THE JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE
Volume 10, Number 3, 2004, pp. 481–498
© Mary Ann Liebert, Inc.

Clinical Research on Acupuncture: Part 2. Controlled Clinical Trials, an Overview of Their Methods

STEPHEN BIRCH, Ph.D., Lic.Ac. M.B.Ac.C.

ABSTRACT

There is almost universal agreement that the quality of clinical trials of acupuncture is poor. There is an urgent need to improve their quality. The author develops here a list of 45 criteria important in the design, implementation, and writing up of controlled clinical acupuncture trials. This list has been compiled after examining the quality assessment criteria used in meta-analyses and systematic reviews of acupuncture, general publications on clinical trial designs and methodological considerations specific to acupuncture trials. Each criterion is discussed with recommendations about use and implementation. Additionally, each criterion is discussed relative to their importance in three types of acupuncture trial, acupuncture versus sham acupuncture, acupuncture versus standard therapy and acupuncture versus no treatment or wait-list. It is hoped that this exploration and systematic presentation of the 45 criteria will contribute to improving the quality of clinical trials of acupuncture. Improved trial quality will lead to greater ease interpreting the results of trials, especially in systematic reviews.

INTRODUCTION

There is almost unanimous agreement that controlled clinical trials of acupuncture have suffered from multiple flaws, and that their methods need to be improved NIH Consensus Conference. Acupuncture, 1998; British Medical Association, 2000; Ernst 1994; Ernst and White, 1996, 1997, 1998, 1999; Ezzo et al., 2000, 2001a; Hammerschlag and Morris, 1997; Kleijnen et al., 1991; Lewith and Machin, 1983; Linde et al., 1996; Lytle, 1993; Mayer, 2000; McLellan et al., 1993; Melchart et al., 1999; Patel et al., 1989; Richardson and Vincent, 1986; ter Riet et al. 1990a, 1990b; Tait et al., 2002; van Tulder et al., 1999; Vickers et al., 2002; Vincent, 1993; Vincent and Richardson, 1987; White and Ernst, 1999; White et al., 1999. There is an urgent need to develop better methods, to apply these methods systematically so that better-quality studies can be conducted, results of studies can be more clearly interpreted and accepted, and fairer comparisons made in meta-analyses or systematic reviews of those studies.

The author has written on a range of themes related to clinical trials of acupuncture. These include issues in publications, definitions, conceptual frameworks, and diversity of practice in the field, and their potential impact on clinical research (Birch 1994, 1995b, 1998, 2003c; Birch and Felt, 1999; Birch and Kaptchuk, 1999; Birch and Tsutani, 1996,* problems in the matching of study design to research questions (Birch 2002b, 2003a, 2003b, 2003e), problems and solutions in the adequacy of treatment in acupuncture clinical trials (Birch 1995a, 1997a, 2003a; Stux and Birch 2001),* problems and solutions in the use of sham acupuncture methods and controlling for the nonspecific effects of treatment (Avants, et al., 1995; Birch, 1995a, 1997a, 1997b, 2003a, 2003e, 2003f; Birch and Jamison, 1998; Birch et al., 2002; Margolin et al., 1995, 1997b; Stux and Birch, 2001),* methods of locating and testing treatment and sham

Foundation for the Study of Traditional East Asian Medicine, Amsterdam, The Netherlands.

*Birch S. An Exploration with Proposed Solutions of the Problems and Issues in Conducting Clinical Research in Acupuncture. Exeter University [doctoral thesis] 1997d.

acupuncture points (Falk et al., 2000; Margolin et al., 1995, 1997b), problems and solutions in the testing of traditional forms of acupuncture (Birch, 1997c, 2003g; MacPherson et al., 2002a), validating diagnosis and treatment selection in traditional forms of acupuncture (Birch and Sherman, 1999; MacPherson et al., 2002a),* problems in the interpretation of clinical trials and systematic reviews of acupuncture (Birch 2000, 2001a, 2001b, 2003a, 2003e, 2003h; Birch et al., 2004; Ezzo et al., 2001b; White et al., 2002), summaries of the treatment (Birch and Sherman, 1999), and clinical trial literature (Birch 2001, 2002a; Birch et al., 1996, 2004; Ezzo et al., 2001b) reviews of (Birch et al., 2004) and problems in the interpretation of reports of adverse effects of acupuncture (Margolin et al., 1997a), and strategies for research in the acupuncture profession (Birch 1994, 2003d).

Drawing together themes developed in these publications, this paper attempts a comprehensive overview of the methods used in clinical trials of acupuncture by systematically describing the components necessary for conducting them. It is hoped that this will be a useful resource for others conducting and writing up controlled clinical trials of acupuncture.

In any clinical trial there are always a number of important items that must be attended to: recruiting the desired number of relevant patients, having sufficient personnel and funds available to conduct the study, consulting with appropriate statistical experts prior to submitting the study for approval or funding, using appropriate validated outcome measures, et cetera. Many texts can be consulted about designing clinical trials (see for example, Meinert, 1986). There are international standards in medical journals that give guidance and checklists for study publication (Altman et al., 2001, AUA 1998; Begg et al. 1996, International Committee of Medical Journal Editors, 1991; Moher et al., 2001), and some recommendations about the design of studies in the field of complementary medicine (Vickers, 2002). The fact that the majority of acupuncture trials have suffered from multiple methodological flaws indicates that simply following these recommendations from medical publications will not be sufficient to address the complexities of conducting acupuncture trials. There are a number of specific issues that also need to be considered and addressed for acupuncture trials in addition to basic study design criteria.

The author pointed out that methods selected in clinical trials of acupuncture are not always suitable for the research question asked (Birch 2002b, 2003b). In addition to ensuring the selection of an appropriate research method to answer the study question, it is essential that the methodology of clinical trials of acupuncture be appropriate and sufficient. A number of papers have been written about the design of acupuncture studies (Birch, 2003b; Hammerschlag, 1998; Hopwood and Lewith, 2003; Lao et al., 2001; Lewith and Vincent, 1996; Margolin et al., 1998a, 1998b; Schnyer and Allen, 2001; Sherman and Cherkin, 2003; Vincent, 1989; Vincent and Lewith, 1995; White, 2002; White and Park, 1999; White et al., 2001; Zaslawski, 2003), each of which details some of the issues that need to be addressed. Below is a more extensive discussion of the design of controlled studies of acupuncture. It is a list with discussions and proposed solutions to a number of major design criteria important for clinical studies of acupuncture. It represents an integration and extension of criteria used in the evaluation of clinical trials of acupuncture.

The list began as an expansion (Birch and Jamison 1998)* of the design criteria by which the Dutch reviewers ter Riet, Kleijnen, and Knipschild evaluated the quality of clinical trials in the late 1980s (Kleijnen et al., 1991; ter Riet et al., 1990a; 1990b). While there are problems with the reviews published by these researchers, not the least of which was improper judgments about efficacy (Morris, 1996), the criteria that they listed for evaluating the quality of clinical trials are themselves a useful framework for designing clinical trials of acupuncture. Hammerschlag and Morris (1997) developed a separate list of criteria for evaluating studies that compared acupuncture to standard care drawing from published sources, a number of these are incorporated. Systematic reviews following the Cochrane Collaboration methodology have used the five-point Jadad scale (Jadad, 1998) to evaluate the quality of acupuncture studies (Ernst and White, 1998; Ezzo et al., 2000; Melchart et al., 1999), these five criteria are also addressed in the list below. Recently various groups have listed criteria that should be described in clinical trials of acupuncture to help ensure adequate treatment and complete reporting of the treatment interventions (Elorriaga Claraco et al., 2003a, 2003b; MacPherson et al., 2002b; Romoli et al., 2003; Stux and Birch, 2001), relevant criteria are incorporated in the list below. Several authors have written about trials that examine traditional forms of acupuncture (Birch, 1997c, 2003g; MacPherson et al., 2002a) their recommendations are included where appropriate. Finally, a number of other sources on the reporting or evaluation of clinical trials were examined to see if any other criteria should be included (Altman et al., 2001; Bailar and Mosteller, 1992; Chalmers and Altman, 1995; Moher et al., 2001; Sackett et al., 1991; Tait et al., 2002) giving the final list of 45 criteria. This paper discusses each and offers possible solutions for many that if addressed would ensure high-quality clinical trials. Within each criterion there are sometimes several items discussed so that the total number of items of interest is considerably more than 45. For example under criterion 29—diffuse noxious inhibitory control (DNIC)—the following 6 items are discussed: nature DNIC control; DNIC effects may not be only needling effects, need exploration of other effects; attempt to match intensity of needling effects; need to assess intensity or perceptions of needling effects; need validated scale to do this; may need acupuncture naïve patients for sham acupuncture studies.

In addition to discussing each criterion, those criteria thought necessary for the three most common and impor-

tant acupuncture study designs are highlighted: (1) acupuncture treatment, compared to some form of sham acupuncture treatment, with attempts to control for placebo and other nonspecific effects (acupuncture treatment versus sham acupuncture design); (2) acupuncture therapy compared to standard therapy, with no efforts to control for placebo and other treatment related nonspecific effects (acupuncture treatment versus standard treatment design), and (3) acupuncture treatment compared with no treatment or wait list group (delayed treatment group), with no efforts to control for placebo and other treatment related nonspecific effects (acupuncture treatment versus no treatment/wait list group design). Table 1 lists those items important in the design of acupuncture treatment versus sham acupuncture, acupuncture treatment versus standard treatment, and acupuncture treatment versus no treatment/wait list group, with items that would benefit from pilot studies highlighted. It is hoped that at the very least the list of criteria serves as a checklist for designing and writing up clinical trials of acupuncture, and possibly could inform future systematic reviews of published studies.

CLINICAL TRIAL DESIGN CRITERIA AND THEIR PROPOSED SOLUTIONS

1. Clear statement of study question, how chosen design can answer the question and citation of relevant literature

Acupuncture has been tested in a variety of different research models, for example in comparison to a sham acupuncture, standard therapy, and a wait-listed treatment. In a recent publication the author showed that there has been a problem with the selection of research models in controlled clinical trials of acupuncture (Birch, 2003b). Of the 29 models examined, as many as 10 were either clearly unable to answer the questions for which they were chosen or it remains questionable and controversial as to how well or whether they can answer the questions for which they were selected (Birch 2002b, 2003b). It is thus important that researchers clearly explain their research questions and how the research models they select can answer those questions, with greater care paid to ensuring the matching of study model to research question (MacPherson et al., 2002b). There have been the additional problems of not reading (Birch, 2003e, 2003f) and misreading (Birch, 2003a) the relevant literature pertinent to the selected study models. This has contributed to the continued use of inappropriate research models in clinical trials of acupuncture, or the use of inadequate methods. Designing and conducting clinical trials of acupuncture is sufficiently complicated that research teams should be much more careful about what they are doing, it is easy to make faulty assumptions that lead to poor quality trials or uninterpretable results (Birch, 1998). One area that can help improve trials is the requirement that the research team read widely and cite relevant literature (see also items 19, 22, 24, and 31). Showing a broad knowledge of the area can help minimize these problems.

