Background: Acupuncture is commonly used to treat back pain, but there is no published meta-analysis of trials of its effectiveness for this condition.

Objective: To perform a meta-analysis of trials of acupuncture for the treatment of back pain.

Methods: A systematic literature search was conducted to retrieve all randomized controlled trials of any form of acupuncture for any type of back pain in humans. The adequacy of the acupuncture treatment was assessed by consulting 6 experienced acupuncturists. The main outcome measure for the meta-analysis was numbers of patients whose symptoms were improved at the end of treatment.

Results: Twelve studies were included, of which 9 presented data suitable for meta-analysis. The odds ratio of improvement with acupuncture compared with control intervention was 2.30 (95% confidence interval, 1.28-4.13). For sham-controlled, evaluator-blinded studies, the odds ratio was 1.37 (95% confidence interval, 0.84-2.25).

Conclusion: Acupuncture was shown to be superior to various control interventions, although there is insufficient evidence to state whether it is superior to placebo.
METHODS

DATA SOURCES

Searches were performed for clinical trials of any form of acupuncture for back pain in 3 computerized databases: MEDLINE (1969-1996); the Cochrane Controlled Trials Register (Issue 1, 1997); and CISCOM (November 1996), a database specializing in complementary medicine including much of the “gray literature,” such as unpublished studies and conference reports. Searches were performed using keywords acupuncture, electroacupuncture, and backache and text-word searches for the above terms together with low-back pain and lumbago. In addition, our files of published articles selected over the years were screened and several experts in different countries were invited to contribute published studies on the topic. The bibliographies of all articles thus retrieved together with reviews of acupuncture treatment of pain were reviewed for further references. Authors of articles published in the past 5 years (1992-1997) were contacted and asked to inform us of any other articles they were aware of. Authors of abstracts were contacted and asked to provide the full reports.

STUDY SELECTION

All articles that reported a randomized controlled trial in which dry needles were inserted into the skin and for which the process was described by the author(s) as “acupuncture” for the treatment of any type of back pain in humans were included. The basis for selecting points for needle insertion was not restricted. Articles published in the English, French, German, Spanish, Italian, or Polish language were included. Trials in which one form of acupuncture was compared with another were excluded. When more than 1 publication described a single trial, only 1 report was included.

DATA EXTRACTION

Data were extracted independently by both of us in a predefined, standardized manner. Differences were settled by discussion. For each trial, numbers of patients in experimental and control groups who were objectively rated as improved or who had returned to work were obtained. Where objective measures were not available, subjective ratings were used. Where more than 1 outcome measure was reported in this way, the least and most favorable results were extracted for separate evaluations. Where necessary, letters were sent to authors requesting these data.

ASSESSMENT OF METHODOLOGICAL QUALITY

The quality of the studies was assessed by a modification of the method described by Jadad et al.11 First, points were awarded by the investigators (E.E. and A.R.W.) in 3 categories: randomization (2 points), blinding (2 points), and description of dropouts and withdrawals (1 point). Studies were rated as “blind” if the control group received an intervention that appears likely to be indistinguishable from acupuncture (ie, sham acupuncture) and if the outcome was assessed in a blind manner, whether by a blinded observer or by the blinded subject’s self-report. Second, points up to a maximum of 2 were awarded for each study according to the adequacy of the acupuncture treatment used, as assessed by blinded experts. For this purpose, an extract was prepared for each study containing only details of patients and interventions, translated into English if necessary. All means of identification were removed. Six experienced medical acupuncturists involved both in clinical practice and in teaching studied these extracts and rated the adequacy of acupuncture in each trial on a visual analog scale (VAS) that consisted of a 100-mm line. The left end of the line was labeled “complete absence of evidence that the acupuncture was adequate” and the right end was labeled “total certainty that the acupuncture was adequate.” Two points were awarded for mean VAS scores of more than 66, one point for scores between 33 and 66, and zero points for scores less than 33 mm.

DATA SYNTHESIS

Meta-analyses of the data were performed using custom-written software (RevMan 3.0, Cochrane Collaboration, Oxford, England). For each trial, the total numbers of patients included in each group and the numbers whose symptoms had improved were entered into tables for comparison. Subsequently, the odds ratio (OR) for each trial was calculated, ie, the ratio of successes to failures in the genuine treatment group, divided by the same ratio in the control group. Weighting studies according to their inverse variance, the program computed a combined OR for all trials in the comparison, using the random effects model of DerSimonian and Laird.14 Confidence intervals (CIs) were calculated from the sums of the individual variances, and were set at 95%. The primary meta-analysis combined all studies that contained data in the appropriate form. Further meta-analyses were performed to compare subsets of trials that were combined according to particular features: blinding, length of follow-up, adequacy of acupuncture, use of formula acupuncture or electrical stimulation, and number of treatment sessions.

