Performing systematic reviews of clinical trials of acupuncture: problems and solutions

Adrian White, Kien Trinh, Richard Hammerschlag

Systematic reviews of randomized controlled trials, rather than narrative reviews, generally provide the approach that is least subject to bias for assessing the efficacy and effectiveness of a therapy. But despite their increasing use, systematic reviews are not free of problems, particularly when used to assess the evidence for acupuncture. Common weaknesses are inadequate literature searches and unclear inclusion criteria, including absence of a definition of 'acupuncture' for delineating the scope of the review. In addition, the adequacy of the acupuncture performed in the trials is often not addressed. Overall, there is little consistency in the criteria used for critiquing the included trials. The validated five-point, and expanded nine-point, Jadad scales for assessing trial quality are discussed, especially in regard to their emphasis on double blinding that can be applied to acupuncture trials if 'double blind' is defined as patient and assessor blinding. Suggestions are made for avoiding each of the above-cited problems for future systematic reviews, which should include an acupuncture specialist in the author team. For the near future, however, there is a greater need for new high-quality RCTs of acupuncture than for additional systematic reviews of existing trials.

Adrian White, MA BM BCh, Senior Lecturer, Department of Complementary Medicine, School of Sport and Health Sciences, University of Exeter, 25, Victoria Park Road, Exeter, EX2 4NT, UK. Tel: 01392 424839; Fax: 01392 424989; E-mail: a.r.white@ex.ac.uk
Kien Trinh, MD, BSc (Math), MSc (Epid), Assistant Clinical Professor, Chair, Medical Acupuncture Program: An Evidence-based Approach, School of Medicine, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada, L8N 3Z5. Tel: 1(905)/648-4425; Fax: 1 (905)/648-4426; E-mail: trinhk@mcmaster.ca
Richard Hammerschlag, PhD, Research Director, Oregon College of Oriental Medicine, 10525 SE Cherry Blossom Drive, Portland, OR 97216, USA. Tel: 503/253-3443, x156; Fax: 503/253-2701; E-mail: rhammerschlag@ocom.edu

Correspondence to: **Adrian White**

INTRODUCTION

Evidence-based medicine has been defined as the conscious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.[1] The practice of evidence-based medicine means integrating individual clinical experience with the best available evidence from clinical research. It does not mean, as commonly interpreted, only using treatments that have been demonstrated as effective in RCTs.

Systematic reviews form the highest level of evidence of efficacy and effectiveness for practising evidence-based medicine, since they reduce bias by following pre-established methods to obtain all available information and to evaluate it even-handedly. Systematic reviews usually use evidence from RCTs, though reviews can include other trial designs such as uncontrolled studies when there are insufficient RCTs of acceptable quality.

A systematic review is an overview of primary studies that contains an explicit statement of objectives, materials, and methods and has been conducted according to explicit and reproducible methodology.[2] The rationale for the place of systematic reviews in health care has been discussed in detail and includes the needs to evaluate evidence critically, condense large amounts of information into digestible quantity and format, integrate important pieces of information, assess the consistency of

the results of different trials and hopefully explain any differences, and generalize the findings.[3] Systematic reviews often have greater power than individual studies and allow more precise estimation of effects.

The purpose of systematic reviews is to improve patient care. The evidence from systematic reviews may save lives. For example, the availability of a cumulative systematic review and meta-analysis could have shown streptokinase to be life-saving in myocardial infarction 20 years before its actual submission to and approval by the United States FDA and its adoption in practice.[3] Unfortunately, some experts continue to recommend treatments which have been shown by systematic reviews to be inferior to others.[4]

As is common in scientific advances, systematic reviews are not an unrelieved boon. The method is to an extent still evolving, and reviews need to be performed with careful attention to the subject. For example, discordant results are not uncommon and reasons for these have been discussed in detail.[5] Large, well-designed and well-performed RCTs provide the most reliable evidence, particularly when subsequently evaluated in a systematic review. Nevertheless, in the absence of large-scale trials in acupuncture, many systematic reviews of acupuncture have been performed. These have been assessed and criticized[6–8] and the topic of systematic reviews in acupuncture was discussed at a conference at Exeter University in July, 2001. This paper summarizes the concerns and recommendations from that conference.