2. Informed consent obtained

Any clinical trial involving human subjects must follow international standards for conducting clinical research. Obtaining informed consent is an important component of these standards. When writing up the study, state that informed consent was obtained.

3. Consent form descriptions

The issue of assessing nonspecific effects of treatment so as to be able to control for them (see items 29 and 30) is further complicated by how the consent form for the study is worded. Some ethics committees or Institutional Review Boards (IRB) insist on full and explicit disclosure of what treatments a patient may be randomized to. The more explicit the description in the consent form, the more likely that the patient will be thinking about what treatment they are receiving (Birch, 1997b; Zaslawski, 1997). Some research has already demonstrated how the informed consent can affect treatment outcome (Kleijnen and de Craen, 1996). The author has encountered a variety of different IRB responses in the studies that he has worked on, ranging from not allowing the use of any term such as “sham” acupuncture because it implied a deceit, to full and explicit description of the control needling procedure with full statement of the hoped for lack of effectiveness as a “sham” and even “placebo” treatment. In a study that uses a “sham transcutaneous electrical nerve stimulation (TENS)” unit as the control, the wording of the consent form can be critical, as ethical considerations might force such an explicit description of the procedure so as to render it useless in sham-controlled studies. In the United States it may be difficult to obtain ethical approval for use of a sham TENS unit as a control procedure because of the demand for explicitness in the consent form, but in the United Kingdom a number of studies have been possible with the sham TENS control (Birch 2003b, 2003f; Vincent and Lewith, 1995). This highlights the influence of cultural and political differences in research method selection, which can affect how research is conducted. Assuming that the wording of the consent form may have an impact on how the patients perceive the treatments they receive and how they think about those treatments, especially when asked to consider them, variation in the explicitness of the consent form descriptions produce variations that may also need to be accounted for (Birch, 1997b; Margolin et al., 1998a; White, 2002; Zaslawski, 1997), and, in certain cases could make it difficult to conduct the study at all because of perceived ethical issues (Kleijnen and de Craen, 1996). It is important to include relevant statements from the consent form in the published study.

Table 1. Design Criteria Needed for Different Study Designs

Criteria	*AT vs. SA/AT vs. ST/AT vs. NT/WL*
1. Clear statement of study question, how chosen design can answer the question & citation of relevant literature	Important for AT vs. SA, AT vs. ST and AT vs. NT/WL studies
2. Informed consent obtained	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
3. Consent form description	Important in AT vs. SA studies. But can also be important in AT vs. ST, and AT vs. NT/WL studies
4. Homogeneity/inclusion & exclusion criteria presented[a]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
5. Pre-stratification[a]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
6. Randomization employed	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
7. Recruitment method and treatment setting described	Important especially for AT vs SA studies, but good to do for AT vs. ST and AT vs. NT/WL studies
8. Patients blinded	Important only for AT vs. SA studies
9. Treatment groups separated	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
10. Evaluator blinded	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
11. Blinding of data-handling study personnel	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
12. Patient demographics presented/comparability of baseline characteristics shown[a]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
13. Rejection data presented	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
14. Adequate sample size[a]	Important for all studies, AT vs. SA, AT vs. ST, and AT vs. NTWL
15. Sample size calculation presented	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
16. Sample size calculations based on data and not estimates[a]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
17. 20% or less loss to follow up/Reduce loss to follow-up[a]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
18. Adequate description of acupuncture procedure	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
19. Thorough citing of the relevant literature	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
20. Adequacy of test treatment[a]	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
21. Acupuncture training stated	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
22. Checking expertise against familiarity with the clinical literature	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
23. Appropriateness of the "sham" control treatment[a]	Important only in AT vs. SA studies
24. Adequate and appropriate existing treatment in control group with protocol described/clinical trial of standard care cited	Important for AT vs. ST studies
25. Standard care (e.g., medication) compliance monitored	Important for AT vs. ST studies
26. Existing treatment in reference group	Important in AT vs ST studies, and may be important in AT vs. SA studies, depending on factors discussed in text
27. Use of untreated reference group	Important for AT vs. NT/WL studies and some AT vs. SA studies
28. Generalizability of the results of the study	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
29. DNIC control[a]	Only important in AT vs. SA studies
30. Controlling for the non-specific effects of treament through appropriate assessments[a]	Important only for AT vs. SA studies
31. Steps taken to validate TBSA described[b]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
32. Reliability of TBSA assessments described[b]	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
33. Follow up $\geq$3 months after treatment (chronic problems)	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
34. Endpoints described	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
35. Endpoint statistics presented/Reader given opportunity to do inferential statistics	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
36. Onset treatment effects compared	Important in AT vs. SA studies, and AT vs. SA studies
37. Duration treatment efects cmpared	Important in AT vs. SA studies, and AT vs. SA studies
38. Follow-up data presented	Important in AT vs. SA, AT vs. SA, and AT vs. NT/WL studies
39. Assessment of dropout patients and presentation of data	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
40. Duplicate assessments employed	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
41. Use of medication	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
42. Tracking other treatments	Important in AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
43. Activities of daily living	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
44. Remarks on side effets	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies
45. Funding sources acknowledged	Important for AT vs. SA, AT vs. ST, and AT vs. NT/WL studies

[a]Indicates that the criterion would benefit from or is dependent on prior testing in pilot studies.

[b]Only relevant if a study utilizes the individualized treatment approaches of a traditionally based system of acupuncture (TBSA).

AT, acupuncture therapy; SA, sham acupuncture; ST, standard therapy; NT, no treatment; WL, wait list; DNIC, diffuse noxious inhibitory control.

4. Homogeneity/inclusion and exclusion criteria presented

As ter Riet and colleagues (1990a) note, ensuring homogeneity of the patient population can be a difficult criterion to adequately deal with. To ensure a homogenous sample, it is important to have adequate inclusion and exclusion criteria that are thoroughly applied. However, to properly exclude all patients with similar complaints but not exactly the precise condition under consideration can be time consuming and difficult. For example, if a study attempts to exclude patients with severe depression, while allowing those with mild depression to participate, or attempts to exclude all patients with a psychiatric condition, it becomes necessary to screen every patient with the Structured Clinical Interview for *DSM* (SCID) part of the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition* (*DSM IV*) to ensure correct exclusion, this can be very time consuming. Multiple exclusion criteria can also be costly to apply, which can strain resources in small studies. Careful consideration of this issue at the pilot study stage can help, by examining how broadly a particular treatment can be applied. Additionally the criteria should allow recruitment of patients for whom the main outcome measures used in the study are most suitable. The study should develop appropriate inclusion and exclusion criteria in patient selection and should state those criteria in the published report.

5. Prestratification

Prestratification refers to the process of stratifying patients at randomization according to important criteria that could impact the outcome such as gender, prior experience with acupuncture treatment, et cetera. If women tended to respond better to treatment than men, it would be important to ensure an equal distribution of female and male patients during randomization to the different treatment arms in the study. Prestratification requires that one has already identified important variables that need to be accounted for to ensure randomization procedures are adequate. Often the criteria about which stratification can be important are not clearly known without first conducting pilot studies. Thus it can be important to conduct pilot studies prior to going to a full clinical trial.

6. Randomization used

Questions have begun to emerge about the effects of randomization on treatment outcome, and therefore the validity of randomization in clinical research [Kaptchuk et al., 1996). Yet other questions have been raised about the efficiency of randomization at distributing important variables equally among treatment groups (Aickin, 2003), with suggestions that it should be replaced by other methods such as the design adaptive allocation method (Aickin, 2003; Walach 2003). But these ideas have not yet reached a critical mass to change the way that medical research is usually conducted. Thus, because it is still generally accepted that randomization is an important method of controlling for a range of variables that can interfere with the interpretation of clinical trial results (Resch and Ernst, 1996), it remains important to apply randomization procedures in clinical research. The five-point Jadad scale, a summary analysis of the quality of clinical trials scores one point for use of randomization and one for using an appropriate randomization procedure, thus scoring two out of a possible total of five points for use of randomization (Ernst and White, 1998; Ezzo, et al., 2000; Melchart et al., 1999). Because this scale is currently used in evaluating the quality of clinical trials, it is thus important that a proper randomization procedure is used and that it is well described in the final published report. If an alternate to randomization such as the design adaptive allocation method is used (Aickin, 2003; Walach, 2003), it will probably be necessary to ensure that it will be scored positively in place of randomization on quality assessment scales such as the Jadad scale.

7. Recruitment method and treatment setting described

Many critics of acupuncture believe that it works primarily because of its nonspecific effects. In the published study it is therefore necessary to describe important variables that could produce bias and thereby increase the nonspecific effects of treatment in the study. Among these are the methods of recruitment and the settings in which treatments are administered. It may also be necessary to make ongoing assessments at the locations where treatment are provided to be able to demonstrate minimization of bias (Avants et al., 2000; Margolin et al., 1998a).

8. Patients blinded

In studies in which patients are assigned to receive one of two acupuncture treatments, it is important that the patients be blinded as to which they receive. It is not enough to simply state that the patients were blinded to which group they were in: it must be demonstrated (Birch and Jamison, 1998; Zaslawski et al., 1997). Whether the patients were blinded to which treatment they received or not should be determined by using a questionnaire to assess this (Vincent 1990a; Vincent and Lewith, 1995; White, 2002) or simply asking patients which treatment they think they received (White et al., 1996; Zaslawski et al., 1997). In addition, other placebo-related nonspecific effects should be assessed so that they can also be controlled for (see items 29 and 30) (Birch 1995a, 2003f; Birch et al., 2002; Birch and Jamison, 1998).

9. Treatment groups separated

This refers to keeping the treatment groups separated to prevent exchanges of information between patients in each

group. This helps prevent contamination of data between treatment groups and helps maintain the blind in acupuncture treatment versus sham acupuncture type studies.

10. Evaluator blinded

It is important in clinical trials that the evaluator be blinded to treatment assignment. In acupuncture this is especially important because studies cannot be conducted with the therapist blinded (British Medical Association 2000; Vincent, 1993). The use of a blinded evaluator, with blinded patients is a recommended study design for acupuncture trials (Lao et al., 2001; Vincent 1993) and is an acceptable alternate to the standard double-blinded study where the patient and therapist are blinded to treatment assignment (Jadad, 1998).