SENSITIVITY ANALYSIS

The primary meta-analysis was performed using the data that were least favorable to acupuncture. A sensitivity analysis was performed by repeating this meta-analysis using the most favorable data.

Table 3 summarizes the quality ratings of all included studies. The range is wide yet the majority of studies scored 3 or more points. The expert acupuncturists held divergent opinions about the adequacy of the treatment in all studies except 1,32 which they agreed was almost totally inadequate.
META-ANALYSES

Nine studies presented data in a form suitable for inclusion in the meta-analyses. The results of the primary meta-analysis are presented in Figure 1. A total of 377 patients was included in the trials, and the overall OR was 2.30 (95% CI, 1.28-4.13). There was no significant heterogeneity between studies ($\chi^2=12.58$, $P>1$).

In 3 studies, the outcome was markedly more positive than in the remainder. These studies have no uniformity of inclusion criteria, acupuncture approach, setting, or end points that could account for the divergence. Alternative (more favorable) outcome data were available in 1 study, yielding a new OR for all studies favoring acupuncture, and Gallacchi et al showed no superiority of acupuncture over sham acupuncture. The results of the meta-analyses of studies grouped according to design features are given in Figure 2. The OR of the 4 sham-controlled, evaluator-blinded studies was 1.37 (95% CI, 0.84-2.25). The results were not meaningfully affected by length of follow-up, quality of acupuncture, type of acupuncture, or number of sessions. There are trends worthy of discussion but CIs of comparisons overlap, indicating that no firm conclusions can be drawn from these data.

META-ANALYSES OF SUBGROUPS

The results of the meta-analyses of studies grouped according to design features are given in Figure 2. The OR of the 4 sham-controlled, evaluator-blinded studies was 1.37 (95% CI, 0.84-2.25). The results were not meaningfully affected by length of follow-up, quality of acupuncture, type of acupuncture, or number of sessions. There are trends worthy of discussion but CIs of comparisons overlap, indicating that no firm conclusions can be drawn from these data.

STUDIES EXCLUDED FROM THE META-ANALYSIS

Three studies presented results in a form that was unsuitable for inclusion in this meta-analysis: Macdonald et al concluded that acupuncture was significantly superior to control intervention, Yue found a positive trend favoring acupuncture, and Gallacchi et al showed no superiority of acupuncture over sham acupuncture. The total sample size of these 3 studies was 95. It seems unlikely that the omission of these studies seriously undermines the accuracy of the meta-analyses.

The funnel plot is presented in Figure 3. The studies are not evenly distributed around the combined OR, but there are too few studies to conclude whether publication bias has influenced the result of the meta-analysis.

COMMENT

Collectively, these data suggest that acupuncture is an effective treatment for back pain. The assessment of acupuncture for back pain is notoriously difficult for several reasons: there are many variations of acupuncture (eg, points used, method of stimulation) that are not necessarily comparable; back pain is not a distinct entity but an ill-defined category of complaints with diverse causes; acute back pain frequently disappears within days with or without treatment; concomitant treatments abound; and there is no objective, universally accepted outcome measure. The trials included in this analysis are heterogeneous in terms of study population, type of acupuncture used, outcome measure used, and length of follow-up. Thus, it is problematic to form a firm judgment.

Our search strategy was as comprehensive as possible. Yet, it is possible that some trials were not located. In particular, negative trials may not be published. We also suspect that complementary medicine journals are heavily biased toward positive results. Thus, systematic reviews could easily become distorted. Unfortunately, the present material gives no firm indication for or against the existence of a publication bias and we are unable confidently to exclude this source of error in our analysis.

One inclusion criterion for this analysis was randomization. We, therefore, dealt with trials of relatively high standard. The quality ratings (Table 3) show that the methodological quality was good in the majority of the studies. Only 2 trials, both published in the 1970s, were of low quality. Thus, the present meta-analysis is based largely on rigorous research, which lends weight to its findings.

All the studies recruited patients with chronic pain or who had failed to respond to conventional therapy. Only 2 studies specifically excluded patients with previous back surgery. Thus, the majority of study participants were associated with a poor prognosis and belonged to a category that is notoriously difficult to treat. The fact that nonetheless the overall result is positive suggests that acupuncture can be helpful even for difficult cases of back pain.

In most studies, the follow-up period was inadequate. Two studies suggest that the result measured immediately after treatment may be sustained if not improved (Figure 2). Unfortunately, this sample is too small to draw any conclusions, and the long-term effects of acupuncture on back pain remain uncertain. This, it seems, could be a fruitful area of future research.

