Acupuncture for Chronic Pain

Individual Patient Data Meta-analysis

Andrew J. Vickers, DPhil; Angel M. Cronin, MS; Alexandra C. Maschino, BS; George Lewith, MD; Hugh MacPherson, PhD; Nadine E. Foster, DPhil; Karen J. Sherman, PhD; Claudia M. Witt, MD; Klaus Linde, MD; for the Acupuncture Trialists’ Collaboration

Background: Although acupuncture is widely used for chronic pain, there remains considerable controversy as to its value. We aimed to determine the effect size of acupuncture for 4 chronic pain conditions: back and neck pain, osteoarthritis, chronic headache, and shoulder pain.

Methods: We conducted a systematic review to identify randomized controlled trials (RCTs) of acupuncture for chronic pain in which allocation concealment was determined unambiguously to be adequate. Individual patient data meta-analyses were conducted using data from 29 of 31 eligible RCTs, with a total of 17,922 patients analyzed.

Results: In the primary analysis, including all eligible RCTs, acupuncture was superior to both sham and no-acupuncture control for each pain condition (P < .001 for all comparisons). After exclusion of an outlying set of RCTs that strongly favored acupuncture, the effect sizes were similar across pain conditions. Patients receiving acupuncture had less pain, with scores that were 0.23 (95% CI, 0.13-0.33), 0.16 (95% CI, 0.07-0.25), and 0.15 (95% CI, 0.07-0.24) SDs lower than sham controls for back and neck pain, osteoarthritis, and chronic headache, respectively; the effect sizes in comparison to no-acupuncture controls were 0.55 (95% CI, 0.51-0.58), 0.57 (95% CI, 0.50-0.64), and 0.42 (95% CI, 0.37-0.46) SDs. These results were robust to a variety of sensitivity analyses, including those related to publication bias.

Conclusions: Acupuncture is effective for the treatment of chronic pain and is therefore a reasonable referral option. Significant differences between true and sham acupuncture indicate that acupuncture is more than a placebo. However, these differences are relatively modest, suggesting that factors in addition to the specific effects of needling are important contributors to the therapeutic effects of acupuncture.


A large number of randomized controlled trials (RCTs) of acupuncture for chronic pain have been conducted. Most have been of low methodologic quality, and, accordingly, meta-analyses based on these RCTs are of questionable interpretability and value. Herein, we present an individual patient data meta-analysis of RCTs of acupuncture for chronic pain, in which only high-quality RCTs were eligible for inclusion. Individual patient data meta-analysis are superior to the use of summary data in meta-analysis because they enhance data quality, enable different forms of outcome to be combined, and allow use of statistical techniques of increased precision.

See Invited Commentary at end of article

Author Affiliations are listed at the end of this article.

Group Information: The members of the Acupuncture Trialists’ Collaboration are listed at the end of this article.
METHODS

The full protocol of the meta-analysis has been published. In brief, the study was conducted in 3 phases: identification of eligible RCTs; collection, checking, and harmonization of raw data; and individual patient data meta-analysis.

DATA SOURCES AND SEARCHES

To identify articles, we searched MEDLINE, the Cochrane Collaboration Central Register of Controlled Trials, and the citation lists of systematic reviews (the full search strategy is shown in the Appendix; http://www.archinternmed.com). There were no language restrictions. The initial search, current to November 2008, was used to identify studies for the individual patient data meta-analysis; a second search was conducted in December 2010 for summary data to use in a sensitivity analysis.

STUDY SELECTION

Two reviewers applied inclusion criteria for potentially eligible articles separately, with disagreements about study inclusion resolved by consensus. Randomized controlled trials were eligible for analysis if they included at least 1 group receiving acupuncture needling and 1 group receiving either sham (placebo) acupuncture or no-acupuncture control. The RCTs must have accrued patients with 1 of 4 indications—nonspecific back or neck pain, shoulder pain, chronic headache, or osteoarthritis—with the additional criterion that the current episode of pain must be of at least 4 weeks duration for musculoskeletal disorders. There was no restriction on the type of outcome measure, although we specified that the primary end point must be measured more than 4 weeks after the initial acupuncture treatment.