QUALITY OF SYSTEMATIC REVIEWS IN ACUPUNCTURE

There are accepted standards for performing systematic reviews. Methodological guidelines were reviewed,[9] subjected to consensus in the Potsdam consultation,[10] and updated in the consensus QUOROM statement.[11] These basic procedures apply to systematic reviews of all therapies but there are special considerations that apply to this type of review of acupuncture.

Assessing the quality of existing systematic reviews is problematic: one attempt to provide an index was limited by the need to use judgement in applying the index (which makes it subject to bias).[12] The QUOROM consensus statement, detailing what should be included in a systematic review, can be used as a checklist for quality, but has not been validated as a quality assessment instrument.[11] Since the quality of systematic reviews is known to affect their conclusions,[13] pragmatic criteria for assessing the performance of a systematic review have been developed by the York Centre for Reviews and Dissemination. These criteria are available on the Cochrane Library and York website (www.york.ac.uk/inst/crd) but this evaluating tool also has not been validated.

It is of value to examine the recent review of systematic reviews of acupuncture by Linde and colleagues[14] as a guideline for the present discussion. Their literature search identified 49 potential reviews of which 39 met their inclusion criteria. Of these included reviews, 22 were on various pain or rheumatological conditions. The problems most frequently noted with the primary studies were small sample size, heterogeneity of patients, interventions and outcome measures, and insufficient quality. The quality of the 39 reviews was assessed by a Yes/No response to five criteria:

- Comprehensive search
- Explicit inclusion criteria
- Formal quality assessment
- Summary of results for each included study
- Meta-analysis.

Some of these criteria are open to subjective interpretation. Some are likely to require specialist information. For this reason, it seems essential that an acupuncturist should be included in the team of reviewers, together with independent researchers without overt conflict of interest. An expanded version of this list will be used as the basis for discussion.

PERCEIVED PROBLEMS WITH SYSTEMATIC REVIEWS IN ACUPUNCTURE

(1) Incomplete literature search

(a) Much complementary medicine literature, including that on acupuncture, is not included on the standard databases, and special efforts should be made to identify and search, by hand if necessary, specialized databases as well as non-database sources such as theses (the so-called 'grey literature').

(b) Much acupuncture research has been reported in languages other than English, particularly Chinese, Japanese and Korean. The major databases used for literature searches contain predominantly English language journals. In order to ensure that searches do not omit eligible data, reviewers must make every attempt to search non-English databases or journals. Translation skills should be available for the reviewer team.

(2) Poorly described or inadequate inclusion criteria

Criteria for inclusion should be clearly stated in the methods section of the review and rigorously applied, in order to avoid bias in performance. For example, abstracts with insufficient descriptions of research design and little data contribute little to a systematic review.[14] Every effort should be made to specify as accurately as possible the research questions that are being addressed, and the outcomes of interest which will determine what the inclusion criteria should be.

(3) Lack of definition of 'acupuncture'

There are many different approaches to acupuncture and each practitioner will have a view on what is and what is not 'acupuncture'. Individual views will largely be formed from personal experience, since there seems no basis on which to form an objective authoritative view. For example, does a single needle inserted into a myofascial trigger point count as acupuncture? Does it count if the point was identified as an *Ah Shi* point? Is 'laser acupuncture' included? On the one hand, it could be regarded as sufficient if the author specifically names the therapy as acupuncture, but on the other hand some interventions are so far from what is currently taught as 'traditional Chinese acupuncture' that they may not count. Remembering that the primary rationale for a systematic review is to improve patient care rather than to validate any particular therapeutic approach, and that the practice of acupuncture combines many different treatment approaches, it seems that authors of reviews might choose to include all forms of treatment which might be described to patients as 'acupuncture'. In any case, the styles of acupuncture included for review should be explicitly stated and, where possible, a secondary analysis should be performed of RCTs analysing each style in a separate sub-group.

Sub-group analyses should be considered in a further instance. The term acupuncture is often used in a broader sense when referring to treatment given in conjunction with other interventions such as moxibustion, dietary advice, etc. While it is understandable that investigators wish to reflect actual practice in clinical research, it is not meaningful to combine in a single analysis studies that have used different varieties of intervention.