11. Blinding of data-handling study personnel

In addition, it is ideal that the personnel responsible for entering the data into a computer, checking the data for accuracy and completion, and the statisticians conducting the data analyses be blinded wherever possible to patient assignment so as to eliminate any other potential sources of bias (Hammerschlag and Morris, 1997; Margolin et al., 1998a). This is not always possible, but by eliminating or reducing potential bias it improves the final interpretation of results if it can be achieved.

12. Patient demographics presented/comparability of relevant baseline characteristics shown

Patient demographics should be presented so that the baseline characteristics of patients in each group can be clearly shown in the final published report. There should not be significant differences between treatment groups at baseline. When such differences are found it is an indication of problems with the randomization procedure, and/or small sample size, and possibly the need for prestratification. Pilot studies could help identify demographic characteristics that can be dealt with in future studies by use of exclusion criteria or prestratification. In studies that attempt to control for the nonspecific effects of treatment, factors that can influence the size of these effects such as prior experience, beliefs and expectations should be assessed at baseline and described (see item 30) (Avants et al., 2000; Birch 2003f; Birch et al., 2002; Birch and Jamison, 1998).

13. Rejection data presented

It is important to briefly describe the number of study applicants that were rejected and reasons for those rejections. This allows reviewers to check proper administration of the study's inclusion and exclusion criteria, and to examine the study for possible sources of bias.

14. Adequate sample size

ter Riet and colleagues in their reviews set a requirement of 50 or more patients be assigned to each group (Kleijnen et al., 1991; ter Riet et al., 1990a, 1990b), this however seems arbitrary. If the study is a pilot study, then it is usual that only small numbers of patients are required for each group, because it is not necessary to achieve statistical significance and finding a trend in the results is sufficient. If the study is a full-scale clinical trial, prior indications of the possible treatment effect size is necessary. Hence pilot studies are usually important to help determine sample size. Additionally, sample size requirements can be heavily dependent on the type of control treatment used, which can also be strongly impacted by ethical concerns. For example, if the medical condition to be treated is relatively mild the ethics committee reviewing the study is more likely to allow testing of questions that require use of some form of sham acupuncture, but the more severe and progressive the disease, the harder to get the committee to agree on the use of sham control studies because of ethical concerns about withholding treatment from the patients (Zaslawski, 2003). Thus, if the condition being treated is severe, a more ethical study design may be to compare the acupuncture to standard treatment (acupuncture treatment versus standard therapy). In the acupuncture treatment versus standard therapy model two approaches can be used, the first simply compares acupuncture treatment to standard therapy, the second adds acupuncture treatment to standard therapy and compares these two to the standard therapy alone. Either way, the study does not have to demonstrate that acupuncture treatment is more effective than standard therapy, it only has to show equivalence. But to do this the sample sizes need to be quite large and the standard therapy must have been previously shown to be effective (Pocock, 2002).

15. Sample size calculation presented

Sample size calculations should be presented in the published report.

16. Sample size calculations based on data and not estimates

With preliminary data from pilot studies, calculations can be made so that sample sizes appropriate to the condition and test treatment under study can be chosen. The routine problem of inadequate sample sizes (NIH Consensus Conference. Acupuncture, 1998; Lewith and Machin, 1983; Lytle, 1993) argues for the use of pilot studies first to help ascertain the correct sample size. A common problem that is encountered in the acupuncture literature in acupuncture treatment versus sham acupuncture studies is the faulty assumption that the "sham" acupuncture is equivalent to placebo and will have a treatment effect of about 30% (Lewith and Machin, 1983). It has been estimated that the "in-

vasive sham" acupuncture is usually a complex combination of placebo and other nonspecific effects (Birch et al., 2002; Lewith and Machin, 1983), with an effect size of approximately 50% (Lewith and Machin, 1983). Despite publication of this issue 20 years ago, many studies using the acupuncture treatment versus sham acupuncture model have routinely undercalculated the sample size requirements of the study because they assumed a 70% versus 30% difference rather than the 70% versus 50% difference that usually occurs in this design. Furthermore, the placebo effects of different therapies can be quite variable (Kaptchuk et al., 1996; Kleijnen and de Craen, 1996; Moerman and Jonas, 2002) making assumptions about the expected size of placebo effects in sample size calculations questionable. An additional problem with sample size calculations is the use of calculations based on estimated effect size differences derived from the outcome measurements. In the absence of pilot data, researchers often base sample size calculations on the theoretical number of patients needed to demonstrate a significant difference between treatment groups in changes on the outcome measurement. Sample size is important to all types of acupuncture studies, but in general an acupuncture therapy versus sham acupuncture design may require a smaller sample size than an acupuncture therapy versus standard treatment design (Birch, 2003f; Lao et al., 2001). However, if the acupuncture is compared to a noninvasive sham sample size requirements will in principle be smaller than when acupuncture is compared to some form of invasive sham because the nonplacebo nonspecific effects should be less (Birch, 2003f).

17. Twenty percent or less loss to follow-up/reduce loss to follow-up

Efforts should be made to curtail dropout and loss to follow-up. Adequate numbers of well-trained staff are necessary to help ensure a low loss to follow-up. The number of 20% or less loss to follow-up used as a criterion by ter Riet and colleagues (1990a, 1990b) is potentially difficult because a control treatment that is truly ineffective might tend to cause those patients to drop out or be lost to follow-up (Margolin et al., 1998a). Pilot studies are often helpful in identifying risk factors that increase drop out, and for developing strategies to minimize drop out.

18. Adequate description of acupuncture procedure

A clear and thorough description of the acupuncture procedure should be given. This should include: points treated, needles used, depth of insertion, techniques used, how long retained, nature of any manipulation, if electrical stimulation is used, what device and/or amplitude, frequency, wave form and duration used, number and frequency of treatment sessions, et cetera (Stux and Birch, 2001; White et al., 2001). A *de facto* standard, the STRICTA guidelines has recently been published for the reporting of treatment information in clinical trials of acupuncture (MacPherson et al., 2002b) these guidelines should always be followed. Other authors have presented similar lists of items that should be reported (Elloriaga Claraco et al., 2003a; Romoli et al., 2003; White et al., 2001). Reviews have found that not only are treatments inadequately described, simply providing an adequate description of the acupuncture procedure is still not enough (Birch, 1997a; Elloriaga Claraco et al., 2003b). Many studies have administered inadequate treatments (Birch, 1995a, 2003a; Stux and Birch, 2001),* yet these treatments have been adequately described, but this does not make the treatment adequate. A more comprehensive solution is necessary.

19. Thorough citing of the relevant literature

Many clinical trials of acupuncture have not cited any reliable sources or have inadequately cited sources for their acupuncture treatments (Birch, 1997a, 2003a, 2003e). It is thus important as a matter of course, that appropriate sources be properly cited. Many studies have tested a really poor acupuncture treatment or an idiosyncratic treatment (Birch, 1995a, 1997a, 2003a; Stux and Birch, 2001). Citing relevant sources ensures that the researchers have at least read relevant sources. This helps readers make clearer decisions about adequacy of treatment (item 20) and training of the acupuncturist (items 21 and 22).

20. Adequacy of the test treatment

The adequacy of the test treatment must be guaranteed. Many studies appear to have administered inadequate treatment in terms of number of treatment sessions, number of points used, specific points used (Birch, 1995a, 1997a, 2003a; Birch and Felt, 1999). Ensuring that an adequate and proper treatment is administered is not an easy problem to solve, in part because of the diversity found in the field (Birch, 1998; Birch and Felt 1999; Birch and Sherman, 1999; MacPherson and Kaptchuk, 1997). It is essential in future clinical trials of acupuncture that realistic and viable solutions are found (Birch, 2001a; Molsberger et al., 2002; Sherman and Cherkin, 2003; Stux and Birch, 2001; White, 2002). A number of groups have attempted to determine treatment patterns and the adequacy of treatment for specific conditions in clinical trials, with varying results (Birch 1995a, 1997a; Birch and Sherman, 1999; Linde et al., 1996; Sherman et al., 2001a, 2001b, 2001c; Stux and Birch, 2001; White, 2002; White and Ernst, 1998; White et al., 2001).* In their systematic review of pain studies, Ezzo and colleagues, following earlier findings (Birch, 1995a)* found that a minimum treatment defined in terms of number of points treated and number of treatment sessions, was significantly associated with a positive outcome (Ezzo et al., 2000). This finding strongly supports the need to establish at least a minimum adequate treatment and better still an optimal adequate treatment (Stux and Birch, 2001). If a particular test treat-

ment has shown promising results in a pilot study, then this is a major support for the selection of that particular treatment, even if it is theoretically not the best or most adequate test treatment for the condition being investigated in the study (Birch, 1997a). Although traditionally trained practitioners might not agree with the use of PC-6 alone as a treatment for postoperative or chemotherapy-related nausea and vomiting (Maciocia, 1993), there is clear international agreement that this treatment has been shown to be effective despite it being theoretically inadequate (NIH Consensus Conference. Acupuncture, 1998; Birch et al., 2004; British Medical Association, 2000; Tait et al., 2002; Vickers et al., 2002). An additional strategy is to use the BRITS [Birch Relevant and Irrelevant Treatment Selection] method, which essentially involves extensive reviews of relevant literatures and the citing of all relevant sources consulted. This or another literature review method can be complemented as needed by practitioner surveys and expert panels to determine optimal treatment (Birch and Sherman, 1999; Sherman and Cherkin, 2003; Sherman et al., 2001c; Stux and Birch, 2001; White, 2002). The approach adopted will ensure that only acupuncture points generally agreed on as being appropriate for the condition being treated will be used, and that relevant techniques are also being used (see Birch 1995a; Stux and Birch, 2001, for details of the method).* Another approach is to develop a manualized approach for the treatment protocols (Schnyer and Allen, 2002). Here treatment protocols are developed based on literature reviews of relevant textbooks, published research, practitioner surveys, and expert panels (Schnyer and Allen, 2002; Sherman and Cherkin, 2003). The protocols that are developed are then manualized, which ensures a consistent description of the treatment protocols so as to ensure standardization, replicability yet flexibility of the treatments (see also items 21 and 32) (Schnyer and Allen, 2002). A recent publication has suggested the potential for problems with the accuracy of point location (Aird et al., 2002). Training prior to the start of a clinical trial, as is suggested in the manualization approach (Schnyers and Allen, 2002) can help resolve this potential problem. Exploring test treatments in pilot studies is always useful.