It has been demonstrated that uncontrolled allocation is the most important source of bias in RCTs, and, as such, we included only those RCTs in which allocation concealment was determined unambiguously to be adequate (further details are in the review protocol). Where necessary, we contacted authors for further information concerning the exact logistics of the randomization process. We excluded RCTs if there was any ambiguity about allocation concealment.

DATA EXTRACTION AND QUALITY ASSESSMENT

The principal investigators of eligible studies were contacted and asked to provide raw data from the RCT. To ensure data accuracy, all results reported in the RCT publication, including baseline characteristics and outcome data, were then replicated.

Reviewers assessed the quality of blinding for eligible RCTs with sham acupuncture control. The RCTs were graded as having a low likelihood of bias if either the adequacy of blinding was checked by direct questioning of patients (eg, by use of a credibility questionnaire) and no important differences were found between groups, or the blinding method (eg, the Streitberger and Kleinhenz sham device) had previously been validated as able to maintain blinding. Randomized controlled trials with a high likelihood of bias from unblinding were excluded from the meta-analysis of acupuncture vs sham; a sensitivity analysis included only RCTs with a low risk of bias.

DATA SYNTHESIS AND ANALYSIS

Each RCT was reanalyzed by analysis of covariance with the standardized principal end point (scores divided by pooled standard deviation) as the dependent variable, and the baseline measure of the principal end point and variables used to stratify randomization as covariates. This approach has been shown to have the greatest statistical power for RCTs with baseline and follow-up measures. The effect size for acupuncture from each RCT was then entered into a meta-analysis using the metan command in Stata software (version 11; Stata Corp): the meta-analytic statistics were created by weighting each coefficient by the reciprocal of the variance, summing, and dividing by the sum of the weights. Meta-analyses were conducted separately for comparisons of acupuncture with sham and no-acupuncture control, and within each pain type. We prespecified that the hypothesis test would be based on the fixed effects analysis because this constitutes a valid test of the null hypothesis of no treatment effect.

SYSTEMATIC REVIEW

We identified 82 RCTs (Figure 1), of which 31 were eligible (Table 1 and Appendix). Four of the studies were organized as part of the German Acupuncture Trials (GERAC) initiative. Four were part of the Acupuncture Randomized Trials (ART) group; 4 were Acupuncture in Routine Care (ARC) studies; and 3 were UK National Health Service acupuncture RCTs. Eleven studies were sham controlled, 10 had no-acupuncture control, and 10 were 3-armed studies, including both sham and no-acupuncture control. The second search for subsequently published studies identified an additional 4 eligible studies, with a total of 1619 patients.

An important source of clinical heterogeneity between studies concerns the control groups. In the sham RCTs, the type of sham included acupuncture needles inserted superficially, sham acupuncture devices with needles that retract into the handle rather than penetrate the skin, and nonneedle approaches,
such as deactivated electrical stimulation\textsuperscript{20} or detused laser.\textsuperscript{27} Moreover, cointerventions varied, with no additional treatment other than analgescics in some RCTs,\textsuperscript{15} whereas other RCTs, both acupuncture and sham groups received a course of additional treatment, such as exercise led by physical therapists.\textsuperscript{24} Similarly, the no-acupuncture control groups varied among usual care, such as an RCT in which control group patients were merely advised to “avoid acupuncture”;\textsuperscript{198} attention control, such as group education sessions;\textsuperscript{28} and guided care, in which patients were given advice as to specific drugs and doses.\textsuperscript{13}

### DATA EXTRACTION AND QUALITY ASSESSMENT

Usable raw data were obtained from 29 of the 31 eligible RCTs, including a total of 17 922 patients from the United States, United Kingdom, Germany, Spain and Sweden. For 1 RCT, the study database had become corrupted;\textsuperscript{29} in another case, the statisticians involved in the RCT failed to respond to repeated enquiries despite approval for data sharing being obtained from the principal investigator.\textsuperscript{30}

The 29 RCTs comprised 18 comparisons with 14 597 patients of acupuncture and no-acupuncture group and 20 comparisons with 5230 patients of acupuncture and sham acupuncture. Patients in all RCTs had access to analgesics and other standard treatments for pain. Four sham RCTs were determined to have an intermediate likelihood of bias from unblinding;\textsuperscript{31,32} the 16 remaining sham RCTs were graded as having a low risk of bias from unblinding. On average, dropout rates were low (weighted mean, 10%). Dropout rates were higher than 25% for only 4 RCTs: those by Molsberger et al\textsuperscript{20,97} (27% and 33%, respectively, but raw data were not received and neither RCT included in main analysis); Carlsson et al\textsuperscript{12} (46%, RCT excluded in a sensitivity analysis for blinding), and Berman et al\textsuperscript{28} (31%). This RCT had a high dropout rate among no-acupuncture controls (43%); dropout rates were close to 25% in the acupuncture and sham groups. The RCT by Kerr et al\textsuperscript{31} had a large difference in dropout rates between groups (acupuncture, 13%; control, 33%) but was excluded in the sensitivity analysis for blinding.