(4) Inconsistent assessment of quality of studies

Certain aspects of quality are generally applicable to all reports of primary research. Randomization, blinding and reporting of dropouts and withdrawals are three design features that have been shown empirically to affect study quality. These features are commonly assessed by the Jadad scale, which has been validated.[15] Studies that do not include one of these features have a clear positive bias in their outcome.[16]

As one author pointed out, the use of the Jadad scale for trials of acupuncture appears at first sight to raise a particular issue.[7] This is that the term 'double-blinding', interpreted by some researchers as the blinding of patient and therapist, counts for two out of the total of five points, yet 'double blinding' seems virtually impossible in acupuncture studies.[7] Thus the maximum score that can be awarded for an acupuncture study appears to be three points (one each for use and appropriateness of randomization and one for reporting dropouts and withdrawals). However, it is now clear that the term 'double-blinding' is not used consistently even in the orthodox literature.[17] Moreover, Jadad intended the term to apply to blinding of (1) patient and (2) the assessor (not therapist).[18] Thus, acupuncture studies are eligible for the maximum two points for blinding as was allowed in one earlier review.[19] The formal Jadad criteria[15] should be revised by dropping the requirement to include the actual words 'double-blind'.

Jadad also published four additional criteria[18] that have been used in reviews of acupuncture.[20] Were co-interventions avoided or controlled for? Was compliance satisfactory? Was the study population adequately homogeneous? Was the therapeutic time equivalent between groups? It is noted that compliance and homogeneity require judgement and are therefore less reliable criteria for quality.

Various authors have suggested other methods of assessing the quality of acupuncture studies.[21–24] Before applying these checklists, however, they need to be carefully validated. It has been shown that different checklists for quality applied to the same set of studies can result in entirely opposite conclusions about the quality of the studies.[25] In conclusion, it seems best to rely on the one method that is known to be valid, i.e. that of Jadad, until other methods that may be more sensitive to acupuncture methodology have been validated.

(5) Lack of assessment of the adequacy of acupuncture

A quality assessment criterion not included in the expanded Jadad scale is the completeness of reporting of the acupuncture protocol. This concern is among those addressed in the new STRICTA (**ST**andards for **R**eporting **I**nterventions in **C**ontrolled **T**rials of **A**cupuncture) guidelines described elsewhere in this issue of *Clinical Acupuncture and*

Oriental Medicine. Clearly, a trial outcome can be questioned if there is insufficient information about the treatment parameters, whether the results favour acupuncture or not. Two attempts to assess 'quality of acupuncture' by asking 'expert' acupuncturists have met with limited success.[26,27] Reasons include divergent views on the appropriateness of different approaches, as well as disagreement on fundamental technical issues such as which points should be used, what depth needles should be inserted, and the comparative importance of these and other issues in giving a treatment.

The formulation of reliable criteria for quality analysis of acupuncture trials is complex. The risks of simply listing criteria have been mentioned and a reliable instrument will require empirical testing, in a similar method as the Jadad score discussed above. While such validated criteria will be helpful, only comparison in clinical trials will solve questions about which approach is of greatest value to the patient in any given condition. New studies should involve a carefully developed treatment schedule based on either, or preferably both, consensus of practitioners and searches of the literature and traditional acupuncture texts.

(6) Impact of low sample size

Many acupuncture studies have been conducted in private practice or with limited resources, which has restricted the number and suitability of the patients included. This leads to under-powered studies and the risk of false-negative conclusions (type II errors). Reviewers should be aware of this risk and deal with it in a manner that does not introduce bias, for example by using the term 'inadequate' rather than 'negative' when formulating their conclusions.

(7) Inappropriate evaluation of results

(a) Sub-group analysis. As with any therapy, trials should be evaluated based on the research question asked. For example, clinical trials comparing acupuncture to a sham needling control should be assessed separately from trials employing usual care as an 'active control' arm. Not only are issues of blinding different in these two research designs, the conclusion that there is 'no difference' between the groups has a very different meaning in the two cases. Thus, separate sub-group analyses should be specified in advance to evaluate trials with different control groups.