21. Acupuncture training stated

The qualifications of the acupuncturist should be described. However, it should be noted that simply mentioning the "good" qualifications of the acupuncturist (ter Riet et al., 1990a) is not a guarantee of quality treatment. Not only have many studies not described the qualifications of the acupuncturists (Hammerschlag and Morris, 1997; ter Riet et al., 1990a), many have described qualifications that sound inadequate (Stux and Birch, 2001), and many acupuncturists whose qualifications have been mentioned have then administered inadequate acupuncture (Birch, 2003a). The necessary qualifications need to be carefully established and fully stated (MacPherson et al., 2002b). International standards could be developed to set a minimum training requirement for acupuncturists who participate in acupuncture research studies, this is a complex area due to variation in training standards in different countries (Birch and Felt, 1999; Stux and Birch, 2001). Another important aspect of establishing the qualifications of the participating acupuncturist(s) in a study is to train them specifically for the study (Schneyer and Allen, 2002). This should involve training in the standards of the treatment to be given, including training and review of diagnosis and treatment techniques, point location, the application and use of the sham device in acupuncture therapy versus sham acupuncture studies. A manualized approach can be advantageous to train the acupuncturists for the study (Schnyer and Allen, 2002). This training can also help minimize differences in how the practitioners handle and interact with the patients, which is important to eliminate bias and improve credibility of the study (Margolin et al., 1998a; Schnyer and Allen, 2002).

22. Cross-checking expertise against familiarity with the clinical literature

One way to double check the stated "expertise" of the clinician administering the acupuncture is to see how familiar they are with the acupuncture literature, by examining the literature cited by the researchers. The more thoroughly the relevant literature is cited, the more likely the acupuncturist is familiar with the relevant treatment methods for the condition under study. While only a rough guide to clinical expertise, it does allow some additional assessment of how familiar the acupuncturist is with the literature, and hence the treatment of the condition under study. The use of the BRITS method or the manualized approach helps ensure some form of a minimum standard for the acupuncture treatment and ensures a thorough review of the literature prior to selecting the treatment (Birch, 1995a; Schnyer and Allen, 2002; Stux and Birch, 2001).

23. Appropriateness of the "sham" control treatment

Specific selection criteria should also be used to ensure that a truly appropriate treatment is selected for the sham acupuncture treatment. The type of sham treatment chosen is dependent on the research question asked (Birch, 2003b, 2003f) and the type of acupuncture treatment to be tested (Birch, 1995a, 2003b). The BRITS method can be used to help ensure that such a treatment is selected when using "irrelevant acupoints" for the control treatment (Stux and Birch, 2001). It is probable that a pilot study is necessary to investigate the appropriateness of the nonpoints when using nonacupoints as treatment sites (Birch, 1995a). Treatment of the same acupoints as used in the test treatment but with a different technique for the control treatment can answer questions about the specific effects of those techniques (Birch, 2003b, 2003f), but proba-

bly should be avoided as a general model for testing the efficacy of the test treatment, or used only after conducting extensive pilot studies to investigate this approach. This last approach essentially involves comparing two forms of acupuncture to each other, which is not usually the intention of researchers using this model (Birch, 1997a; Wyon et al., 1995). In general, if the control treatment is tested first in a pilot study, one can generate data about the relative effectiveness of the sham acupuncture treatment, which will assist in determining if indeed it is an appropriate sham treatment, and play an important role in sample size calculations. Recently a number of noninvasive sham controls have been developed and tested (Fink et al., 2001; Karst et al., 2003; Park et al., 2000; Sherman et al., 2002; Streitberger and Kleinhenz, 1998). While these hold promise in some respects, they also have limitations in what they can be used for, thus they should be used only when it is clear that their use matches the question for which that sham treatment model is being selected (item 1) (Birch, 2003f). (For details about selection of various sham controls see Birch, 2003b, 2003f; Birch et al., 2002; Hammerschlag, 1998; Hopwood and Lewith, 2003; Lewith and Vincent; 1996; Vincent and Lewith, 1995; White, 2002; White et al., 2001). Another important issue with all sham treatments is to avoid the mislabeling of the sham treatment as a placebo treatment (Birch et al., 2002). All of the sham interventions tested so far are thought to activate physiologic nonplacebo nonspecific effects and some activate specific effects (Birch, 2003f; Birch et al., 2002; White, 2002).

24. Adequate and appropriate existing treatment in control group with protocol described/clinical trial of standard care cited

In a study that compares acupuncture to an existing or standard treatment or where acupuncture is added to that existing treatment, it is also important that the existing treatment be adequate and appropriate for the condition. This will require the citing of appropriate literature such as clinical trials of that therapy (Hammerschlag and Morris, 1997) to verify that the existing treatment is a standard treatment, and a clear statement about the qualifications of the personnel administering that treatment. Details of the treatment, frequency and duration should also be described.

25. Standard care (e.g., medication) compliance monitored

For a fair comparison of acupuncture to a standard therapy, it will be important to monitor treatment compliance. Poor compliance of the standard therapy is not usually considered to be adequate standard therapy, exceptions being for example in addiction, where utilization of and hence retention in treatment can be an outcome measure (see discussion in item 39). For a therapy that is administered by a therapist such as physical therapy, it is easy to monitor compliance. For a therapy such as medication, monitoring compliance may be more complex and appropriate strategies should be developed. In situations where poor compliance is encountered, strategies should be developed to help deal with this.

26. Existing treatment modality in reference group

This criterion from ter Riet and colleagues (1990a) is not strictly speaking accurate. It depends on the question asked in the study as to whether it is necessary or desirable that an existing therapy be used in the reference group (Resch and Ernst, 1996). The administration of an existing treatment modality in the reference group was suggested as a means of satisfying certain ethical considerations. But depending on the nature of the question, these sometimes have to be put aside. If the study design requires an untreated reference group, this criterion is inappropriate (Ernst and Resch, 1995; Resch and Ernst, 1996), see next criterion. If a reference group receives some form of control acupuncture, it may also be inappropriate. It can also depend on the severity and morbidity of the condition and patient population being studied as to whether it is necessary to maintain the existing treatment modality (Zaslawski, 2003). For example, in the treatment of asthma and in stroke rehabilitation, it is generally accepted that acupuncture is an adjunctive therapy, not a replacement for the standard therapy (NIH Consensus Conference, Acupuncture, 1998, Birch, 2001; Jobst, 1996; Naeser, 1996). Thus it is essential in such conditions that if acupuncture is used it be used *in addition* to the standard therapy. Such requirements allow for the use of a design where the standard therapy is compared to the use of the standard therapy with the addition of the acupuncture treatment (Johansson et al., 1993), which is probably the optimal design for the use of acupuncture for these conditions. This criterion is important in all studies of acupuncture, but it depends on many factors as to how important it is in a particular study.

27. Use of an untreated reference group

In a study that compares acupuncture to no treatment (no treatment/wait list group) it is of course necessary to include an untreated comparison group in the study. If a placebo-controlled study specifically attempts to quantify the size of the placebo effect then the study must also control for a number of nonplacebo related nonspecific effects such as the natural course of the disease, regression towards the mean, etc. To do this, it is recommended that an untreated control arm be added, wherever ethically possible (Davis, 2002; Ernst and Resch, 1995; Resch and Ernst, 1996).

28. Generalizability of the results of the study

The issue of the generalizability of the results in acupuncture studies is complex (Birch, 1997a; Sherman et al., 2001b, 2001c). In addition to selecting a truly adequate acupunc-

ture treatment and a truly appropriate control acupuncture treatment, it is additionally important to ensure, if possible, that the results are generalizable to the practice of acupuncture. If one is testing a specific and idiosyncratic treatment, it can be difficult to generalize the results to the practice of acupuncture. Within a specific model of practice the same issues can emerge, as there is often considerable variability in the published literature within a particular model of practice (Birch and Sherman, 1999). However, use of the BRITS method or some modification of it to select the test and control treatments, especially in conjunction with practitioner surveys and expert panels, to further reinforce the judgments, seems to be a reasonable strategy for addressing this issue (Birch and Sherman, 1999; Stux and Birch, 2001). Additionally, more limited applications of acupuncture should be explicitly acknowledged. A study investigating auricular acupuncture should be described as auricular acupuncture not acupuncture, a study using electroacupuncture should describe itself as electroacupuncture and not acupuncture.

29. DNIC control

DNIC mechanisms are heterosegmental pain-control mechanisms that occur regardless of the location of treatment stimulation (Le Bars et al., 1988). ter Riet and colleagues (1990a, 1990b) used "avoidance of DNIC" as a criterion for assessing clinical trials of acupuncture. However, simply *avoiding* DNIC is unrealistic and places an unfair burden on the study, possibly excluding common methods of treatment from study (Birch, 1995a), because many forms of acupuncture are naturally more painful than others (Birch, 1998; Birch and Felt, 1999) and are thus more likely to activate DNIC mechanisms. To avoid these is to not be able to test them. Rather one should "control" for the DNIC effects (Avants et al., 2000; Birch 1995a, 2003f; Birch and Jamison, 1998; Margolin et al., 1998a). Furthermore, controlling for only DNIC may not be sufficient to control for the range of possible nonplacebo, nonspecific physiologic effects that may occur (Birch, 2003f; Birch et al., 2002). It may therefore be important to explore the general effects of different needle techniques in preclinical studies in order to understand what other physiologic changes might result from needle insertion (Birch and Felt, 1999; Itaya et al., 1987; Pomeranz, 1996, 1998). It is proposed that DNIC be controlled for by attempting to match intensity of stimulation in the "active" and control "sham" treatments, coupled with assessment of the perceived intensities of those stimuli. This model has been used already in a number of controlled clinical trials (Avants et al., 2000; Birch and Jamison, 1998; Margolin et al., 2002). It should however be noted that in a "noninvasive sham" control study it may be difficult to adequately control for any DNIC effects because they will probably not be comparable between the noninvasive needling control and the acupuncture treatment (Birch, 1995a, 2003f).* This difficulty may persist even when patients seem unable to distinguish between the invasive and noninvasive sham treatments (Park et al., 2002). Similarly the nonplacebo nonspecific effects (such as DNIC) of a stronger stimulus type acupuncture, needling with *de qi* and a weaker stimulus type invasive sham control minimal needling, may not be comparable (Birch, 1995a, 2003f).* It may also be necessary to use acupuncture-naive patients in sham-controlled studies because some evidence suggests that the expectations of sensations associated with the treatment can show significant differences between acupuncture-experienced and acupuncture-naïve patients (Zaslawski C, personal communication). Pilot studies (Margolin et al., 1995, 1997b, 1998a; Park et al., 2002) will be important in the process of developing appropriate control procedures and measures. It would be useful to develop an instrument similar to that used for assessing credibility of treatment (Avants et al., 2000; Birch and Jamison, 1998, see next item).