### META-ANALYSIS

Forest plots for acupuncture against sham acupuncture and against no-acupuncture control are shown separately for each of the 4 pain conditions in Figure 2 and Figure 3. Meta-analytic statistics are shown in Table 2. Acupuncture was statistically superior to control for all

### Table 1. Characteristics of Included Studies\textsuperscript{a}

<table>
<thead>
<tr>
<th>Indication</th>
<th>Pain Type</th>
<th>Control Group</th>
<th>Primary Outcome Measure</th>
<th>Time Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic headache (n=7)</td>
<td>Migraine (n=2)\textsuperscript{15,18}; tension-type headache (n=3)\textsuperscript{14,18,19}; both (n=2)\textsuperscript{21,33}</td>
<td>Sham control (n=4)\textsuperscript{15,14,18,19}; no-acupuncture control (n=6); ancillary care (n=1)\textsuperscript{22}; usual care (n=4)\textsuperscript{16,18,23,36}; non-specific advice (n=1)\textsuperscript{23}</td>
<td>Constant-Murley score (n=2)\textsuperscript{25,41}; VAS (n=2)\textsuperscript{40,97}; Roland-Morris Disability Questionnaire (n=2)\textsuperscript{36}; WOMAC (n=2)\textsuperscript{12,22}; WOMAC (n=1)\textsuperscript{21,24,38,95,96}; Hannover Functional Neck Pain and Disability Questionnaire (n=1)\textsuperscript{35}; von Korff pain score (n=1)\textsuperscript{22}; SF-36 Bodily pain (n=1)\textsuperscript{23}</td>
<td>1 mo (n=1)\textsuperscript{22}; 3 mo (n=2)\textsuperscript{7,13}; 6 mo (n=2)\textsuperscript{12,14}; 12 mo (n=1)\textsuperscript{22}</td>
</tr>
<tr>
<td>Nonspecific musculoskeletal pain (back and neck) (n=15)</td>
<td>Back (n=10)\textsuperscript{6}; neck (n=5)\textsuperscript{20,26,27,36,37}</td>
<td>Sham control (n=10)\textsuperscript{6}</td>
<td>VAS (n=7)\textsuperscript{15,27,30-32,37}; Roland-Morris Disability Questionnaire (n=2)\textsuperscript{15,27,30-32,37}; Neck Pain and Disability Questionnaire (n=1)\textsuperscript{35}; Hannover Functional Pain subscore (n=1)\textsuperscript{26}; Northwick Park Neck Pain Questionnaire (n=1)\textsuperscript{94}; Von Korff pain score (n=1)\textsuperscript{22}; SF-36 Bodily pain (n=1)\textsuperscript{23}</td>
<td>1 mo (n=4)\textsuperscript{27,31,37}; 2 mo (n=3)\textsuperscript{34,35}; 3 mo (n=3)\textsuperscript{28,34,35,36}; 6 mo (n=2)\textsuperscript{21,32}; 24 mo (n=1)\textsuperscript{32}</td>
</tr>
<tr>
<td>Osteoarthritis (n=9)</td>
<td></td>
<td>Sham control (n=8)\textsuperscript{11,17,24,28,30,35}</td>
<td>Oxford Knee score questionnaire (n=1)\textsuperscript{75}; WOMAC (n=2)\textsuperscript{12,22}; WOMAC Pain subscore (n=6)\textsuperscript{11,24,28,30,35,36}</td>
<td>2 mo (n=2)\textsuperscript{17,38}; 3 mo (n=3)\textsuperscript{22,38,39,96}; 6 mo (n=3)\textsuperscript{12,24,28}</td>
</tr>
<tr>
<td>Shoulder pain (n=4)</td>
<td></td>
<td>Sham control (n=4)\textsuperscript{20,40,41,97}; no-acupuncture control (n=1); usual care (n=1)\textsuperscript{25,41}</td>
<td>Constant-Murley score (n=2)\textsuperscript{25,41}; VAS (n=2)\textsuperscript{97}</td>
<td>1 mo (n=2)\textsuperscript{25,41}; 6 mo (n=2)\textsuperscript{25,41}</td>
</tr>
</tbody>
</table>