(b) Best evidence synthesis. In performing systematic reviews, it is clear in many instances that the data cannot be combined since the outcomes of interest are quite different from study to study. In these instances, the best evidence synthesis method proposed by Slavin can be used.[28] This approach involves four levels of increasingly acceptable evidence ranging from Level 1 'strong evidence' based on multiple, relevant, high-quality RCTs with generally consistent outcomes, to Level 4 'no evidence' resulting from only one relevant, low quality RCT, no relevant high-quality RCTs or RCTs with inconsistent outcome. This method, of combining the results based on the consistency of the RCTs identified, has been used in some systematic reviews of acupuncture.[29,30] However, the decisions to determine the strength of evidence with the best evidence synthesis method are somewhat arbitrary and may be open to bias as they depend on the proportion of positive to negative studies. This qualitative approach can, therefore, be influenced by publication biases. It is believed that negative trials were more likely not to be published despite sound methodological design in these studies.[19] Thus, for those treatments found to be effective, it is possible that negative trials that have not been identified could change the direction of the conclusion. In consequence, this method of quality assessment may pose confusion for clinicians seeking the results of systematic reviews to guide their clinical decisions if the recommendations change from time to time due to shifts in the critical proportion for effectiveness.

(c) Sensitivity analysis. One way to improve a review of studies that are of variable quality is to perform a sensitivity analysis. This involves comparing subgroups of studies according to any features that may reasonably affect the treatment outcome. For example, patients in a trial should receive a general level of sufficient treatment, even if there is disagreement on what is 'sufficient'. Ezzo and colleagues assessed the effect of number of treatment sessions, number of points needled, whether *de qi* was elicited, and whether acupuncture treatments were standardized or individualized.[29] The number of treatment sessions was the only variable found to be significantly associated with treatment outcome, with six or more treatments being more effective than less than six. In a similar approach for studies of acupuncture for smoking cessation, no statistically significant influence was found for type of acupuncture or number of attendances, though the number of studies involved was too small to rule out false-negative results.[31] Sensitivity analysis should be considered as the way to make best use of the limited information available in acupuncture studies and to provide some indication of which treatment variable should be included in future studies.

CONCLUSIONS

Examination of systematic reviews that have assessed the effectiveness of acupuncture for a wide range of conditions has identified a number of weaknesses that must be considered when performing future reviews of this type. Systematic reviews of acupuncture require knowledge and application of the general methods of reviews as well as awareness of special issues regarding this traditional treatment modality, including the diversity of practice styles considered as acupuncture, difficulties in determining adequacy of treatment, and the wide range of control procedures employed in such trials. Although the quality of systematic reviews of acupuncture has improved in recent years, with increasing attention given to the above issues, the need for the near future is not for additional systematic reviews but for new, high-quality RCTs of acupuncture that meet the design and reporting standards called for in recent articles[32–34] and in the STRICTA guidelines mentioned above.

