

Review

Reviews of acupuncture for chronic neck pain: pitfalls in conducting systematic reviews

P. White¹, G. Lewith^{1,2}, B. Berman³ and S. Birch⁴

¹University of Southampton, Royal South Hants Hospital, Southampton, ²University of Southampton, Southampton, UK, ³University of Maryland, School of Medicine, Baltimore, USA and ⁴Stichting (Foundation) for the Study of Traditional East Asian Medicine, W.G. Plein 330, 1054 SG Amsterdam, The Netherlands

Abstract

This paper examines some of the problems specifically associated with conducting research into acupuncture and how this can lead to further problems with subsequent systematic reviews. Studies for the treatment of chronic neck pain have been used as examples of how presented information can be misleading to an acupuncture-naïve reader and how researchers must be sensitive to these problems when compiling their inclusion and exclusion criteria. The problems associated with scoring trials are discussed and further work to increase the scope of scoring mechanisms is recommended in order to produce meaningful systematic reviews in the future.

KEY WORDS: Acupuncture, Neck pain, Systematic review.

The efficacy of the treatment of chronic mechanical neck pain with acupuncture has yet to be convincingly proven. The prime reason for this is the lack of basic evidence, in the form of good-quality randomized controlled trials (RCTs). Adding to the confusion is a plethora of systematic reviews on acupuncture and pain that have included a number of clinical trials on the use of acupuncture to treat neck pain, some of which should have perhaps been omitted, either because they did not involve acupuncture or because they are of such poor quality that they tell us little that is constructive about its clinical effects. Systematic reviews can be a very valuable tool because they can give an indication of the quality of trials and also form a summary of the available literature on a particular subject. In this format, evidence is easily accessible and digestible to researchers and, perhaps more importantly, to clinicians and purchasers, who often simply require a reliable summary of evidence in order to make policy or informed clinical decisions. The aim of all clinical research must surely be to assess an intervention or process or the relationship between variables, ultimately in order to be

able to improve clinical practice and health-care for patients/consumers. As such, evidence gathered must have external validity (generalizability) [1] and it must be relevant to clinical practice. This article examines some of the fundamental pitfalls in which some of these systematic reviews have been caught, thus raising issues about their credibility, generalizability and subsequent conclusions. Our aim is to critically review the foundations of systematic reviews within the field of acupuncture with particular reference to neck pain, and to open up a debate that will allow us to formulate new and more appropriate methodology.

There have been many systematic reviews which have included trials relating to the efficacy of acupuncture for the treatment of neck pain along with other trials involving painful conditions [2–11]. The broad general conclusions of these are summarized in Table 1.

With many of these trials being examined over and over again by so many investigators, we must ask why there is a need to revisit this topic. Why must we scrutinize the results of these reviews with scepticism and why does the answer to the acupuncture and neck pain conundrum remain so elusive? The answer to these questions may lie, in part, with the quality and number of RCTs as well as the systematic review methodology and the subsequent selection and analysis of these trials.

Only one review has attempted to deal solely with the question of acupuncture and neck pain. This was carried out by White and Ernst [9]. As our particular interest

Submitted 22 October 2001; revised version accepted 16 April 2002.

Correspondence to: P. White, School of Medicine, Complementary Medicine Research Unit, Community Clinical Sciences, Mailpoint OPH, Royal South Hants Hospital, Brintons Terrace, Southampton SO14 0YG, UK.

TABLE 1. Systematic review of acupuncture and pain including neck pain

Prime author	Summary of results
Mendelson [2]	Some evidence that acupuncture has a marked analgesic effect
Richardson [3]	Generally, 50–80% of patients receive benefit from acupuncture
Patel [4]	Pooled results showed significance in favour of acupuncture
ter Riet [5]	The efficacy of acupuncture for chronic pain remains doubtful
Aker [6]	Lack of evidence for many conservative treatments of neck pain
Ernst [7]	In the main, acupuncture is not superior to sham needling
Kjellman [8]	Few trials on neck problems are of high methodological quality
White [9]	Acupuncture treatment for neck pain is not supported by evidence
Smith [10]	No convincing evidence for acupuncture for back or neck pain
Ezzo [11]	Inconclusive proof that acupuncture is more effective than placebo

was in neck pain, their review will be, by way of example, the focus of this article. However, we concentrate on the science involved in systematic reviews. It must be pointed out that both White and Ernst are well-respected and thoughtful investigators in the field of complementary medicine, being both prolific and thorough in their work. This paper therefore highlights the fact that systematic reviews can be difficult, as even the most diligent and seasoned of researchers can fall foul of the inherent pitfalls.