Abbreviations: SF-36, 36-Item Short Form questionnaire; VAS, visual analog scale; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

\textsuperscript{a}Trial level information is provided in the eAppendix (http://www.archinternmed.com). The table includes the 31 trials identified in the initial search plus the 4 recently identified trials for which summary data were used.

\textsuperscript{b}Ancillary care: Program of care received by both acupuncture and no-acupuncture control groups (eg, trial comparing physiotherapy plus acupuncture with physiotherapy alone).

\textsuperscript{c}Usual care: Protocol did not specify treatments received in control group (eg, trials with “waiting list control”).

\textsuperscript{d}Guideline care: Patients in the control group received care according to national guidelines.

\textsuperscript{e}Abstracted from references 12, 15, 19, 23, 30-32, 34, 35, and 94.

\textsuperscript{f}Abstracted from references 12, 15, 26, 27, 30-32, 34, 37, and 94.

\textsuperscript{g}Non-specific advice: Patients in control group receive general advice and support (“attention control”).
analyses \((P < .001)\). Effect sizes are larger for the comparison between acupuncture and no-acupuncture control than for the comparison between acupuncture and sham: 0.37, 0.26, and 0.15 in comparison with sham vs. 0.55, 0.57, and 0.42 in comparison with no-acupuncture control for musculoskeletal pain, osteoarthritis, and chronic headache, respectively.

For 5 of the 7 analyses, the test for heterogeneity was statistically significant. In the case of comparisons with sham acupuncture, the RCTs by Vas et al\(^{37,38,41}\) are clear outliers. For example, the effect size of the RCTs by Vas et al for neck pain is about 5 times greater than meta-analytic estimate. One effect of excluding these RCTs in a sensitivity analysis (Table 3 and Table 4) is that there is no significant heterogeneity in the comparisons between acupuncture and sham. Moreover, the effect size for acupuncture becomes relatively similar for the different pain conditions: 0.23, 0.16, and 0.15 against sham, and 0.55, 0.57, and 0.42 against no-acupuncture control for back and neck pain, osteoarthritis, and chronic headache, respectively (fixed effects; results similar for the random effects analysis).

To give an example of what these effect sizes mean in real terms, a baseline pain score on a 0 to 100 scale for a typical RCT might be 60. Given a standard deviation of 25, follow-up scores might be 43 in a no-acupuncture group, 35 in a sham acupuncture group, and 30 in patients receiving true acupuncture. If response were defined in terms of a pain reduction of 50% or more, response rates would be approximately 30%, 42.5%, and 50%, respectively.

The comparisons with no-acupuncture control show evidence of heterogeneity. This seems largely explicable in terms of differences between the control groups used. In the case of osteoarthritis, the largest effect was in the study by Witt et al,\(^{17}\) in which patients in the waiting list control received only rescue pain medication, and the smallest was in the study by Foster et al,\(^{24}\) which involved a program of exercise and advice led by physical therapists. For the musculoskeletal analyses, heterogeneity is driven by 2 very large RCTs\(^{19,20}\) (n = 2565 patients and n = 3118 patients, respectively) for back and neck pain. If only back pain is considered (Table 3 and Table 4), heterogeneity is dramatically reduced and is again driven by one RCT, by Brinkhaus et al,\(^{15}\) with waiting list control. In the headache meta-analysis, Diener et al\(^{13}\) had much smaller differences between groups. This RCT involved providing drug therapy according to national guidelines in the no-acupuncture group, including initiation of β-blockers as migraine prophylaxis. There was disagreement within the collaboration about whether this constituted active control. Excluding this RCT reduced evidence of heterogeneity \((P = .04)\) but had little effect on the effect size \((0.42-0.45)\).