The subgroup analyses revealed interesting trends. The OR of unblinded studies is larger than that of studies in which blinding had been introduced. This trend might indicate the importance of patient and therapist...
expectations in terms of clinical outcome and suggests that acupuncture (like most hands-on interventions) is associated with a powerful placebo effect. In 2 studies, acupuncture was rated as inadequate, yet tended to be more effective than that judged as adequate. This seems counterintuitive at first glance but might be explained quite easily. In both cases, the descriptions of the acupuncture were minimal. Difficulty in judging the trials was reflected in the large SDs of the experts’ scores (Table 3). Thus, the acupuncture technique was poorly described in these reports, but may still have been of good quality. Further, it is relevant to note that both individualizing the selection of points for treatment and repeating the acupuncture more than 4 times were associated with larger effect sizes. The latter trend would be compatible with the dose-response curve of an effective therapy.

In a review of 7 acupuncture trials of acupuncture for pain, including back pain, Deyo concluded that “acupuncture is little more effective than placebo therapy which mimics active treatment.” Our subgroup analysis of 4 sham-controlled, evaluator-blinded studies yielded a combined OR of 1.37 (95% CI, 0.84-2.25) in favor of genuine acupuncture. Although this finding is not conclusive, it does suggest that further trials are justified to determine whether acupuncture works through specific or nonspecific effects.

It has been suggested that sham acupuncture is an unfair comparison since needling the skin can relieve pain through the process of diffuse noxious inhibitory control. This effect can be minimized by needling non-meridian, nontender points. Future studies should include both sham-control groups and other comparison groups in parallel. The optimal study in acupuncture for back pain should be randomized and fully blinded, with an adequate acupuncture technique and follow-up. It should include a sample size based on power calculation, should evaluate success with a widely accepted outcome measure, and should test acupuncture on a homogeneous subtype of back pain that has previously been suggested to respond favorably. In light of the poor reporting quality of some acupuncture trials (see above), the optimal study should, of course, detail all aspects of the method accurately.

If one accepts that acupuncture is an effective form of treatment for back pain, one might ask about its value compared with other forms of treatment. So far, only exercise treatments, transcutaneous electrical nerve stimulation (TENS), and spinal manipulation have been evaluated in a systematic way. Koes et al concluded that it is “uncertain whether exercise is better than other conservative treatments for back pain or whether a specific type of back pain is more effective.” The review by Gadsby
and Flowerdew51 “provides some evidence to support the use of TENS.” Finally, a meta-analysis of spinal manipulation found that it was of short-term benefit.52 The last findings are so convincing that several national guidelines now recommend spinal manipulation as a first-line treatment.53,54 It is interesting, therefore, to see that the overall OR calculated by Shekelle et al52 for all 7 randomized control trials meta-analyzed is similar to the one emerging from this article (OR, 2.0 [95% CI, 1.48-2.77] for spinal manipulation compared with 2.30 [95% CI, 

<table>
<thead>
<tr>
<th>Intervention (No. of Treatments)</th>
<th>Primary End Points</th>
<th>Follow-up</th>
<th>Result (Significance)</th>
<th>No. Improved in Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula EA vs sham acup (3)</td>
<td>Evaluator’s rating of change† including pt report</td>
<td>None</td>
<td>No intergroup differences</td>
<td>7/15 vs 6/15</td>
</tr>
<tr>
<td>Formula acup vs sham acup vs physiotherapy (not stated)</td>
<td>Pain, range of movement</td>
<td>None</td>
<td>Acup was superior to physiotherapy but not to sham</td>
<td>NA vs NA</td>
</tr>
<tr>
<td>Formula acup + sodium chloride injection vs sham electrical stimulation (4)</td>
<td>Global assessment by 2 physicians†</td>
<td>None</td>
<td>Acup superior to controls (P&lt;.05)</td>
<td>14/18 vs 10/16</td>
</tr>
<tr>
<td>Individualized traditional acup vs waiting-list controls (11)</td>
<td>Pain score (combined pt and evaluator rating†)</td>
<td>10-15 wk</td>
<td>Pain reduction: acup, 51%; controls, 2%</td>
<td>19/23 vs 5/16</td>
</tr>
<tr>
<td>Needling at muscle motor point (mean 7.9) + standard physiotherapy vs physiotherapy alone</td>
<td>Pain and work status†</td>
<td>12-42 wk</td>
<td>Needling superior to controls (P&lt;.01)</td>
<td>18/29 vs 4/27</td>
</tr>
<tr>
<td>Formula acup vs 2 forms of sham acup vs 5 forms of laser acup (8)</td>
<td>VAS for pain</td>
<td>None</td>
<td>All groups improved, no intergroup differences</td>
<td>NA vs NA</td>
</tr>
<tr>
<td>Formula acup vs sham acup (5)</td>
<td>Pain VAS (1) on standing 10 min,†(2) on resting†</td>
<td>VAS for pain</td>
<td>Acup superior to sham for severe pain, supported by SLR changes</td>
<td>(1) 11/15 vs (2) 6/15</td>
</tr>
<tr>
<td>Superficial needling (with or without EA) vs sham TENS (10)</td>
<td>None</td>
<td>Pain reduction greater after needling (P&lt;.01)</td>
<td>(2) 13/15 vs (2) 5/15</td>
<td></td>
</tr>
<tr>
<td>Formula acup vs lidocaine injections (8)</td>
<td>VAS for pain (&gt;33% relief†)</td>
<td>4, 12, and 16 wk</td>
<td>No significant difference between groups</td>
<td>44/77 vs 41/77</td>
</tr>
<tr>
<td>Individualized EA (6) vs TENS (15) vs sham TENS (15): + education and exercise program</td>
<td>VAS for pain and disability + physician’s assessment + return to work†</td>
<td>Immediate and 6 mo</td>
<td>EA superior to TENS (P&lt;.11); no difference between TENS and sham TENS</td>
<td>10/17 vs 10/18</td>
</tr>
<tr>
<td>Trigger point needling vs vapocoolant spray + acupressure vs lidocaine injection vs lidocaine + corticosteroid injection (1)</td>
<td>Pt assessment of improvement†</td>
<td>None</td>
<td>Needling and acupressure yielded the best results (P&lt;.05)</td>
<td>11/20 vs 8/16</td>
</tr>
<tr>
<td>Flexible formula acup vs low-frequency EA vs high-frequency EA by pt choice (average 6.8) vs waiting-list controls</td>
<td>Activity related to pain, mobility, verbal descriptors of pain, pt assessment of improvement†</td>
<td>Immediate, 6 wk, 6 mo</td>
<td>After 6 wk all EA groups superior to untreated controls (P&lt;.05); after 6 mo this was only the case for low-frequency EA</td>
<td>17/30 vs 4/10</td>
</tr>
</tbody>
</table>