Literature search

In order to provide an unbiased report of the current literature available, the literature search must be exhaustive and include all relevant trials, including the grey literature (i.e. unpublished studies or studies published in non-Medline journals). However within the field of acupuncture there may be some bias, which could confound the outcome. Vickers *et al.* [12] produced a review that examined the results of acupuncture RCTs published in different countries. They examined a total of 252 trials from 27 countries between 1966 and 1995 and found that 10 countries produced only positive results in the published literature. The most prolific producer of trials that fit the criteria used by Vickers *et al.* was the USA, with 47 trials, 53% of which showed a positive outcome. Second in the list was China, with 36 published trials, 100% of which showed positive results. They noted that Russia/USSR also had a high positive publication rate, 10 out of 11 trials (91%) being positive. The trials from East Asia and Eastern Europe as a whole tended to show only positive results. This indicates a clear publication bias by country and therefore we must take this into account in our interpretation of the outcome of any systematic review.

The White and Ernst review was comprehensive and searched the Medline, Embase, Cochrane Library and Ciscam databases for trials between 1966 and 1997, and they also searched their own files for grey literature, thus maximizing the potential for studies to be included. This was in contrast to the review by Mendelson [2], who gave no indication as to how the literature search was done or how exhaustive it was. However, despite White and Ernst's thoroughness, there were still a further four trials that were not mentioned in the review: Lee *et al.* [13], Ahonen *et al.* [14], Lundeberg *et al.* [15] and Wang [16]. These four studies were either Medlined or published in acupuncture and pain journals and should therefore have been easily accessible. As White and Ernst produced a table of other trials that were excluded from their review, along with reasons, it is not known if these four were overlooked or excluded.

Ahonen *et al.* [14] dealt with headache and neck pain and Wang's [16] study focused on neck dysfunction. Neither trial was randomized and therefore would not have fulfilled White and Ernst's entry criteria. Lee *et al.* [13] dealt with pain from multiple sites. Lundeberg *et al.* [15] studied patients with pain from the head. However, the fact that these were not mentioned is a clear and unexplained omission. It is of course accepted, as White and Ernst [17] suggest, that conducting such a search is a difficult process and mistakes are inevitable.

Inclusion/exclusion criteria

The subject to be reviewed must be clearly stated and trials that are included in a systematic review must be appropriate to the subject being examined. This implies the use of great clarity and rigour with respect to the inclusion and exclusion criteria.

White and Ernst did indeed have a clear research agenda, i.e. to 'summarise the existing evidence for or against the hypothesis that acupuncture is an efficacious therapy for neck pain'. Their inclusion criteria for neck pain trials were that they had to be randomized and controlled, and they had to use used acupuncture, electro-acupuncture or lasers over acupuncture points. Their search revealed only 32 possibly relevant trials, 18 of which were excluded for various reasons, e.g. non-randomization or no control group. This left 14 trials, which White and Ernst scored using the 0–5 Jadad scoring system (see below). Those scoring two or less (poor quality) were: Junnila [18], Loy [19], Petrie and Langley [20], Petrie and Hazleman [21], Kisiel and Lindh [22], David *et al.* [23] and Irnich *et al.* [24] (this trial was unpublished when included in the systematic review). The higher-scoring studies are summarized in Table 2 (adapted from White and Ernst [9]).

This raises several issues in interpreting the trial data in relation to the primary aim of this systematic review. The inclusion of lasers is curious as this does not involve the use of needles and is therefore not acupuncture. It would seem reasonable, when specifically purporting to be examining acupuncture efficacy, to only include trials that actually used acupuncture, i.e. needle

TABLE 2. Higher-scoring trials from the White and Ernst review

Study (no. of subjects)	Design	Score	Treatment	Treatment method (no. of treatments) ^a	Control	Control method ^b	Result
Gallacchi <i>et al.</i> [30]	Parallel	3	15	Formula (8)	15	Sham acupuncture	Acu = sham
Coan <i>et al.</i> [40]	Parallel	3	15	Classical (36–48)	15	Waiting list	Acu > wait list
Emery and Lythgoe [31]	Crossover	4	10	Formula (3)	10	Dummy needling	Acu = dummy needling
Kreczi and Klingler [26]	Crossover	3	21	Laser to acupoints (1)	21	Sham laser	Laser > sham
Ceccherelli <i>et al.</i> [27]	Parallel	3	13	Laser to acupoints (12)	14	Sham laser	Laser > sham
Lundeberg <i>et al.</i> [38]	Parallel	3	14	Formula (1)	14	Superficial needling	Acu = superficial
Thomas <i>et al.</i> [39]	Crossover	3	44	Formula (1)	44	Superficial needling	Acu = superficial

^a‘Formula’ indicates that specific, preset acupuncture points were used, whereas ‘classical’ suggests a much more flexible approach, as in traditional Chinese acupuncture.