The Table 3 and Table 4 show several prespecified sensitivity analyses. Neither restricting the sham RCTs to those with low likelihood of unblinding nor adjustment for missing data had any substantive effect on our main estimates. Inclusion of summary data from RCTs for which raw data were not obtained (2 RCTs) or which were published recently (4 RCTs) also had little impact on either the primary analysis (Table 3 and Table 4) or the analysis with the outlying RCTs by Vas et al\(^{37,38,41}\) excluded (data not shown).

To estimate the potential impact of publication bias, we entered all RCTs into a single analysis and com-
RCTs by Vas et al,37,38,41 which were influenced by the outlying control (\(P = .02\)) but not no-acupuncture control (\(P = .72\)). However, these analyses are influenced by the outlying RCTs by Vas et al,37,38,41 which were smaller than average, and by indication, because the shoulder pain RCTs were small and had large effect sizes. Tests for asymmetry were nonsignificant when we excluded the RCTs by Vas et al,37,38,41 and shoulder pain studies (\(n = 15; \ P = .07\)) and when small studies were also excluded (\(n < 100\) and \(n = 12\), respectively; \(P = .30\)). Nonetheless, we repeated our meta-analyses excluding RCTs with a sample size of less than 100. This had essentially no effect on our results. As a further test of publication bias, we considered the possible effect on our analyses if we had failed to include high-quality, unpublished studies. Only if there were 47 unpublished RCTs with \(n = 100\) patients showing an advantage to sham of 0.25 SD would the difference between acupuncture and sham lose significance.

A final sensitivity analysis examined the effect of pooling different end points measured at different periods of follow-up. We repeated our analyses including only pain end points measured at 2 to 3 months after randomization. There was no material effect on results: effect sizes increased by 0.05 to 0.09 SD for musculoskeletal and osteoarthritis RCTs and were stable otherwise.

As an exploratory analysis, we compared sham control with no-acupuncture control. In a meta-analysis of 9 RCTs,11-13,15-18,24,28 the effect size for sham was 0.33 (95% CI, 0.27-0.40) and 0.38 (95% CI, 0.20-0.56) for fixed and random effects models, respectively (\(P < .001\) for tests of both effect and heterogeneity).

Figure 3. Forest plots for the comparison of true and sham acupuncture. A, Osteoarthritis; B, chronic headache; C, musculoskeletal pain; D, shoulder pain.

In an analysis of patient-level data from 29 high-quality RCTs, including 17 922 patients, we found statistically significant differences between both acupuncture vs sham and acupuncture vs no-acupuncture control for all pain types studied. After excluding an outlying set of studies, meta-analytic effect sizes were similar across pain conditions.

The effect size for individual RCTs comparing acupuncture with no-acupuncture control did vary, an effect that seems at least partly explainable in terms of the type of control used. As might be expected, acupuncture had a smaller benefit in patients who received a program of ancillary care—such as physical therapist–led exercise33—than in patients who continued to be treated with usual care. Nonetheless, the average effect, as expressed in the meta-analytic estimate of approximately 0.5 SD, is of clear clinical relevance whether considered either as a standardized difference100 or when converted back to a pain scale. The difference between acupuncture and sham is of lesser magnitude, 0.15 to 0.23 SD.

**COMMENT**

In an analysis of patient-level data from 29 high-quality RCTs, including 17 922 patients, we found statistically significant differences between both acupuncture vs sham and acupuncture vs no-acupuncture control for all pain types studied. After excluding an outlying set of studies, meta-analytic effect sizes were similar across pain conditions.

The effect size for individual RCTs comparing acupuncture with no-acupuncture control did vary, an effect that seems at least partly explainable in terms of the type of control used. As might be expected, acupuncture had a smaller benefit in patients who received a program of ancillary care—such as physical therapist–led exercise33—than in patients who continued to be treated with usual care. Nonetheless, the average effect, as expressed in the meta-analytic estimate of approximately 0.5 SD, is of clear clinical relevance whether considered either as a standardized difference100 or when converted back to a pain scale. The difference between acupuncture and sham is of lesser magnitude, 0.15 to 0.23 SD.