30. Controlling for the nonspecific effects of treatment through appropriate assessments

There are potential problems with attempting placebo controlled studies of acupuncture resulting in large part to trying to control for the nonspecific effects of treatment, which include placebo and other physiologic effects. As Walach (2001) pointed out, these studies are conducted to highlight the specific effects of a treatment, and are seen as negative when the specific effects are not significantly greater than the nonspecific effects. In acupuncture studies, the nonspecific effects may be quite large because they involve both placebo and physiologic nonspecific effects (Birch et al., 2002; Ernst and Resch, 1995; Resch and Ernst, 1996). Additionally the enhancement of the healing processes that placebo possibly utilizes appears to be a desired part of the treatment, not an effect to be controlled for (Birch et al., 2002; Walach, 2001). Thus it is possible that in many instances the specific effects of the treatment may be quite small compared to the nonspecific effects, but that cumulatively the sum of these effects may be quite large; this is not the same as showing that the treatment is ineffective (Walach, 2001). Great care needs to be taken whenever placebo controlled studies are attempted. However, given these potential problems, when such studies are attempted, the credibility of the two acupuncture treatments (acupuncture treatment and sham acupuncture) should be assessed at intervals during the treatment phase using a series of standardized questions (Birch and Jamison, 1998; Margolin et al., 1998a; Petrie and Hazelman, 1985; Vincent, 1990a; White, 2002). While there have been questions about the validity of the specific questions used in the instrument developed for acupuncture studies by Vincent and others (White, 2002; White et al., 1996; Zaslawski, 1997), there is little disagreement about the need for some form of credibility measure (Birch and Jamison, 1998; Margolin et al., 1998a; Vincent and Lewith, 1995; White, 2002; White et al., 1996;

Wood and Lewith, 1998; Zaslawski et al., 1997). In addition, it may be useful to investigate other factors such as prior beliefs about treatment, expectations before treatment, prior experiences with treatment. It is also important to assess how the therapists are perceived, are they viewed as clinically capable, is their behavior seen to be the same in the different treatment arms, are they liked more in one treatment arm than another, are there more interactions between patients and practitioners in one treatment arm compared to the other? Various assessment models and methods have been explored and developed (Avants et al., 2000; Birch, 2003f; Birch et al., 2002; Birch and Jamison, 1998; Camp et al., 1995; Choi and Tweed, 1996; Kreitler et al., 1987; Margolin et al., 1998a; Roth et al., 1997). At the same time, nonplacebo-related nonspecific physiologic effects should be controlled for through the use of appropriate assessments (Birch and Jamison, 1998; see item 29). When appropriate valid assessments are made, and the scores for these questions are similar or equal in the different treatment groups, if there is a difference in treatment effects between the groups, it is likely that these differences are not the result of the nonspecific effects. Using appropriate statistical analyses, one can examine whether or to what degree these nonspecific effects may have been responsible for those between group treatment effect differences (Birch and Jamison, 1998). But it is also important to be aware of the potential for misinterpretation of the study results (Walach, 2001). This is a complex area that requires further work and discussion.

31. Steps taken to validate the traditionally based system of acupuncture tested

A number of trials of acupuncture have been conducted claiming that they were testing classic or traditional acupuncture often without description of the actual interventions used and with no citation of any relevant texts (Coan et al., 1980) to substantiate the claim that the treatments they applied were related to a particular tradition (Birch, 1997c). Acupuncture is a diverse medical system with many styles and methods used (Birch, 1998; Birch and Felt, 1999; MacPherson and Kaptchuk, 1997), even within the same tradition of practice (Birch and Sherman, 1999). It is thus important that research teams state clearly what traditional style of acupuncture they are testing and provide clear documentation of that style (Birch, 1997c, 2003g; MacPherson et al., 2002a, 2002b).

32. Reliability of traditionally based system of acupuncture assessments described

The use of traditional diagnostic methods and categories in order to individualize treatment is an important aspect of most traditionally based systems of acupuncture (Birch and Felt, 1999; MacPherson et al., 2002a; Zhang et al., 2003). In clinical trials of acupuncture for specific health problems, these categories become secondary inclusion/exclusion criteria (Birch, 1997c; Lao et al., 2001; MacPherson et al., 2002a; Schnyer and Allen, 2002; Zhang et al., 2003).* In order to guarantee administration of the correct treatment that matches the pattern of diagnosis for each patient, the reliability of these diagnostic assessments must be demonstrated (Birch, 1997c; Hogeboom et al., 2001; Lao et al., 2001; Lewith and Vincent, 1996; MacPherson et al., 2002a; Schnyer and Allen, 2002; Sherman and Cherkin, 2003; Vincent, 1993; White, 2002; Zhang et al., 2003).* To date only a few studies exploring the interrater reliability of diagnosis have been conducted (Hogeboom et al., 2001; Zell et al., 2000).* The current situation requires that any clinical trial of a traditional form of acupuncture will need to include an interrater reliability study to demonstrate that the diagnostic assessments were reliable and that patients received the appropriate treatment for the traditional style of practice (Vincent, 1993; Zhang et al., 2003). Some strategies have been offered in how to develop reliability, including the development of a manualized approach for the treatment protocol and using this as a training tool for study practitioners (Schyner and Allen, 2001, 2002).

33. Follow-up of 3 months or more after treatment

Unless looking only at short-term effects, it is important that adequate long-term follow-up be used in the study. Follow-up should be conducted at least 3 months after the completion of treatment, preferably up to 1 year. It depends on many factors as to what are the best lengths of follow-up that should be used. Crossover study designs have been used on occasion where long-term effects of treatment are being evaluated (Birch, 2002b, 2003b; Blom et al., 1996; Vincent, 1990b), but this is a mistake, because the effects of each therapy cannot be separated, making it impossible to interpret long-term effects of treatment in crossover designs.

34. Endpoints described

In a study where pain is the target of treatment, either primary or secondary, appropriate standardized and validated pain measurement tools should be used. It would also be useful to examine the frequency and intensity, sensory and affective components, emotional and support components, and examining various functional components of the pain using validated pain assessment questionnaires (Jamison, 1991). It would also be desirable to use range-of-motion measures and muscle pressure sensitivity (Vincent, 1993) to provide other more objective measures to match the subjective pain reports. If the study is examining the effects of acupuncture on something other than pain, appropriate standardized and validated outcome measurement methods should be used for each kind of symptom examined. Various outcome measurement tools exist in relation to specific health problems or as general health measures (e.g., McDowell and Newell, 1987).

35. Endpoint statistics presented/reader given opportunity to do inferential statistics

The results and statistical analyses of the primary endpoints should be presented in such a way and in sufficient detail to allow the reader to make their own interpretation and do their own inferential statistics. The primary endpoints are the focus of the answer to the question for which the study was designed hence they should be clearly presented.

36. Onset treatment effects compared

It is helpful to know about the onset of treatment effects, especially in a study that compares acupuncture to standard therapy (Hammerschlag and Morris, 1997). If both therapies are overall of similar efficacy but one has a more rapid onset (Lee et al., 1992) the one with more rapid onset can be considered to have certain advantages, and possibly be more clinically useful.

37. Duration treatment effects compared

Similar to the previous criterion in studies that compare acupuncture to standard therapy it is useful to compare the duration of the treatment effects (Hammerschlag and Morris, 1997). If the symptom improvement is similar in both treatments, but the effects last longer in one, that therapy could possibly be clinically more useful.

38. Follow-up data presented

The clinical significance of a therapy is often determined by examination of its long-term effects. It is thus important to clearly present the follow-up data for the reader to evaluate.

39. Assessment of dropout patients and presentation of data

It is ideal but not always possible to continue assessing dropout patients after they discontinue treatment. One way of handling the possibility of a higher dropout it to use questionnaires during the treatment phase of the study to examine patient understanding, expectations and belief in the therapy, the relative discomfort of the therapy, and the perceived effectiveness of the therapy (Avants et al., 2000; Birch and Jamison, 1998). If the acupuncture treatment and the control treatment groups showed the same or similar ratings of credibility, expectations and amounts of discomfort associated with them, and there were a higher drop out or loss to follow-up in one of the treatment groups but the patient reports of improvement in that group were less than in the other group, we can suspect that this higher dropout or loss to follow-up might be associated with the treatment being less effective in that group. This differential dropout or loss to follow-up rate could be used as an outcome measure, if the treatments are perceived as equally credible and painful but not equally effective. In substance abuse clinical trials, retention in treatment is itself often an outcome measure (Bullock et al., 1989), which makes criterion 17 less important in these studies, so long as appropriate assessments are made to be able to allow interpretation of differential drop out rates. One can also ask patients that discontinue in the study why they withdraw. Data for patients who withdraw or dropout of the study should be presented.

40. Duplicate assessments used

Where possible, similar measurements or assessments of the primary endpoints should be made. If more than one measurement or assessment shows the same result, this strengthens conclusions drawn from that result.

41. Use of medication

Use of medication should be monitored continuously or at regular intervals during the study using a standardized medication record. But, the tracking and calculation of medications consumed can be very complex, as different medications are not easily compared. A comparative quantitative method can sometimes be needed (Steedman et al., 1992).

42. Tracking other treatments

It is also important to track any other treatments used by the patient outside the study. If one of the study treatments is more effective than other study treatments, it is possible that patients receiving that treatment might stop using non-study treatments or use them less. Conversely, patients who receive a less effective study treatment might seek other non-study treatments more frequently. In a study comparing acupuncture to control acupuncture and a no-acupuncture group, several patients in the no-acupuncture group sought and received acupuncture outside of the study (Birch and Jamison, 1998), it is better to have tracked this. Not tracking the utilization of other therapies could confound the analysis of outcomes in studies of acupuncture. It is thus highly desirable to track utilization of other treatments during the treatment and follow-up phases of the study. Tracking the use of other therapies will be an important part of any assessment of cost effectiveness resource utilization, et cetera. These are most easily assessed in studies that compare acupuncture to standard therapy (Hammerschlag, 1998), thus tracking utilization of other therapies will be important in any cost effectiveness studies.

43. Activities of daily living

It would be useful to assess activities of daily living using a standardized instrument designed specifically for such measures. However, care should be taken to ensure that an

appropriate instrument is used. If the target symptom in the study is not debilitating, an instrument that measures moderate to severe debility may not be sensitive for the patients in the study. It is important to select the appropriate measurement tool for each study. This issue is part of the larger issue in clinical trials of acupuncture that of assessing broad health and quality of life changes. In principle, traditionally based systems of acupuncture do not simply treat the symptoms of a patient, but seek to improve the overall health of the patient as well (Birch, 1997c; Birch and Felt, 1999). Assessing whether and how well this objective is achieved would be good for traditionally based systems of acupuncture, but a validated outcome measure still needs to be developed that is sensitive to the expected changes.

44. Remark on side-effects

It is important in any clinical trial that any side effects of treatment be noted. It is often in the clinical trial phase development of a new therapy that side-effects are discovered and side-effect profiles developed. The causal relationship between an observed side-effect and the therapy is most easily clarified in controlled studies, as opposed to clinical practice (Margolin et al., 1997a). Acupuncture studies should use the same model. However, it will be important to discriminate at least between minor and severe side-effects and between side-effects of the therapy and improperly administered therapy (MacPherson, 1999b; Yamashita et al., 1999). This task is somewhat hampered by the lack of clear standard definitions and a classification scheme in the international literature on side-effects or adverse events (MacPherson, 1999a), however, it is important to attempt to differentiate between possible causes of each event.