^b‘Sham’ indicates needling at non-acupuncture points; ‘dummy needling’ refers to some form of ‘staged’ acupuncture in which no penetration occurred; ‘superficial needling’ suggests minor penetration (usually 2–3 mm) at acupuncture sites.

puncture, as a treatment. The logic behind including forms of treatment other than acupuncture is questionable. Cummings [25] pointed out that, to combine studies that use an active and invasive treatment, such as needling, with others in which the intervention is imperceptible, such as laser treatment, is ‘clearly a mistake’. This is particularly pertinent when including an entirely unproven treatment, such as laser treatment, which, if it is specifically effective, may be mobilizing an entirely different neurological and neurohumoural mechanism from that activated by acupuncture. White and Ernst later justified their inclusion of laser therapy by suggesting that ‘laser is used as therapy and patients need to know if there is evidence that it is effective’ [17]. Patients do indeed need to know about this form of therapy. However, they also need to know about the use of ultrasound, pulsed electromagnetic energy, cervical collars and a host of other interventions, but there would similarly be little justification or value in including trials on these subjects in a review of acupuncture. The studies by Kreczi and Klingler [26] and the trial by Ceccherelli *et al.* [27] both only used laser to acupuncture points and therefore should also be excluded. This reduces the number of reviewable trials to 12. The systematic review of Smith *et al.* [10] also included laser treatment, in our view incorrectly, whereas ter Riet *et al.* [5] and Ezzo *et al.* [11] deliberately excluded this type of treatment.

As efficacy may vary from condition to condition [28, 29], it may well be very prudent to limit reviews to one specific painful site or pathology. White and Ernst [9] rightly excluded two trials because there was pain in multiple sites. However, of the 14 trials that were included in the review, four should not have been because they involved neck and/or back pain—Gallacchi *et al.* [30], Junnila [18], Emery and Lythgoe [31] and Kreczi and Klingler [26]; three of these attained higher Jadad scores in White and Ernst’s review. If the authors’ stated exclusion criteria had been followed, a further three trials would also have been excluded, leaving nine. Ezzo *et al.* [11], in their review of chronic pain, similarly combined multiple diagnoses and pain from multiple

sites. Whilst this would appear to be a reasonable step if specifically looking at chronic pain, it does not further illuminate the question of efficacy if acupuncture is indeed condition-specific. Patel *et al.* [4], however, whilst still examining acupuncture for chronic pain in general, actually split specific conditions into discrete subgroups, thereby enabling separate analysis. Similarly, there might also be some value in analysing trials separately according to follow-up period, i.e. the acute effects of acupuncture might be different from the longer-term effects, and this might have repercussions on the results of a systematic review.

Treatment

The type of treatment given is vitally important and must be such that it can be considered adequate by general consensus of those who use acupuncture. The question of what is adequate is, of course, a matter for debate and may vary widely from practitioner to practitioner, but this important problem needs to be addressed. A method for developing treatment protocols, the ‘BRITS’ method, has been suggested as a way forward for improving the quality of treatment in acupuncture trials [32]. It recommends, among other things, a comprehensive review of the literature in order to ascertain an optimum consensus about the treatment protocol. Clearly, trials with too few treatment sessions, too few needles and an unusual or unacceptable needling technique must be considered suboptimal; this might be analogous to treating an infection with one small dose of an inappropriate antibiotic. As Birch [33] stated, ‘a clinical trial will be a fair test of acupuncture only if an adequate treatment is administered’. If the number of needles used is too few, this could seriously impair efficacy, as often the combination of acupuncture points used may give a much more powerful effect than each point used individually [34]. It is also thought that acupuncture may have a cumulative effect [11, 35]; if this is so, treatment regimes must reflect clinical practice if we are to arrive at a realistic evaluation of its clinical

effects. Ironically, White and Ernst in an earlier paper [36] stated that 'It is important to examine the adequacy of acupuncture treatments to correctly assess studies in the field and to develop improved protocols for future research'.

Indeed, Ezzo *et al.* [11] have suggested that six sessions, each using six needles, are consistently associated with positive outcomes. Alternatively, Stux and Birch [37] have suggested that a minimum of 11 points and 10 sessions should be used for the treatment of neck pain. Returning to the White and Ernst [9] example, the trial by Emery and Lythgoe [31] must once again be excluded as this used three sessions; Kreczi and Klingler [26], Lundeberg *et al.* [38] and Thomas *et al.* [39] used only one session and the trials by Coan *et al.* [40], Loy [19] and Irnich *et al.* [24] do not provide sufficient information to make an informed decision as to their adequacy, although it is noted that, in the table produced by White and Ernst [9], the study of Coan *et al.* shows that 36–48 treatment sessions were given. However, there is no mention of this in the original paper by Coan *et al.* [40]. This further reduces the number of admissible trials to four. Whilst many of the reviews that included neck pain trials (Table 1) noted the specific treatment given, no author made an analysis of the adequacy of the treatment or excluded trials on the grounds of poor treatment. Ter Riet *et al.* [5], in their scoring system, included a section entitled 'Adequacy of treatment'. However, this related to an inclusion of the description of treatment given rather than its adequacy as a therapy. Ezzo *et al.* [11] commented that 'issues of dosing and an adequate/optimal acupuncture procedure need to be examined' but unfortunately failed to do this in their systematic review. Patel *et al.* [4] suggested that the choice between formula and classical traditional Chinese acupuncture may influence the outcome, but failed to make detailed comment on this in his systematic review. This also highlights a problem relating to the experience of the investigator. If a trial or systematic review is undertaken, it is imperative that the investigator has a good working knowledge of the complexities of the treatment regime used so that informed and logical decisions can be made as to its adequacy and therefore external validity. It is also vital, therefore, that a clear methodology for evaluating treatment adequacy is developed and incorporated into further systematic reviews of acupuncture [41]. Such standards must be developed strictly in consultation with the field, and should recognize the different modes of practice used in the studies under review.