REFERENCES

1. Sackett D.L., Rosenberg W.M., Gray J.A., Haynes R.B., Richardson W.S. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–72.
2. Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). BMJ 1997; 315:672–675.
3. Mulrow C.D. Rationale for systematic reviews. BMJ 1994;309:597–599.
4. Antman E.M., Lau J., Kupelnick B., Mosteller F., Chalmers T.C. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992;268:240–248.
5. Jadad A.R., Cook D.J., Browman G.P. A guide to interpreting discordant systematic reviews. Can Med Assoc J 1997;156:1411–1416.
6. Cummings M. Teasing apart the quality and validity in systematic reviews of acupuncture. Acupunct Med 2000;18:104–107.
7. Birch S. Systematic reviews of acupuncture – are there problems with these? Clin Acupunct Oriental Med 2001;2:17–22.
8. Ezzo J., Lao L., Berman B. Assessing clinical efficacy of acupuncture: what has been learned from systematic reviews of acupuncture? In: Stux G., Hammerschlag R., eds. *Clinical Acupuncture: Scientific basis*, pp 113–130. Berlin: Springer-Verlag, 2001.
9. Oxman A.D. Checklists for review articles. BMJ 1994;309:648–651.
10. Cook D.J., Sackett D.L., Spitzer W.O. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam consultation on meta-analysis. J Clin Epidemiol 1995;48:167–171.
11. Moher D., Cook D.J., Eastwood S., Olkin I., Rennie D., Stroup D.F. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896–1900.
12. Oxman A.D., Guyatt G.H. Validation of an index of the quality of review articles. J Clin Epidemiol 1991;44: 1271–1278.
13. Moher D., Pham B., Jones A., Cook D.J., Jadad A.R., Moher M. et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609–613.
14. Ernst E., White A.R. Acupuncture for back pain: a meta-analysis of randomized controlled trials. Arch Intern Med 1998;158:2235–2241.
15. Jadad A.R., Moore R.A., Carrol D. et al. Assessing the quality of reports of randomised clinical trials: is blinding necessary? Controlled Clin Trials 1996;17: 1–12.
16. Schulz K.F., Chalmers I., Hayes R.J., Altman D.G. Empirical evidence of bias. JAMA 1995;273:408–412.
17. Park J., White A.R., Stevinson C., Ernst E. Who are we blinding? A systematic review of blinded controlled trials. Perfusion 2001;14:296–304.
18. Jadad A.R. Randomised controlled trials. London: BMJ Books, 1998.
19. White A.R., Ernst E. A systematic review of randomized controlled trials of acupuncture for neck pain. Br J Rheumatol 1999;38:143–147.
20. Ezzo J., Hadhazy V., Birch S., Lao L., Kaplan G., Hochberg M. et al. Acupuncture for osteoarthritis of the knee: a systematic review. Arthritis Rheum 2001;44: 819–825.
21. Ter Riet G., Kleijnen J., Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. J Clin Epidemiol 1990;1191–1199.
22. Birch S. An overview of acupuncture in the treatment of stroke, addiction and other health problems. In: Stux G., Hammerschlag R., eds. *Clinical Acupuncture: Scientific basis*, pp 148–155. Berlin: Springer-Verlag, 2001.
23. Hammerschlag R., Morris M.M. Clinical trials comparing acupuncture with biomedical standard care: a criteria-based evaluation of research design and reporting. Complement Ther Med 2001;5:133–140.
24. Smith L.A., Oldman A.D., McQuay H.J., Moore R.A. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. Pain 2000;86:119–132.
25. Juni P., Witschi A., Bloch R., Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054–1060.
26. Linde K., Worku F., Stör W., Wiesner-Zechmeister M., Pothmann R., Weinschütz T. et al. Randomised clinical trials of acupuncture for asthma: a systematic review. Forsch Komplementërmed 1996;3.
27. White A.R., Ernst E. A trial method for assessing the adequacy of acupuncture treatments. Altern Ther Health Med 1998;4:66–71.
28. Slavin R.E. Best evidence synthesis: an intelligent alternative to meta-analysis. J Clin Epidemiol 1995;48:9–18.
29. Ezzo J., Berman B., Hadhazy V., Jadad A.R., Lao L., Singh B.B. Is acupuncture effective for the treatment of chronic pain? A systematic review. Pain 2000;86:217–225.
30. Tulder M.W.v., Cherkin D.C., Berman B., Lao L., Koes B.W. Acupuncture for low back pain (Cochrane Review). *Cochrane Library, Issue 1*, Oxford: Update Software, 2000.
31. White A.R., Resch K.L., Ernst E. A meta-analysis of acupuncture techniques for smoking cessation. Tob Control 1999;8:393–397.

32. Ernst E., White A.R. A review of problems in clinical acupuncture research. Am J Chin Med 1997;25:3–11.
33. Lao L., Ezzo J., Berman B., Hammerschlag R. Assessing clinical efficacy of acupuncture: considerations for designing future acupuncture trials. In: Stux G., Hammerschlag R., eds. *Clinical Acupuncture: Scientific basis*, pp 187–210. Berlin: Springer-Verlag, 2001.
34. Stux G., Birch S. Proposed standards of acupuncture treatment for clinical studies. In: Stux G., Hammerschlag R., eds. *Clinical Acupuncture: Scientific basis*, pp 171–186. Berlin: Springer-Verlag, 2001.