45. Funding sources acknowledged

It is always proper to disclose all funding sources for a study, both private and public. This helps reveal possible sources of bias or other potential problems that can affect interpretation of the results.

DISCUSSION

In the preceding, a number of important issues and criteria thought necessary for conducting and presenting clinical trials of acupuncture have been discussed. It is hoped that these discussions will assist future researchers in the design, conduct, and publication of better-quality clinical trials of acupuncture.

These discussions have detailed 45 criteria that are important to consider when designing controlled studies of acupuncture. Table 1 summarizes which of the 45 criteria are important for the three types of study design, acupuncture versus sham acupuncture, acupuncture versus standard therapy, and acupuncture versus no treatment or wait list. Those criteria where the use of pilot studies will be important to determine how best to address them have been highlighted.

Most acupuncture trials were conducted without having first attempted pilot studies to better define solutions to the complex problems involved in acupuncture research. It should be routine to have pilot data available before larger controlled studies of acupuncture are conducted (Linde and Jonas, 1999; Margolin, 1999). The reasons for this are not simply to do with the need to identify critical factors that may not have emerged in the phase of formulation and design of the trial, but also for ethical reasons. It is possibly unethical to enter human subjects into a trial without having adequate knowledge of the possible outcomes and problems that arise in implementing the study. The only way to prepare oneself before conducting full-scale controlled clinical trials is to conduct pilot studies or be able to point to other studies from which one can reasonably estimate needs and requirements for the study. In the above discussions a number of critical areas were identified where pilot studies will likely be necessary to be able to develop sufficient answers for those criteria. But pilot studies will probably be necessary in general to test an overall study design, not only to address specific criteria for the study.

There is a continued demand for placebo-controlled studies of acupuncture (NIH Consensus Conference, Acupuncture, 1998, British Medical Association, 2000). In certain areas, such as when acupuncture is used as an adjunctive therapy, for example addiction, asthma, angina and stroke, such demands may no longer be considered reasonable, because the difficulties of interpreting results in light of probable interaction of the effects of the different treatments (Birch et al., 2002; Kleijnen et al., 1994; Kleijnen and de Craen, 1996). Placebo studies can sometimes be difficult because of possible ethical problems (Ezzo et al., 2001a; Linde et al., 1996) or the probable interaction of treatment effects of the different specific and nonspecific effects of the treatment (Birch, 2003f; Birch et al., 2002; Hyland, 2003; Kaptchuk et al., 1996). Placebo-controlled studies may also run contrary to the specific treatment goals of some acupuncture treatments (Birch et al., 2002), and may lead to paradoxical conclusions (Walach, 2001). However, decisions about the need for placebo-controlled studies are not made by single research teams, but by the research community as a whole, and often researchers must satisfy the demands of ethics committees, grant reviewers, and peer review. As long as placebo-controlled studies are demanded, it will be necessary to continue designing studies that attempt placebo controls. Any time that the specific effects of an acupuncture treatment are to be assessed, some form of control of the nonspecific effects will be necessary (Birch et al., 2002). Greater care needs to be given to consideration of when placebo-controlled studies should and can be conducted (Birch, 2003b; Birch et al., 2002). Greater care needs to be

given to definitions and discussions of the different treatment effects of acupuncture, for example no longer calling the sham treatment a placebo treatment (Birch et al., 2002). Additionally, much greater care needs to be given in the selection of the control treatment (Birch 2003b, 2003f). Table 1 details the criteria thought necessary for conducting a placebo-controlled study, the acupuncture treatment versus sham acupuncture model.

To date few studies have tested or attempted to test traditional forms of acupuncture. Most of those that did, performed poor studies or used inadequate study design (Birch, 1997c). If a particular study attempts to test the efficacy of a traditional form of acupuncture, one that individualizes treatments based on individual assessments of patient differences guided by traditional models and principles of practice, it will be necessary to pay attention to several of the criteria listed above, in particular 31 and 32, validating the model of practice tested and testing the reliability of diagnosis (Birch, 1997c, 2003g; MacPherson et al., 2002a; White, 2002). If the study attempts to test the theoretical basis of the tradition, that for example, the diagnosis yields a treatment that has specific treatment effects (specific sites and techniques [Birch, 1995a, 1997c]) that maximize the overall treatment effects then the study will need to use the acupuncture treatment versus sham acupuncture model. In order to investigate the specific effects, one must control for the nonspecific effects (Birch, 2003f; Birch et al., 2002). Thus the criteria listed in Table 1 necessary for the acupuncture treatment versus sham acupuncture design will be necessary for studies testing the specific effects of traditional forms of acupuncture, with special attention paid to controlling for the broad spectrum of nonspecific effects associated with the test and sham treatments (Birch, 1997c; Birch and Jamison, 1998). But when a study does not attempt to investigate the specific effects of a traditionally based treatment, comparison to standard therapy (pragmatic trial) is often a good model to understand the value of treatment (MacPherson et al., 2002a).

These 45 criteria could also become the basis for evaluating the quality of clinical trials of acupuncture. ter Riet and colleagues (1990a, 1990b) used a list of 18 criteria, Hammerschlag and Morris (1997) used a list of 25 criteria, the Jadad scale uses 5 criteria (Jadad 1998). It may be possible to develop a new list of criteria incorporating some of the above.

CONCLUSION

We have seen that there are many complex issues that need to be addressed if high-quality trials of acupuncture are to yield interpretable results. This paper has extended earlier lists of study design components and evaluation criteria and presented a list of 45 criteria with discussions of each. It is hoped that the discussions here will be helpful for those designing, conducting, and writing up controlled clinical trials of acupuncture. Suggestions were made for which criteria are likely to be important in studies that compare acupuncture to a sham, a standard therapy, or a no-treatment or wait-listed control, with further suggestions of which will benefit from pilot studies. Many of the criteria discussed here will also be important in clinical trials of other nonpharmacologic therapies, and thus could form the starting point for the development of similar lists for other therapies.

ACKNOWLEDGMENTS

Thanks to Chris Zaslawski, Ph.D., candidate, for reviewing the manuscript and for his useful comments and suggestions.

REFERENCES

Aickin M. Beyond randomization. J Altern Complement Med 2003;8:765–772.

Aird M, Cobbin DM, Rogers C. A study of the relative precision of acupoint location methods. J Altern Complement Med 2002;8:635–642.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T, for the CONSORT Group. The Revised CONSORT statement for reporting randomized controlled trials: Explanation and elaboration. Ann Intern Med 2001;134 (8):663–694.

AUA, (American Urological Association). J Urol 1998;169: 2279–2282.

Avants SK, Margolin A, Chang P, Kosten TR, Birch S. Acupuncture for the treatment of cocaine addiction in methadone-maintained patients: Investigation of a needle puncture control. J Subst Abuse Treat 1995;12:95–105.

Avants SK, Margolin A, Holford TR, Kosten TR. A randomized controlled trial of auricular acupuncture for cocaine dependence. Arch Intern Med 2000;160:2305–2312.

Bailar JC, Mosteller F. Medical Uses of Statistics. Boston, NEJM Books, 1992.

Begg CB, Cho MK, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. JAMA 1996;276:637–679.

Birch S. Prelude to a trial: Preparatory studies for clinical trial research of acupuncture and East Asian herbal medicine. Proceedings of the First Symposium of the Society for Acupuncture Research. Bethesda, Society for Acupuncture Research, 1994:38–50.

Birch S. Testing the clinical specificity of needle sites in controlled clinical trials of acupuncture. Proceedings of the Second Society for Acupuncture Research. Georgetown: Society for Acupuncture Research, 1995a:274–294.

Birch S. Introduction. In: Manaka Y, Itaya K, Birch S. Chasing the Dragon's Tail. Brookline, MA: Paradigm Publications, 1995b: ix–xxviii.