Controls

The choice of control is important as this enables different questions to be answered. It is felt that if the question to be answered relates to the specific efficacy of acupuncture *vs* placebo or inactive treatment, it would be reasonable to exclude trials which used an active control as this may trigger analgesic mechanisms, such

as diffuse noxious inhibitory control (DNIC) [42, 43]. This implies that sham acupuncture may not be an appropriate placebo/control [44] to determine the specific efficacy of acupuncture in the context of an RCT. Once again, returning to the White and Ernst review [9], the trial by Gallacchi *et al.* [30] used sham acupuncture as the control whereas Lundeberg *et al.* [38] and Thomas *et al.* [39] used superficial acupuncture, which, it could be argued, might still have a physiological effect [45] through the mechanism of DNIC. Loy [19], Kisiel and Lindh [22] and David *et al.* [23] all used an alternative but active treatment, i.e. physiotherapy. Lastly, the trial by Coan *et al.* [40] used a waiting list control and therefore did not control for non-specific effects. Of the original 14 trials therefore included in this systematic review, only two, i.e. Petrie and Langley [20] and Petrie and Hazleman [21], were appropriate for inclusion in order to answer the question posed. We feel that two studies are insufficient to conduct a meaningful systematic review. These two trials incidentally provide contradictory evidence as to the efficacy of acupuncture for the treatment of neck pain.

The nature and format of a study to be reviewed must also be taken into account and the wisdom of including it must be examined when attempting to answer a specific question relating to efficacy. Three of the 14 trials included by White and Ernst employed a crossover design. Whilst this in itself is not a flaw, it is not known whether the effects (if any) of acupuncture are short-lived or more long-term. In a crossover study there may therefore be problems with cross-contamination of results due to carry-over. A washout period of several months may be appropriate to negate any such possible effect [44, 46], and therefore the inclusion and subsequent interpretation of crossover trials must be accompanied by guarded and cautious conclusions. Interestingly, Smith *et al.* [10] circumvented this problem in their systematic review by only considering data from the first treatment period where a crossover trial was used.

Scoring

It is useful to assign scores to trials in order to ascertain their relative credibility and assess their methodological competence. Failure to consider trial quality may certainly introduce bias into the results of any pooled analysis [47], and if the prime studies are flawed then any conclusions drawn by a meta-analysis will be invalid [48]. However, if this is to have any real meaning, the trials must be homogeneous and therefore comparable, and the scoring system must be relevant to the pragmatic clinical practice of acupuncture as well as to the carefully defined rigorous science of an RCT.

The 14 studies reviewed by White and Ernst were scored using a modified (and therefore unvalidated) Jadad scale [49] of 0–5, with one point being allocated for each aspect of good design, as follows:

- (i) study described as randomized: 1 point;
- (ii) adequate randomization technique: 1 point (if method inappropriate, deduct 1 point);
- (iii) subject blinding (i.e. control indistinguishable from acupuncture): 1 point;
- (iv) evaluator blinded to treatment: 1 point;
- (v) description of withdrawals and dropouts: 1 point.

White and Ernst report that no study achieved 5 points; only one achieved 4 points (good quality), six scored 3 (acceptable) and the others scored 2 or less (poor quality). All of the 'better' studies (Table 2) had low numbers of subjects, suggesting low statistical power and a great risk of a type II error [50].

According to White and Ernst, the trials by Petrie and Langley [20] and Petrie and Hazleman [21] scored two each on the Jadad scale. Ezzo *et al.* [11], using the same system, scored the first of these as 1. We would argue that the second of these trials should have attained a score of 3. It was described as randomized, although the method of randomization was not stated. There was a clear description of withdrawals and a blinded assessor was used. The question of adequate patient blinding in this trial is also contentious. White and Ernst suggest that this trial had poor blinding. This cannot be substantiated from the original papers as, although the active and control treatments were different, both groups (allegedly) believed that they were receiving real treatment, although this was not formally tested. These anomalies highlight the problems of using an insensitive scoring regime in acupuncture trials where no convincing placebo exists and where the issue of therapist blinding similarly remains a problem. Indeed, inter-rater reliability for the Jadad scale has been tested formally [51] and shown to be poor, yielding a kappa score of 0.37–0.39. It also raises real and confusing issues in relation to the authors' interpretation of the Jadad scale and their subsequent individual scores for each study.