**OVERVIEW OF FINDINGS**

We saw some evidence that small studies had larger effect sizes for the comparison with sham (\(P = .02\)) but not no-acupuncture control (\(P = .72\)). However, these analyses are influenced by the outlying RCTs by Vas et al,37,38,41 which were smaller than average, and by indication, because the shoulder pain RCTs were small and had large effect sizes. Tests for asymmetry were nonsignificant when we excluded the RCTs by Vas et al,37,38,41 and shoulder pain studies (\(n = 15; \ P = .07\)) and when small studies were also excluded (\(n < 100\) and \(n = 12\), respectively; \(P = .30\)). Nonetheless, we repeated our meta-analyses excluding RCTs with a sample size of less than 100. This had essentially no effect on our results. As a further test of publication bias, we considered the possible effect on our analyses if we had failed to include high-quality, unpublished studies. Only if there were 47 unpublished RCTs with \(n = 100\) patients showing an advantage to sham of 0.25 SD would the difference between acupuncture and sham lose significance.

A final sensitivity analysis examined the effect of pooling different end points measured at different periods of follow-up. We repeated our analyses including only pain end points measured at 2 to 3 months after randomization. There was no material effect on results: effect sizes increased by 0.05 to 0.09 SD for musculoskeletal and osteoarthritis RCTs and were stable otherwise.

As an exploratory analysis, we compared sham control with no-acupuncture control. In a meta-analysis of 9 RCTs,11-13,15-18,24,28 the effect size for sham was 0.33 (95% CI, 0.27-0.40) and 0.38 (95% CI, 0.20-0.56) for fixed and random effects models, respectively (\(P < .001\) for tests of both effect and heterogeneity).

Figure 3. Forest plots for the comparison of true and sham acupuncture. A, Osteoarthritis; B, chronic headache; C, musculoskeletal pain; D, shoulder pain.

In an analysis of patient-level data from 29 high-quality RCTs, including 17 922 patients, we found statistically significant differences between both acupuncture vs sham and acupuncture vs no-acupuncture control for all pain types studied. After excluding an outlying set of studies, meta-analytic effect sizes were similar across pain conditions.

The effect size for individual RCTs comparing acupuncture with no-acupuncture control did vary, an effect that seems at least partly explainable in terms of the type of control used. As might be expected, acupuncture had a smaller benefit in patients who received a program of ancillary care—such as physical therapist–led exercise33—than in patients who continued to be treated with usual care. Nonetheless, the average effect, as expressed in the meta-analytic estimate of approximately 0.5 SD, is of clear clinical relevance whether considered either as a standardized difference100 or when converted back to a pain scale. The difference between acupuncture and sham is of lesser magnitude, 0.15 to 0.23 SD.

**OVERVIEW OF FINDINGS**

In an analysis of patient-level data from 29 high-quality RCTs, including 17 922 patients, we found statistically significant differences between both acupuncture vs sham and acupuncture vs no-acupuncture control for all pain types studied. After excluding an outlying set of studies, meta-analytic effect sizes were similar across pain conditions.

The effect size for individual RCTs comparing acupuncture with no-acupuncture control did vary, an effect that seems at least partly explainable in terms of the type of control used. As might be expected, acupuncture had a smaller benefit in patients who received a program of ancillary care—such as physical therapist–led exercise33—than in patients who continued to be treated with usual care. Nonetheless, the average effect, as expressed in the meta-analytic estimate of approximately 0.5 SD, is of clear clinical relevance whether considered either as a standardized difference100 or when converted back to a pain scale. The difference between acupuncture and sham is of lesser magnitude, 0.15 to 0.23 SD.
Table 2. Primary Analyses

<table>
<thead>
<tr>
<th>Indication</th>
<th>Studies, No.</th>
<th>Acupuncture vs Sham Acupuncture</th>
<th>Acupuncture vs No-Acupuncture Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FE (95% CI)</td>
<td>Heterogeneity: P Value</td>
</tr>
<tr>
<td>Nonspecific musculoskeletal pain (back and neck)</td>
<td>8</td>
<td>0.37 (0.27-0.46)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>5</td>
<td>0.26 (0.17-0.34)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic headache</td>
<td>4</td>
<td>0.15 (0.07-0.24)</td>
<td>.31</td>
</tr>
<tr>
<td>Shoulder pain</td>
<td>3</td>
<td>0.62 (0.46-0.77)</td>
<td>.44</td>
</tr>
</tbody>
</table>

Abbreviations: FE, fixed effects; RE, random effects.

Effect sizes are standardized differences; \( P < .001 \) for the overall effects for all comparisons.