Birch S. Issues to consider in determining an adequate treatment

in a clinical trial of acupuncture. Complement Ther Med 1997a;5:8–12.
Birch S. Credibility of treatment in controlled trials of acupuncture [letter to the editor]. J Altern Complement Med 1997b;3:315–317.
Birch S. Testing the claims of traditionally based acupuncture. Complement Ther Med 1997c;5:147–151.
Birch S. Diversity and acupuncture: Acupuncture is not a coherent or historically stable tradition. in AJ. Vickers (ed.) Examining Complementary Medicine: The Sceptical Holist. Cheltenham: Stanley Thomas, 1998:45–63.
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. Berlin: Springer Verlag, 2001:131–149.
Birch S. Letter to the editor. BMJ 2000. Online document at: intl.bmj.org/cgi/eletters/321/7252/11/b#EL7
Birch S. Systematic reviews of acupuncture—Are there problems with these? Clin Acupunct Orien Med 2001a;2:17–22.
Birch S. Letter to the editor. Clin Acupunct Orien Med 2001b;2: 113–115.
Birch S. Treatment of pain by acupuncture. In: Cassidy CM, ed. Contemporary Chinese Medicine and Acupuncture. Edinburgh: Churchill Livingstone, 2002a:239–253.
Birch S. Letter to the editor. Acubriefs, 2002b;3:9. Online document at: www.acubriefs.com/newsletters/newsletter_09.02.htm
Birch S. Acupuncture and bronchial asthma, a long-term randomized study [letter to the editor]. J Altern Complement Med 2003a;8:751–754.
Birch S. Overview of models used in controlled acupuncture studies and thoughts about questions answerable by each. Clin Acupunct Orient Med 2003b;3:207–217.
Birch S. Trigger point—Acupuncture point correlations revisited. J Altern Complement Med 2003c;9:91–103.
Birch S. Developing a research strategy for the acupuncture profession: research questions, resources necessary to answer them and guidelines for matching resources to types of research. Clin Acupunct Orient Med 2003d;4:29–33.
Birch S. Letter to the editor. Acubriefs 2003e;4:7. Online document at: www.acubriefs.com/newsletters/newsletter_07.03.htm
Birch S. Controlling for non-specific effects of acupuncture in clinical trials. Clin Acupunct Orient Med 2003f;4:59–70.
Birch S. Testing traditionally based systems of acupuncture. Clin Acupunct Orient Med 2003g;4:84–87.
Birch S. Problems with systematic reviews of acupuncture: What should we do about these? Clin Acup Orien Med 2003h; 4:105–108.
Birch S, Felt RO. Understanding Acupuncture. Edinburgh: Churchill Livingstone, 1999.
Birch S, Hammerschlag R, Berman B. Acupuncture in the treatment of pain. J Altern Complement Med 1996;2:101–124.
Birch S, Hammerschlag R, Trinh K, Zaslawski C. The non-specific effects of acupuncture treatment: When and how to control for them. Clin Acupunct Orient Med 2002;3:20–25.
Birch S, Jamison RN. A controlled trial of Japanese acupuncture for chronic myofascial neck pain: Assessment of specific and nonspecific effects of treatment. Clin J Pain 1998;14:248–255.
Birch S, Kaptchuk T. The history, nature and current practice of acupuncture: An East Asian perspective. In: Ernst E, White A, eds. Acupuncture a Scientific Appraisal. Oxford: Butterworth-Heinemann, 1999:11–30.
Birch S, Keppel Hesselink J, Jonkman FAM, Hekker TAM, Bos A. Clinical research on acupuncture: Part 1. What have reviews of the efficacy and safety of acupuncture told us so far? J Altern Complement Med 2004;10:468–480.
Birch S, Sherman K. Zhong Yi Acupuncture and low back pain: Traditional Chinese medical acupuncture differential diagnoses and treatments for chronic lumbar pain. J Altern Complement Med 1999;5:415–425.
Birch S, Tsutani K. Survey of English-language publications on acupuncture. Complement Ther Med 1996;4:172–177.
Blom M, Dawidson I, Fernberg J-O, Johnson G, Angmar-Mansson B. Acupuncture treatment of patients with radiation-induced xerostomia. Oral Oncol Eur J Cancer 1996;328:182–190.
British Medical Association. Acupuncture: Efficacy, Safety and Practice. London: Harwood Academic Publishers, 2000.
Bullock ML, Culliton PD, Olander RT. Controlled trial of acupuncture for severe recidivist alcoholism. Lancet 1989;1:1435–1439.
Camp AV, Philips DI, Collier SM, Kirk AP. The influence of knowledge, attitudes and personality characteristics on the outcome of acupuncture treatment [abstr]. Br J Rheum 1995:51.
Chalmers I, Altman D. Systematic Reviews. London: BMJ Publishing Group, 1995.
Choi PYL, Tweed A. The holistic approach in acupuncture treatment: Implications for clinical trials. J Psychosom Res 1996;41: 349–356.
Coan R, Wong G, Ku SL, Chan YC, Wang L, Ozer FT, Coan PL. The acupuncture treatment of low back pain: A randomized controlled treatment. Am J Chin Med 1980:VIII:181–189.
Davis CE. Regression to the mean or placebo effect? The science of the placebo. Towards an interdisciplinary research agenda. In Guess HA, Kleinman A, Kusek JW, Engel LW, eds. London: British Medical Journal Books, 2002:158–166.
Elloriaga Claraco A, Fargas-Babjak A, Hanna SE. The reporting of clinical acupuncture research: what do clinicians need to know? J Altern Complement Med 2003a;9:143–149.
Elloriaga Claraco A, Hanna SE, Fargas-Babjak A. Reporting of clinical details in randomized controlled trials of acupuncture for the treatment of migraine/headaches and nausea/vomiting. J Altern Complement Med 2003b;9:151–159.
Ernst E. Acupuncture research: Where are the problems? Acupunct Med 1994;12:93–97.
Ernst E, Resch KL. Concept of true and perceived placebo effects. BMJ 1995;311:551–553.
Ernst E, White AR. Acupuncture as an adjuvant therapy in stroke rehabilitation? Wien Med Wschr 1996;146:556–558.
Ernst E, White AR. A review of problems in clinical acupuncture research. Am J Chin Med 1997;XXV:3–11.
Ernst E, White AR. Acupuncture for back pain: A meta-analysis of randomized controlled trials. Arch Intern Med 1998;158: 2235–2241.
Ernst E, White AR. Acupuncture as a treatment for temporomandibular joint dysfunction. A systematic review or randomized trials. Arch Otolaryngol Head Neck Surg 1999;125:269–272.
Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain: A systematic review. Pain 2000;86:217–225.
Ezzo J, Lao LX, 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. Berlin: Springer Verlag, 2001a:113–130.

Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, Berman B. Acupuncture for osteoarthritis of the knee: A systematic review. Arthritis Rheum 2001b;44:819–825.

Falk CX, Birch S, Avants SK, Tsau Y, Margolin A. Preliminary results of a new method for locating auricular acupuncture points. Acupunct Electrother Int Res J 2000;25:165–177.

Fink M, Gutenbrunner C, Rollink J, Karst M. Credibility of a newly designed placebo needle for clinical trials in acupuncture research. Forsch Komplementarmed Klass Naturheilkd 2001; 8:368–372.

Hammerschlag R. Methodological and ethical issues in clinical trials of acupuncture. J Altern Complement Med 1998;4:159–171.

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;5:133–140.

Hogeboom CJ, Sherman KJ, Cherkin DC. Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists. Complement Ther Med 2001;9: 154–166.

Hopwood V, Lewith G. Acupuncture trials and methodological considerations. Clin Acupunct Orient Med 2003;3:192–199.

Hyland ME. Extended network generalized entanglement theory: Therapeutic mechanisms, empirical predictions and investigations. J Altern Complement Med 2003;9:919–936.

International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991;324:424–428.

Itaya K, Manaka Y, Ohkubo C, Asano M. Effects of acupuncture needle application upon cutaneous microcirculation of rabbit ear lobe. Acupunct Electrother Res Int J 1987;12:45–51.

Jadad AR. Randomized Controlled Trials: A User's Guide. London: BMJ Publishing Group, 1998.

Jamison RN. Psychological assessment of chronic pain patients. Pain Digest 1991;1:203–211.

Jobst KA. Acupuncture in asthma and pulmonary disease: An analysis of efficacy and safety. J Altern Complement Med 1996;2:179–206.

Johansson K, Lindgren I, Widner H, Wiklung I, Johansson BB. Can sensory stimulation improve the functional outcome in stroke patients? Neurology 1993;43:2189–2192.

Kaptchuk TJ, Edwards RA, Eisenberg DM. Complementary medicine: Efficacy beyond the placebo effect. In: Ernst E, ed. Complementary Medicine: An Objective Appraisal. Oxford: Butterworth Heinemann, 1996:42–70.

Karst M, Scheinichen D, Rueckert T, Wagner T, Wiese B, Piepenbrock S, Fink M. Effect of acupuncture on the neutrophil respiratory burst: A placebo-controlled single-blind study. Complement Ther Med 2003;11:4–10.

Kleijnen J, de Craen AJM. The importance of the placebo effect: A proposal for further research. In: Ernst E, ed. Complementary Medicine: An Objective Appraisal. Oxford: Butterworth Heinemann, 1996:31–41.

Kleijnen J, de Craen AJM, van Everdingen J, Krol L. Placebo effect in double blind clinical trials: A review of interactions with medications. Lancet 1994;ii:1347–1349.

Kleijnen J, ter Riet G, Knipschild P. Acupuncture and asthma; a review of controlled trials. Thorax 1991;46:799–802.

Kreitler S, Kreitler H, Carasso R. Cognitive orientation as predictor of pain relief following acupuncture. Pain 1987;28:323–341.

Lao L, Ezzo J, Berman BM, Hammerschlag R. Assessing clinical efficacy of acupuncture: Considerations for designing future acupuncture trials. In Stux G, Hammerschlag R, eds. Clinical Acupuncture: Scientific Basis. Berlin: Springer Verlag, 2001: 187–209.

Le Bars D, Willer JC, de Broucker T, Villanueva L. Neurophysiological mechanisms involved in the pain-relieving effects of counter-irritation and related techniques including acupuncture. In Pomeranz B, Stux G, eds. Scientific Bases of Acupuncture. Berlin: Springer Verlag, 1988:79–112.

Lee YH, Lee WC, Chen MT, Huang JK, Chung C, Chang LS. Acupuncture in the treatment of renal colic. J Urol 1992; 147:16–18.

Lewith GT, Machin D. On the evaluation of the clinical effects of acupuncture. Pain 1983;16:111–127.

Lewith G, Vincent C. On the evaluation of the clinical effects of acupuncture: a problem reassessed and a framework for future research. J Altern Complement Med 1996;2:79–90.

Linde K, Jonas WB. Evaluating complementary and alternative medicine: The balance of rigor and relevance. In: Jonas WB, Levin JS, eds. Essentials of Complementary and Alternative Medicine. Philadelphia: Lippincott, Williams and Wilkins, 1999:57–71.

Linde K, Worku F, Stor W, Wiesner-Zechmeister M, Pothmann R, Weinschutz T, Melchart D. Randomized clinical trials of acupuncture for asthma—A systematic review. Forsch Kompl 1996;3:148–155.

Lytle CD. An overview of acupuncture; U.S. Department of Health and Human Services, Public Health Service, Food and Drug Administration, Center for Devices and Radiological Health, 1993.

Maciocia G. Letter to the editor. Complement Ther Med 1993; 1:221–222.

MacPherson H. Fatal and adverse events from acupuncture: Allegation, evidence and the implications. J Altern Complement Med 1999a;5:47–56.

MacPherson H. How safe is acupuncture? Developing the evidence on risk. J Altern Complement Med 1999b;5:223–224.

Macpherson H, Kaptchuk TJ. Acupuncture in Practice. New York: Churchill Livingstone, 1997.

MacPherson H, Sherman K, Hammerschlag R, Birch S, Lao L, Zaslawski C. The clinical evaluation of East Asian systems of medicine. Clin Acupunct Orient Med 2002a;3:16–19.

MacPherson H, White A, Cummings M, Jobst K, Rose K, Niemtzow R. Standards for reporting interventions in controlled trials of acupuncture: The STRICTA recommendations. J Altern Complement Med 2002b;8:85–89.

Margolin A. Liabilities involved in conducting randomized clinical trials of CAM therapies in the absence of preliminary, foundational studies: a case in point. J Altern Complement Med 1999;5:103–104.

Margolin A, Avants SK, Birch S. Letter to the editor. Complement Ther Med 1997a;5:53–54.

Margolin A, Avants SK, Chang P, Birch S, Kosten TR. A single-blind investigation of four auricular needle puncture configurations. Am J Chin Med 1995;2:105–114.

Margolin A, Avants SK, Birch S, Falk CX, Kleber HD. Methodological investigations for a multisite trial of auricular acupuncture for cocaine addiction. A study of active and control auricular zones. J Subst Abuse Treat 1997b;13:471–481.

Margolin A, Avants SK, Kleber HD. Rationale and design of the

cocaine alternative treatments study (CATS): A randomized, controlled trial of acupuncture. J Altern Complement Med 1998a;4:405–418.

Margolin A, Avants SK, Kleber HD. Investigating alternative medicine therapies in randomized controlled trials. JAMA 1998b;280:1626–1628.