Internal vs external validity

Unfortunately, studies using the Jadad scale are perhaps prone to drawing the wrong conclusions if it is thought that this scale, of itself, is a reflection of the quality of the trial in general. Jadad *et al.* [49] suggested that assessing the validity of the primary studies is one of the key components of systematic reviews. In designing their scoring mechanism, they stated that their aim was to be able to assess the scientific 'quality, defined as the likelihood of the trial design to generate unbiased results and approach the therapeutic truth'. Although Jadad *et al.* did indeed use the word 'quality', it is important that their definition of quality and its inherent limitation is taken on board if this scoring mechanism is to be used. The emphasis must be on the generation of unbiased results in terms of randomization and blinding, as this is what their score measures and nothing else. It does not give any indication as to the overall quality or validity of a trial, neither does it give any general

indication of 'therapeutic truth', unless such truth rests completely and solely in the ability of a trial to be unbiased. Jadad *et al.* specifically state that factors such as 'clinical relevance' are not assessed by their measure. As factors other than bias are deliberately omitted, it is of prime importance that those using this scale to assess 'quality' do not extrapolate their findings and suggest that a trial is good or bad simply by viewing the score. To do so would be to take a simplistic view of the research process and would certainly not be sensitive to many of the specific problems associated with acupuncture research. Jadad *et al.* tested their instrument by having 36 trials assessed by 14 raters. Whilst all of these were 'pain' trials, Jadad *et al.* do not give any indication as to the nature of these trials; for example, they may all have been pharmaceutical trials, which would not have highlighted some of the problems associated with manual treatments. Unfortunately, in their discussion Jadad *et al.* seem to move away from their previously rigid and well thought out aims of producing a bias-sensitive score when they suggest that the scoring mechanism would be useful for patients to evaluate the evidence presented to them by health professionals. Whilst researchers are trained and have experience in evaluating evidence and therefore must be expected to be able to understand the complexities and pitfalls involved in weighing up evidence presented in a trial, the average patient could easily be misled if it is suggested that the Jadad score provides a tool to assess the validity of presented evidence, as this implies a global reflection of quality and trial design. A trial with a high score on the Jadad system, whilst being unbiased in terms of randomization and blinding, might be subject to a multitude of other methodological problems and errors, which would cast doubt on the issue of internal validity. As has been pointed out, validity depends on much more than the proper conduct of the randomization process [52]. If a trial, and therefore a systematic review, is to be relevant, it must also have external validity and it must therefore be possible to extrapolate the results in a way that is meaningful for clinical practice. The treatment used must be pragmatic and the outcomes must be validated and relevant to the question being asked. If these important factors are not inherent within a scoring mechanism, they must at least be taken into consideration during the inclusion/exclusion process if they are to have any external validity. Failure to take these factors into account is a misrepresentation of the truth and of the underlying science.

There is a plethora of scoring mechanisms and guidance scales that can be employed to grade RCTs, and a cursory glance at the literature reveals at least 26 different scoring techniques [5, 51–75]. Some of these are very narrow in their scope and concentrate on a few items, e.g. three in the Jadad system, whereas others contain many more criteria, e.g. 34 items for a scale developed by Reisch *et al.* [72]. Moher *et al.* [76] have pointed out that when scales assign higher scores for double blinding, this 'automatically discriminates against trials in which masking may be inappropriate

TABLE 3. Summary of conclusions by Ernst and White [7]

Reference	Authors	Quotation from Conclusions
79	Bhatt-Sanders	'methodological flaws invalidate the conclusions'
3	Richardson and Vincent	'the call must be made for more and better studies before firm conclusions can be drawn'
80	Filshie and Morrison	'acupuncture merits serious consideration and further investigation'
4	Patel <i>et al.</i>	'bias precluded a conclusive finding'
5	ter Riet <i>et al.</i>	'The efficacy of acupuncture in the treatment of chronic pain remains doubtful'
81	Resch and Ernst	'the effectiveness of acupuncture has still not been definitely demonstrated'

