific quality problems differed considerably between the different areas.
- Larger trials published more recently in journals listed in Medline and in English language had better methodological quality.

## References

[1] Eisenberg DM, Davis RB, Ettner SL *et al*. Trends in alternative medicine use in the United States, 1990–1997. *JAMA* 1998;**280:**1569–75.

[2] MacLennan AH, Wilson DH, Taylor AW *et al*. Prevalence and cost of alternative medicine in Australia. *Lancet* 1996;**347:**569–73.

[3] Moher D, Pham B, Jones A *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352:**609–13.

[4] Schulz KF, Chalmers I, Hayes RJ *et al*. Empirical evidence of bias. *JAMA* 1995;**273:**408–12.

[5] 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.

[6] Linde K, Scholz M, Ramirez G *et al*. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol* 1999;**52:**631–36.

[7] Kleijnen J, Knipschild P. Mistletoe treatment for cancer. Review of controlled trials in humans. *Phytomedicine* 1994;**1:**255–60.

[8] Rosted P. The use of acupuncture in dentistry: a review of the scientific validity of published papers. *Oral Diseases* 1998;**4:** 100–04.

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

[10] Assendelft WJJ, Koes BW, van der Heijden GJMG *et al*. The effectiveness of chiropractic for treatment of low back pain: an update and attempt at statistical pooling. *J Manipulative Physiol Ther* 1996;**19:** 499–507.

[11] Bloom BS. *Best Evidence for Alternative Medicine*. Book of Abstracts, 14th Annual Meeting International Society of Technology Assessment in Health Care (ISTAHC), 1998.

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

[13] Melchart D, Linde K, Fischer P *et al*. Echinacea for the prevention and treatment of the common cold (Cochrane review). In: *The Cochrane Library*, Issue 1, 1999. Oxford: Update Software.

[14] Linde K, Worku F, Stör W *et al*. Randomized clinical trials of acupuncture for asthma—a systematic review. *Forsch Komplementärmed* 1996;**3:**148–55.

[15] Melchart D, Linde K, Fischer P *et al*. Acupuncture for recurrent headaches: a systematic review of randomized controlled trials. *Cephalalgia* 1999;**19:**779–86.

[16] Jadad AR, Moore RA, Carrol D *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;**17:**1–12.

[17] Moher D, Fortin P, Jadad AR *et al*. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347:** 363–66.

[18] Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996;**156:**661–66.

[19] Olkin I. Invited commentary: Re: A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140:**297–99.

[20] Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994;**140:**300–01.

[21] Jüni P, Witschi A, Bloch R *et al*. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282:**1054–60.

[22] Van Tulder MW, Cherkin DC, Berman B, Lao L, Koes BW. Acupuncture for low back pain (Cochrane review). In: *The Cochrane Library*, Issue 1, 2000. Oxford: Update Software.

[23] Hammerschlag R, Morris MM. Clinical trials comparing acupuncture with biomedical standard care: a criteria-based evaluation of research design and reporting. *Complement Ther Med* 1997;**5:**133–40.

[24] Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;**274:**1456–58.

[25] Egger M, Zellwäger-Zähner T, Schneider M *et al*. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350:**326–29.

[26] Begg C, Cho M, Eastwood S *et al*. Improving the quality of reporting of randomized trials: the CONSORT statement. *JAMA* 1996;**276:**637–39.



*International Journal of Epidemiology* 2001;**30**:531–532

# Commentary: Science friction—complementary/alternative medicine on the stony road from opinion to evidence

E Ernst

Linde and colleagues present an assessment of 207 randomized clinical trials (RCT) of complementary/alternative medicine (CAM) which points to 'relevant methodological weaknesses' in the evidence supporting CAM.[1] Why is this important?

Complementary/alternative medicine is used by more and more people[2] and continues to grow at a rate that sends a shiver down the spine of many scientifically-minded physicians. A US think-tank recently concluded that 'by 2010 at least two-thirds (of the US population) will be using one or more of the approaches we now consider complementary or alternative'.[3] Is this the advent of the 'age of unreason'?