Table 3. Meta-analysis of Acupuncture for Back Pain: Quality Assessment Scores of Included Studies

<table>
<thead>
<tr>
<th>Source, y</th>
<th>Expert Assessment Score, Mean (SD)†</th>
<th>Adequacy of Acupuncture‡</th>
<th>Randomization</th>
<th>Blinding</th>
<th>Dropouts</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplan et al,17 1983</td>
<td>46.5 (30.6)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Edelstein et al,18 1976</td>
<td>49.0 (31.9)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Garvey et al,22 1989</td>
<td>31.3 (34.9)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Mendelson et al,39 1983</td>
<td>50.7 (33.1)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Thomas and Lundeberg,43 1994</td>
<td>69.5 (23.5)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Coan et al,40 1980</td>
<td>26.5 (34.2)</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Gunn et al,41 1980</td>
<td>53.3 (33.0)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Gallacicci et al,42 1981</td>
<td>45.3 (37.0)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Lehmann et al,43 1983</td>
<td>41.5 (26.8)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Macdonald et al,44 1983</td>
<td>43.3 (24.1)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Lopacz and Grealweyski,45 1979</td>
<td>9.3 (5.3)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Yue,44 1978</td>
<td>18.8 (35.6)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*In descending order of quality.
†Adequacy of acupuncture was assessed by expert panel on visual analog scale (VAS) of 0 to 100.
‡Points awarded for VAS score less than 33, zero points; VAS score of 33 to 66, 1 point; and VAS score greater than 66, 2 points.
safety and costs are other factors. In 2 recent surveys, we found adverse reactions of acupuncture to be considerably less frequent than those of spinal manipulation. Unfortunately, no fully conclusive data about the true prevalence of adverse effects for either method exist. This, it seems, is another area of important research for the future. The costs of acupuncture may well be lower than for spinal manipulation: treatment sessions are usually shorter and often less frequent. When 1020 episodes of back pain were investigated, chiropractors had a greater mean number of visits per episode than any other profession, which is supported by other data. Considering the socioeconomic importance of back pain, it would be relevant to establish the relative cost-effectiveness of all major treatments for this condition.

In conclusion, the combined result of all studies shows acupuncture to be superior to various control interventions. The combined results of 4 sham-controlled, evaluator-blinded studies did not show acupuncture to be superior to placebo; further studies are required to conclude with certainty whether acupuncture has specific effects in addition to its nonspecific effects.

Accepted for publication April 13, 1998.

We gratefully acknowledge V. Camp, MD, H. Dyczek, ACN, P. Marcus, MD, R. James, MD, D. Payne, MD, F. Staebler, MD, and A. Ward, MD, for help with aspects of this study.

Corresponding author: Edzard Ernst, MD, PhD, FRCP(Edin), Department of Complementary Medicine, School of Postgraduate Medicine and Health Science, University of Exeter, 25 Victoria Park Rd, Exeter EX2 4NT, England (e-mail: E. Ernst@ex.ac.uk).

REFERENCES