or impossible'. This would, of course, be particularly relevant to acupuncture trials, where the complexities of a treatment regime are fundamentally different to those found in pharmacological trials. Perhaps it would be more useful in this instance to use a scale which is more sensitive to the subject matter and takes these complexities into account. If scoring is to play an integral part in the overall process of a systematic review, then the choice of instrument must be made with thoughtful consideration for the subject matter. This in itself can be fraught with difficulties, as the choice of scoring mechanism can alter the outcome of the systematic review or meta-analysis. Juni *et al.* [48] conducted a study which examined 25 scoring scales, where they repeated a meta-analysis using these different scoring measures in order to assess whether the choice of scale affected the outcome of the meta-analysis. The meta-analysis they selected was one dealing with a comparison between standard heparin and low molecular weight heparin (LMWH) for the prevention of post-operative thrombosis. They calculated the median score for each trial (17 trials in total) and expressed this as a percentage of the total score for each system. This gave a very large spread of results, ranging from 38.5 to 82.9%, thus illustrating the large discrepancies between these scales. When trials were weighted for quality by each scale, they found that six scales showed that LMWH was not superior to standard heparin, whilst a further seven scales showed that it was superior. Therefore, depending on the scale chosen, the effect size could be manipulated to show either a positive or negative result. They further commented that trial quality could be affected by characteristics such as the setting, the characteristics of the patients and treatments, and that the incorporation of quality scores as weights lacks statistical or empirical justification. They concluded by suggesting that relevant methodological aspects should be identified and assessed individually.

Conclusion

It can be seen that, despite many trials, the question of the efficacy of acupuncture for the treatment of chronic mechanical neck pain remains unanswered. The feel of the state of knowledge within acupuncture research at the present moment was captured by Ernst and White [7] in their review, and their conclusions are summarized in Table 3.

We feel that the only way forward in the search for answers in relation to the clinical effectiveness of acupuncture is not to re-examine existing trials in systematic reviews, as clearly there are too few trials of sufficient quality and homogeneity to be able to draw any conclusions at all and the strength of any systematic review will ultimately rely on the quality of the primary research on which it is based [77]. Instead, the focus must be on actually conducting more clinical trials. A good systematic review can only build on the foundations provided by well-conducted trials. If the trials do not exist or are of poor quality then it does not matter how many times they are reviewed—they will consistently fail to produce meaningful conclusions. Systematic reviews and meta-analysis are important tools as many people use these to gain an overall sense of how this therapy may be used and valued in clinical practice. Policy foundation for evidence-based practice guidelines, economic evaluations and research agendas may be based on this important area of research. But this, as in any other area of research, may be prone to flaws [78]. It therefore behoves those who conduct these reviews to do so responsibly and with sensitivity to the subject matter. They must use logical and defensible inclusion and exclusion criteria, some of which must include a careful and clinically relevant consideration of the treatment used. Any scoring system used must be able to differentiate between trials of good methodology (based on current clinical practice), which may have tackled these problems logically, and those that are simply biased and of poor quality. Lastly, it must be borne in mind that many 'scientific' principles are not cast in stone to be propounded as absolute truths but are open to interpretation and must be used with caution and updated in the light of further research.

Acknowledgements

Peter White was funded by the The Henry Smith Charity (UK) and the Hospital Savings Association (UK).

References

- Downs S, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52:377–84.