As a popular consumer-based movement, CAM is almost entirely opinion-based. We recently evaluated the recommendations of seven leading CAM books for specific medical conditions.[4] The results demonstrate vividly the dominance of opinion over evidence. Firstly, the recommendations of these seven authors yielded close to zero consistency. Secondly, treatments were recommended which, according to evidence from reliable RCT, are ineffective, in some cases even contra-indicated. Thirdly, treatments which were of proven effectiveness were not recommended by some of these authors.

Such data suggest that the time is overdue to replace opinion by evidence. Therefore the reminder by Linde and colleagues,[1] that much of the CAM evidence lacks methodological rigor, is welcome, timely and important. Most probably it also is generalizable across all areas of CAM. US researchers looked at more than 5000 trials of CAM but only 258 met their inclusion criteria. They concluded that 'the overall quality of evidence for CAM RCT is poor'.[5] We evaluated 2938 RCT from the Chinese literature, found major methodological shortcomings, and

Department of Complementary Medicine, School of Postgraduate Medicine and Health Sciences, University of Exeter, 25 Victoria Park Road, Exeter EX2 4NT, UK. E-mail: E.Ernst@ex.ac.uk

concluded that 'the quality of trials of traditional Chinese medicine must be improved urgently'.[6]

While we scientists lament the low average quality of RCT, providers of CAM very often have quite a different agenda and rarely feel the need for scientific scrutiny at all. They often argue that CAM, for a number of reasons, defies the straight jacket of reductionistic science. Most readers of the *International Journal of Epidemiology* will agree that this attitude must be based on misunderstandings; perhaps the best evidence in support is provided by the fact that Linde[1] and others[5,6] had few no problems locating RCT. What is true, however, is that scientists are constantly and miserably failing to get their points across to advocates of CAM.

Randomized clinical trials of CAM are often more difficult and methodologically more challenging than RCT of other types of interventions. Due to the nature of most CAM modalities and the conditions they are used for, such RCT often need to be large, of long duration and require expensive therapists' time. In turn, this means that CAM research is expensive and requires high levels of expertise in terms of trial design. The demoralizing facts demonstrate, however, that research funds for CAM are rare as gold-dust[7] and the infrastructure or culture for CAM research is largely non-existent. Unless fortunes change dramatically, CAM research has therefore little chance of improving. So, is there no hope at all?

Unsurprisingly the solution to these problems lies in creating sufficient resources for supporting CAM research to a level warranted by its popularity. As very few commercial interests are applicable to CAM, the bulk of this money will have to come from official (e.g. government) funds. It is high time that this message is absorbed into political will. On the one hand, most governments seem to feel that, as long as CAM can be contained within the private sector, it will not draw money from their budgets. On the other hand, politicians permanently feel the need to be popular, and lip service to CAM provides one way of fulfilling this need. But *vis-à-vis* the abundance of open questions and the growing prevalence of CAM use, lip service no longer suffices—it is time to stop talking and start researching in earnest, with rigour and adequate support. To ignore this challenge is nothing less than to ignore the need of the public.

## References


1 Linde K, Jonas WB, Melchart D, Willich S. The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. *Int J Epidemiol* 2001;**30:**526–31.

2 Ernst E. Prevalence of use of complementary/alternative medicine: a systematic review. *Bull World Health Organ* 2000;**78:**252–57.

3 The Institute of Alternative Futures. *The Future of Complementary and Alternative Approaches in US Health Care*. 1998 NCMIC Insurance Company.

4 Ernst E, Eisenberg D. *Complementary and Alternative Medicine—A Desk Top Guide*. Mosby, 2001.

5 Bloom BS, Retbi A, Dahan S, Jonsson E. Evaluation of randomized controlled trials on complementary and alternative medicine. *Int J Tech Assess Health Care* 2000;**16:**13–21.

6 Tang J, Zhan S, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;**319:**160–61.

7 Ernst E. Funding research into complementary medicine: the situation in Britain. *Complement Ther Med* 1999;**7:**250–53.