Table 3. Sensitivity Analyses: Acupuncture vs Sham Acupuncture

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Indication</th>
<th>Studies, No.</th>
<th>Acupuncture vs Sham Acupuncture</th>
<th>Acupuncture vs No-Acupuncture Control</th>
<th>P Value for OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusion of VAS trials b</td>
<td>Nonspecific musculoskeletal pain</td>
<td>7</td>
<td>0.23 (0.13-0.33)</td>
<td>.51</td>
<td>0.23 (0.13-0.33)</td>
</tr>
<tr>
<td>Separate pain types</td>
<td>Osteoarthritis</td>
<td>4</td>
<td>0.16 (0.07 to 0.25)</td>
<td>.15</td>
<td>0.17 (0.00 to 0.35)</td>
</tr>
<tr>
<td></td>
<td>Back pain</td>
<td>5</td>
<td>0.20 (0.09 to 0.31)</td>
<td>.40</td>
<td>0.20 (0.09 to 0.32)</td>
</tr>
<tr>
<td></td>
<td>Neck pain</td>
<td>3</td>
<td>0.83 (0.64 to 1.01)</td>
<td>&lt;.001</td>
<td>0.82 (−0.11 to 1.75)</td>
</tr>
<tr>
<td>Inclusion of trials for which raw data not obtained c</td>
<td>Nonspecific musculoskeletal pain</td>
<td>10</td>
<td>0.30 (0.21 to 0.38)</td>
<td>&lt;.001</td>
<td>0.48 (0.14 to 0.81)</td>
</tr>
<tr>
<td></td>
<td>Osteoarthritis</td>
<td>6</td>
<td>0.22 (0.14 to 0.30)</td>
<td>&lt;.001</td>
<td>0.31 (0.02 to 0.60)</td>
</tr>
<tr>
<td></td>
<td>Shoulder pain</td>
<td>4</td>
<td>0.57 (0.44 to 0.69)</td>
<td>.41</td>
<td>0.57 (0.44 to 0.69)</td>
</tr>
<tr>
<td>Only trials with low likelihood of bias for blinding d</td>
<td>Non-specific musculoskeletal pain</td>
<td>5</td>
<td>0.36 (0.25 to 0.46)</td>
<td>&lt;.001</td>
<td>0.57 (0.00 to 1.14)</td>
</tr>
<tr>
<td></td>
<td>Chronic headache</td>
<td>3</td>
<td>0.14 (0.03 to 0.25)</td>
<td>.18</td>
<td>0.12 (−0.02 to 0.27)</td>
</tr>
<tr>
<td>Multiple imputation for missing data</td>
<td>Nonspecific musculoskeletal pain</td>
<td>8</td>
<td>0.36 (0.27 to 0.46)</td>
<td>&lt;.001</td>
<td>0.52 (0.15 to 0.88)</td>
</tr>
<tr>
<td></td>
<td>Osteoarthritis</td>
<td>5</td>
<td>0.25 (0.16 to 0.33)</td>
<td>&lt;.001</td>
<td>0.37 (0.03 to 0.71)</td>
</tr>
<tr>
<td></td>
<td>Chronic headache</td>
<td>4</td>
<td>0.16 (0.07 to 0.25)</td>
<td>.38</td>
<td>0.16 (0.07 to 0.25)</td>
</tr>
<tr>
<td></td>
<td>Shoulder pain</td>
<td>3</td>
<td>0.62 (0.46 to 0.78)</td>
<td>.46</td>
<td>0.62 (0.46 to 0.78)</td>
</tr>
</tbody>
</table>

Abbreviations: FE, fixed effects; OE, overall effects; RE, random effect.

Effect sizes are standardized differences.

b Fewer than 3 trials for shoulder pain.

c No trials for chronic headache.

d For osteoarthritis and shoulder pain, as for Table 2: all trials have a low likelihood of bias for blinding.