2. Mendelson G. Acupuncture analgesia. I. Review of clinical studies. *Aust N Z J Med* 1977;7:642–8.
3. Richardson PH, Vincent CA. Acupuncture for the treatment of pain: a review of evaluative research. *Pain* 1986;24:15–40.
4. Patel M, Gutzwiller F, Paccaud F, Marazzi A. A meta-analysis of acupuncture for chronic pain. *Int J Epidemiol* 1989;18:900–6.
5. ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990;43:1191–9.
6. Aker PD, Gross AR, Goldsmith CH, Peloso P. Conservative management of mechanical neck pain: systematic overview and meta-analysis. *Br Med J* 1996;313:1291–6.
7. Ernst E. Acupuncture as a symptomatic treatment of osteoarthritis. A systematic review. *Scand J Rheumatol* 1997;26:444–7.
8. Kjellman G, Skargren E, Oberg B. A critical analysis of randomized clinical trials on neck pain and treatment efficacy. A review of the literature. *Scand J Rehab Med* 1999;31:139–52.
9. White AR, Ernst E. A systematic review of randomized controlled trials of acupuncture for neck pain. *Rheumatology* 1999;38:143–7.
10. Smith L, Oldman A, McQuay R, Moore A. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. *Pain* 2000;86:119–32.
11. Ezzo J, Berman BM, Hadhazy V, Jadad A, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86:217–25.
12. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19:159–66.
13. Lee PK, Anderson TW, Modell JH, Saga SA. Treatment of chronic pain with acupuncture. *J Am Med Assoc* 1975;232:1133–5.
14. Ahonen E, Hakumaki M, Mahlamaki S, Partanen J, Riekkinen P, Sivenius J. Effectiveness of acupuncture and physiotherapy on myogenic headache: a comparative study. *Acupunct Electrother Res* 1984;9:141–50.
15. Lundeberg T, Hurtig T, Lundeberg S, Thomas M. Long-term results of acupuncture in chronic head and neck pain. *Pain Clinic* 1988;2:15–31.
16. Wang D. Seventy-five cases of stiff neck treated by acupuncture at acupoint yanglao (SI 6). *J Tradit Chin Med* 1994;14:269–71.
17. White A, Ernst E. Systematic reviews of acupuncture — is a more profitable discussion possible? *Clin Acupunct Oriental Med* 2001;2:111–5.
18. Junnila SYT. Acupuncture therapy for chronic pain. A randomized comparison between acupuncture and pseudo-acupuncture with minimal peripheral stimulus. *Am J Acupunct* 1982;10:259–62.
19. Loy TT. Treatment of cervical spondylosis. Electroacupuncture versus physiotherapy. *Med J Aust* 1983;2:32–4.
20. Petrie JP, Langley GB. Acupuncture in the treatment of chronic cervical pain. A pilot study. *Clin Exp Rheumatol* 1983;1:333–6.
21. Petrie JP, Hazleman BL. A controlled study of acupuncture in neck pain. *Br J Rheumatol* 1986;25:271–5.
22. Kisiel C, Lindh C. Smartlindring med fysikalisk terapi och manuell akupunktur. *Sjukgymnasten* 1996;12:24–31.
23. David J, Modi S, Aluko A, Robertshaw C, Farebrother J. Chronic neck pain: a comparison of acupuncture treatment and physiotherapy. *Br J Rheumatol* 1998;37:1118–22.
24. Irnich D, Behrens N, Molzen *et al.* Randomized, placebo-controlled, multicentre trial of acupuncture for the treatment of chronic neck pain. *Res Complement Natural Classical Med* 7;43(abstract).
25. Cummings M. Teasing apart the quality and validity in systematic reviews of acupuncture. *Acupunct Med* 2000;18:104–7.
26. Kreczi T, Klingler D. A comparison of laser acupuncture versus placebo in radicular and pseudoradicular pain syndromes as recorded by subjective responses of patients. *Acupuncture Electrother Res* 1986;11:207–16.
27. Ceccherelli F, Altafini L, Lo Castro G, Avila A, Ambrosio F, Giron G. Diode laser in cervical myofascial pain: a double blind study versus placebo. *Clin J Pain* 1989;5:301–4.
28. Ernst E. Is acupuncture effective for pain control? (letter). *J Pain Symptom Manage* 1994;9:72–4.
29. Lewith G, Vincent C. Evaluation of the Clinical Effects of Acupuncture. A problem Reassessed and a Framework for Future Research. *Pain Forum* 1995;4:29–39.
30. Gallacchi G, Muller W, Plattner C, Schnorrenberger C. Akupunktur und Laserstrahlbehandlung beim Zervikal und Lumbalsyndrom. *Schweiz med Wschr* 1981;111:1360–6.
31. Emery P, Lythgoe S. The effect of acupuncture on ankylosing spondylitis. *Br J Rheumatol* 1986;25:132–3.
32. Birch S. Testing the clinical specificity of needle sites in controlled clinical trials of acupuncture. *Proceedings of the Second Annual Meeting, Society for Acupuncture Research* 1995:274–94.
33. Birch S. Issues to consider in determining an adequate treatment in a clinical trial of acupuncture. *Complement Ther Med* 1997;5:8–12.
34. Hopwood V, Lovesey M, Mokone S. *Acupuncture and related techniques in physical therapy*. Edinburgh: Churchill Livingstone, 1997.
35. Pomeranz B. Bruce Pomeranz, PHD. Acupuncture and the raison d'être for alternative medicine [interview by Bonnie Horrigan]. *Altern Ther Health Med* 1996;2:85–91.
36. White AR, Ernst E. A trial method for assessing the adequacy of acupuncture treatments. *Altern Ther Health Med* 1998;4:66–71.
37. Stux G, Birch S. Proposed standards of acupuncture treatment for clinical studies. In Stux G, Hammerschlag R (eds) *Clinical acupuncture: Scientific basis*, pp. 171–85. Berlin: Springer-Verlag, 2000.
38. Lundeberg T, Eriksson SV, Lundeberg S, Thomas M. Effect of acupuncture and naloxone in patients with osteoarthritis pain. A sham acupuncture controlled study. *The Pain Clinic* 1991;4:155–61.
39. Thomas M, Eriksson SV, Lundeberg T. A comparative study of diazepam and acupuncture in patients with osteoarthritis pain: a placebo controlled study. *Am J Chin Med* 1991;19:95–100.
40. Coan RM, Wong G, Coan PL. The acupuncture treatment of neck pain: a randomized controlled study. *Am J Chin Med* 1982;9:326–32.
41. Birch S. Systematic reviews of acupuncture — are there problems with these? *Clin Acupunct Orient Med* 2001;2:17–22.