Margolin A, Kleber HD, Avants SK, Konefal J, Gawin F, Stark E, Sorenson J, Midkiff E, Wells E, Jackson TR, Bullock M, Culliton PD, Boles S, Vaughan R. Acupuncture for the treatment of cocaine addiction. A randomized controlled trial. JAMA 2002; 287:55–63.

Mayer DJ. Acupuncture: An evidence-based review of the clinical trial literature. Annu Rev Med 2000;51:49–63.

McDowell I, Newell C. Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford: Oxford University Press, 1987.

McLellan AT, Grossman DS, Blain JD, Haverkos HW. Acupuncture treatment for drug abuse: A technical review. J Subst Abuse Treat 1993;10:569–576.

Meinert CL. Clinical Trials: Design, Conduct and Analysis. Oxford: Oxford University Press, 1986.

Melchart D, Linde K, Fischer P, White A, Allais G, Vickers A, Berman B. Acupuncture for recurrent headaches: A systematic review of randomized controlled trials. Cephalalgia 1999;19: 779–786.

Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. Ann Intern Med 2002;136: 471–476.

Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191–2001.

Molsberger A, Bowing G, Haake M, Meier U, Winkler J, Molsberger F. Acupuncture in diseases of the locomotor system. Status of research and clinical applications. Orthopoade 2002;31: 536–543.

Morris MM. Overview of acupuncture in chronic pain clinical research. J Altern Complement Med 1996;2:125–127.

Naeser MA. Acupuncture in the treatment of paralysis due to central nervous system damage. J Altern Complement Med 1996;2:211–248.

NIH Consensus Conference. Acupuncture. JAMA 1998;280: 1518–1524.

Park JB, White A, Lee H, Yamashita H, Ernst E. Credibility of a newly developed sham needle. J Jpn Soc Acup Moxibust 2000;50:111–114.

Park JB, White A, Stevinson C, Ernst E, James M. Validating a new non-penetrating sham acupuncture device: Two randomised controlled trials. Acupunct Med 2002;20:168–174.

Patel M, Gutzwiller F, Paccaud F, Marazzi A. A meta-analysis of acupuncture for chronic pain. Int J Epidemiol 1989;18:900–906.

Petrie J, Hazelman B. Credibility of placebo transcutaneous nerve stimulation and acupuncture. Clin Exp Rheum 1985;3:151–153.

Pocock SJ. The pros and cons of non-inferiority (equivalence) trials. The science of the placebo. Towards an interdisciplinary research agenda. In Guess HA, Kleinman A, Kusek JW, Engel LW, eds. London: British Medical Journal Books, 2002: 236–248.

Pomeranz B. Scientific research into acupuncture for the relief of pain. J Altern Complement Med 1996;2:53–60.

Pomeranz B. Scientific basis of acupuncture. In: Stux G, Pomeranz B. Basics of Acupuncture. Berlin, Springer-Verlag, 4th ed., 1998:6–72.

Resch KL, Ernst E. Research methodologies in complementary medicine: Making sure it works. In Ernst E, ed. Complementary Medicine. An Objective Appraisal. Oxford: Butterworth Heinemann, 1996:18–30.

Richardson PH, Vincent CA. Acupuncture for the treatment of pain: A review of evaluative research. Pain 1986;24:15–40.

Romoli M, Allais G, Giovanardi CM, Menniti Ippolito F. What are the most important factors in the evaluation of an acupuncture treatment: Developing a possible scale to be used for systematic reviews on acupuncture. Clin Acup Orient Med 2003;4:109–113.

Roth LU, Maret-Maric A, Adler RH, Neuenschwander BE. Acupuncture points have subjective (needling sensation) and objective (serum cortisol increase) specificity. Acupunct Med 1997;15:2–5.

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. Boston: Little, Brown and Company, 2nd ed., 1991.

Schnyer RN, Allen JJB. Acupuncture in the Treatment of Depression. A Manual for Practice and Research. Edinburgh: Churchill Livingstone, 2001.

Schneyer RN, Allen JJB. Bridging the Gap in Complementary and Alternative Medicine Research: Manualization as a Means of Promoting Standardization and Flexibility of Treatment in Clinical Trials of Acupuncture. J Altern Complement Med 2002;8:623–634.

Sherman KJ, Cherkin DC. Challenges of acupuncture research: Study design considerations. Clin Acupunct Orient Med 2003;3:200–206.

Sherman KJ, Cherkin DC, Hogeboom CJ. Developing protocols for acupuncture studies: An efficacy trial of low back pain. Altern Ther 2001a;7:111.

Sherman KJ, Cherkin DC, Hogeboom CJ. The diagnosis and treatment of patients with chronic low back pain by traditional Chinese medical acupuncturists. J Altern Complement Med 2001b;7:641–650.

Sherman KJ, Hogeboom CJ, Cherkin DC. How traditional Chinese medicine acupuncturists would diagnose and treat chronic low back pain: Results of a survey of licensed acupuncturists in Washington State. Complement Ther Med 2001c;9:146–153.

Sherman KJ, Hogeboom CJ, Cherkin DC, Deyo RA. Description and validation of a noninvasive placebo acupuncture procedure. J Altern Complement Med 2002;8:11–19.

Steedman SM, Middaugh SJ, Kee WG, Carson DS, Harden RN, Miller MC. Chronic-pain medications: Equivalence levels and method of quantifying usage. Clin J Pain 1992;8:204–214.

Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. Lancet 1998;352:364–365.

Stux G, Birch S. Proposed standards of acupuncture treatments for clinical studies. In Stux G, Hammerschlag R, eds. Clinical Acupuncture: Scientific Basis. Berlin: Springer Verlag, 2001: 171–185.

Tait PL, Brooks L, Harstall C. Acupuncture: Evidence from Systematic Reviews and Meta-analyses. Alberta, Canada: Alberta Heritage Foundation for Medical Research, 2002.

ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: A criteria-based meta-analysis. J Clin Epidemiol 1990a;43: 1191–1199.

ter Riet G, Kleijnen J, Knipschild P. A meta-analysis of studies into the effect of acupuncture on addiction. Br J Gen Pract 1990b;40:379–382.

van Tulder MW, Cherkin DC, Berman B, Lao L, Koes BW. The effectiveness of acupuncture in the management of acute and chronic low back pain. A systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine 1999;24:1113–1123.

Vickers AJ. Inspiration and perspiration: what every researcher needs to know before they start. In Lewith G, Jonas WB, Walach H, eds. Clinical Research in Complementary Therapies. Edinburgh: Churchill Livingstone, 2002:47–58.

Vickers A, Wilson P, Kleijnen J. Effectiveness bulletin: Acupuncture. Qual Saf Health Care 2002;11:92–97.

Vincent CA. The methodology of controlled trials of acupuncture. Acupunct Med 1989;VI:9–13.

Vincent CA. Credibility assessment in trials of acupuncture. Complement Med Res 1990a;4:8–11.

Vincent CA. The treatment of tension headache by acupuncture: A controlled single case design with time series analysis. J Psychosomat Res 1990b;34:553–561.

Vincent CA. Acupuncture as a treatment for chronic pain. In Lewith GT, Aldridge D, eds. Clinical Research Methodology for Complementary Therapies; London: Hodder and Stoughton, 1993:289–308.

Vincent C, Lewith G. Placebo controls for acupuncture studies. J R Soc Med 1995;88:199–202.

Vincent CA, Richardson PH. Acupuncture for some common disorders: A review of evaluative research. J R Coll Gen Pract 1987;37:77–81.

Walach H. The efficacy paradox in randomized controlled trials of CAM and elsewhere: Beware of the placebo trap. J Altern Complement Med 2001;7:213–218.

Walach H. Critical factors for credibility in clinical trials: comments on Aickin's "beyond randomization." J Altern Complement Med 2003;8:773–775.

White AR. Acupuncture research methodology. In Lewith G, Jonas WB, Walach H, eds. Clinical Research in Complementary Therapies. Edinburgh: Churchill Livingstone, 2002:307–323.

White AR, Eddleston C, Hardie R, Resch KL, Ernst E, A pilot study of acupuncture for tension headache, using a novel placebo. Acupunct Med 1996;14:11–15.

White AR, Ernst E. A trial method for assessing the adequacy of acupuncture treatments. Altern Ther Health Med 1998;4:66–71.

White AR, Ernst E. A systematic review of randomized controlled trials of acupuncture for neck pain. Rheumatology 1999; 38:143–147.

White AR, Filshie J, Cummings TM. Clinical trials of acupuncture: Consensus recommendations for optimal treatment, sham controls and blinding. Complement Ther Med 2001;9:237–245.

White A, Park JB. Protocols for clinical trials of acupuncture. Acupunct Med 1999;17:54.

White AR, Resch KL, Ernst E. A meta-analysis of acupuncture techniques for smoking cessation. Tobacco Control 1999;8: 393–397.

White P, Lewith G, Berman B, Birch S. Reviews of acupuncture for chronic neck pain: Pitfalls in conducting systematic reviews. Rheumatology 2002;41:1224–1231.

Wood R, Lewith G. The credibility of placebo controls in acupuncture studies. Complement Ther Med 1998;6:79–82.

Wyon Y, Lindgren R, Lundeberg T, Hammar M. Effects of acupuncture on climacteric vasomotor symptoms, quality of life, and urinary excretion of neuropeptides among postmenopausal women. Menopause 1995;2:3–12.

Yamashita H, Tsukayama H, Tanno Y, Nishijo K. Adverse events in acupuncture and moxibustion treatment: A six-year survey at a national clinic in Japan. J Altern Complement Med 1999;5: 229–236.

Zaslawski C. Response to letter to the editor. J Altern Complement Med 1997;3:317–321.

Zaslawski C. The impact of ethics on the development and implementation of acupuncture clinical trials. Clin Acupunct Orient Med 2003;4:109–113.

Zaslawski C, Rogers C, Garvey M, Ryan D, Yang CX, Zhang SP. Strategies to maintain the credibility of sham acupuncture used as a control treatment in clinical trials. J Altern Complement Med 1997;3:257–266.

Zell B, Hirata J, Marcus A, Ettinger B, Pressman A, Ettinger KM. Diagnosis of symptomatic postmenopausal women by traditional Chinese medicine practitioners. Menopause 2000;7:129–134.

Zhang GG, Bausell B, Lao LX, Handwerger B, Berman BM. Assessing the consistency of traditional Chinese medical diagnosis: An integrative approach. Altern Ther 2003;9:66–71.

Address reprint requests to:
Stephen Birch, Ph.D., Lic.Ac., M.B.Ac.C
Foundation for the Study of Traditional East Asian Medicine
W.G. Plein 330
1054 SG Amsterdam
The Netherlands

E-mail: 71524.3461@compuserve.com