42. Le Bars D, Villanueva L, Willer J, Bouhassira D. Diffuse noxious inhibitory controls (DNIC) in animals and man. *Acupunct Med* 1991;9:47–56.
43. Bing Z, Villanueva L, Le Bars D. Acupuncture and diffuse noxious inhibitory controls: naloxone-reversible depression of activities of trigeminal convergent neurons. *Neuroscience* 1990;37:809–18.
44. Lewith GT, Machin D. On the evaluation of the clinical effects of acupuncture. *Pain* 1983;16:111–27.
45. Vincent C, Lewith G. Placebo controls for acupuncture studies. *J R Soc Med* 1995;88:199–202.
46. Lewith GT. Can we assess the effects of acupuncture? [editorial]. *Br Med J Clin Res* 1984;288:1475–6.
47. Moher D, Pham B, Jones A *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? *Lancet* 1998;352:609–13.
48. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *J Am Med Assoc* 1999;282:1054–60.
49. Jadad A, Moore R, Carrol D *et al.* Assessing the quality of reports of randomised clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
50. Patel MS. Problems in the evaluation of alternative medicine. *Soc Sci Med* 1987;25:669–78.
51. Clark H, Wells G, Huet C *et al.* Assessing the quality of randomized trials: reliability of the Jadad scale. *Control Clin Trials* 1999;20:448–52.
52. Chalmers T, Smith H, Blackburn B *et al.* A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;2:31–49.
53. Andrew E. Method for assessment of the reporting standard of clinical trials with Roentgen contrast media. *Acta Radiol Diagn* 1984;25:55–8.
54. Beckerman H, de Bie R, Bouter L, de Cuyper H, Oostendorp R. The efficacy of laser therapy for musculoskeletal and skin disorders. *Phys Ther* 1992;72:483–91.
55. Brown S. Measurement of quality of primary studies for meta-analysis. *Nurs Res* 1991;40:352–5.
56. Chalmers I, Adams M, Dickersin K. A cohort study of summary reports of controlled trials. *J Am Med Assoc* 1990;263:1401–5.
57. Cho M, Bero L. Instruments for assessing the quality of drug studies published in the medical literature. *J Am Med Assoc* 1994;272:101–4.
58. Colditz G, Miller J, Mosteller F. How study design affects outcomes in comparisons of therapy. *Stat Med* 1989;8:441–54.
59. Detsky A, Naylor C, O'Rourke K, McGeer A, L'Abbe K. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255–65.
60. Evans M, Pollock A. A score system for evaluating random controlled clinical trials of prophylaxis of abdominal surgical wound infection. *Br J Surg* 1985;72:256–60.
61. Goodman S, Berlin J, Fletcher S, Fletcher R. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994;121:11–21.
62. Gotzsche P. Methodology and overt and hidden bias in reports of 196 double blind trials of non-steroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31–56.
63. Imperiale T, McCullough A. Do corticosteroids reduce mortality from alcoholic hepatitis. *Ann Intern Med* 1990;113:299–307.
64. Jonas W. The likelihood of validity evaluation method. Unpublished Work: cited in [76].
65. Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homeopathy. *Br Med J* 1991;302:316–23.
66. Koes BW, Assendelft W, van der Heijden G, Bouter L, Knipschild P. Spinal manipulation and mobilisation for back and neck pain: a blinded review. *Br Med J* 1991;303:1298–303.
67. Levine J. Trial assessment procedure scale (TAPS). In Spilker B (ed), *Guide to clinical trials*, pp. 780–6. New York: Raven Press, 1991.
68. Linde K, Clausius N, Ramirez G. Are the clinical effects of homeopathy placebo effects. *Lancet* 1997;350:834–43.
69. Nurmohamed M, Rosendaal F, Buller H. Low molecular weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152–6.
70. Onghena P, Van Houdenhove B. Antidepressant induced analgesia in chronic non-malignant pain. *Pain* 1992;49:205–19.
71. Poynard T. Evaluation de la qualité méthodologique des essais thérapeutiques randomisés. *Presse Med* 1988;17:315–8.
72. Reisch J, Tyson J, Mize S. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989;84:815–27.
73. Smith K, Cook D, Guyatt G, Madhavan J. Respiratory muscle training in chronic airflow limitation: a meta-analysis. *Am Rev Respir Dis* 1992;145:533–9.
74. Spitzer W, Lawrence V, Dales R. Links between passive smoking and disease. *Clin Invest Med* 1990;13:17–42.
75. Hammerschlag R, Morris MM. Clinical trials comparing acupuncture with biomedical standard care: a criteria-based evaluation of research design and reporting [corrected] [published erratum appears in *Complement Ther Med* 1997;5:253]. *Complement Ther Med* 1997;5:133–40.
76. Moher D, Jadad A, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16:62–73.
77. Khan K, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996;156:661–6.
78. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354:1896–900.
79. Bhatt-Sanders D. Acupuncture for rheumatoid arthritis: an analysis of the literature. *Arthritis Rheum* 1985;14:225–31.
80. Filshie J, Morrison P. Acupuncture for chronic pain: a review. *Palliative Med* 1988;2:1–14.
81. Resch KL, Ernst E. Wirksamkeitsnachweise Komplementärer Therapien. *Fortsschr Med* 1995;113:41–7.