Table 4. Sensitivity Analyses: Acupuncture vs No-Acupuncture Control

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Indication</th>
<th>Studies, No.</th>
<th>Acupuncture vs No-Acupuncture Control</th>
<th>P Value for OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate pain types b</td>
<td>Back pain</td>
<td>5</td>
<td>0.46 (0.40-0.51)</td>
<td>.004</td>
</tr>
<tr>
<td>Inclusion of trials for which raw data not obtained c</td>
<td>Nonspecific musculoskeletal pain</td>
<td>9</td>
<td>0.55 (0.51-0.58)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Osteoarthritis</td>
<td>8</td>
<td>0.58 (0.51-0.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Chronic headache</td>
<td>6</td>
<td>0.42 (0.38-0.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Multiple imputation for missing data</td>
<td>Nonspecific musculoskeletal pain</td>
<td>7</td>
<td>0.55 (0.51-0.58)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Osteoarthritis</td>
<td>6</td>
<td>0.57 (0.50-0.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Chronic headache</td>
<td>5</td>
<td>0.42 (0.38-0.46)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviations: FE, fixed effects; OE, overall effects; RE, random effect.

Effect sizes are standardized differences.

b No trials for neck pain.

c Fewer than 3 trials for shoulder pain.

d No trials for shoulder pain.
STUDY LIMITATIONS

Neither study quality nor sample size seems to be a problem for this meta-analysis, on the grounds that only high-quality studies were eligible and the total sample size is large. Moreover, we saw no evidence that publication bias, or failure to identify published eligible studies, could affect our conclusions.

Because the comparisons between acupuncture and no-acupuncture cannot be blinded, both performance and response bias are possible. Similarly, while we considered the risk of bias of unblinding low in most studies comparing acupuncture and sham acupuncture, health care providers obviously were aware of the treatment provided, and, as such, a certain degree of bias of our effect estimate for specific effects cannot be entirely ruled out. However, it should be kept in mind that this problem applies to almost all studies on non-drug interventions. We would argue that the risk of bias in the comparison between acupuncture and sham acupuncture is low compared with other non-drug treatments for chronic pain, such as cognitive therapies, exercise, or manipulation, which are rarely subject to placebo control.

Another possible critique is that the meta-analyses combined different end points, such as pain and function, measured at different times. However, results did not change when we restricted the analysis to pain end points measured at a specific follow-up time, 2 to 3 months after randomization.

COMPARISON WITH OTHER STUDIES

Many prior systematic reviews of acupuncture for chronic pain have had liberal eligibility criteria, accordingly included RCTs of low methodologic quality, and then came to the circular conclusion that weaknesses in the data did not allow conclusions to be drawn. Other reviews have not included meta-analyses, apparently owing to variation in study end points. We have avoided both problems by including only high-quality RCTs and obtaining raw data for individual patient data meta-analysis. Some more recent systematic reviews have published meta-analyses and reported findings that are broadly comparable with ours, with clear differences between acupuncture and no-acupuncture control and smaller differences between true and sham acupuncture. Our findings have greater precision: all prior reviews have analyzed summary data, an approach of reduced statistical precision when compared with individual patient data meta-analysis. In particular, we have demonstrated a robust difference between acupuncture and sham control that can be distinguished from bias. This is a novel finding that moves beyond the prior literature.

INTERPRETATION

We believe that our findings are both clinically and scientifically important. They suggest that the total effects of acupuncture, as experienced by the patient in routine clinical practice, are clinically relevant, but that an important part of these total effects is not due to issues considered to be crucial by most acupuncturists, such as the correct location of points and depth of needling. Several lines of argument suggest that acupuncture (whether real or sham) is associated with more potent placebo or context effects than other interventions. Yet, many clinicians would feel uncomfortable in providing or referring patients to acupuncture if it were merely a potent placebo. Similarly, it is questionable whether national or private health insurance should reimburse therapies that do not have specific effects. Our finding that acupuncture has effects over and above those of sham acupuncture is therefore of major importance for clinical practice. Even though on average these effects are small, the clinical decision made by physicians and patients is not between true and sham acupuncture but between a referral to an acupuncturist or avoiding such a referral. The total effects of acupuncture, as experienced by the patient in routine practice, include both the specific effects associated with correct needle insertion according to acupuncture theory, nonspecific physiologic effects of needling, and nonspecific psychological (placebo) effects related to the patient's belief that treatment will be effective.

In conclusion, we found acupuncture to be superior to both no-acupuncture control and sham acupuncture for the treatment of chronic pain. Although the data indicate that acupuncture is more than a placebo, the differences between true and sham acupuncture are relatively modest, suggesting that factors in addition to the specific effects of needling are important contributors to therapeutic effects. Our results from individual patient data meta-analyses of nearly 18,000 randomized patients in high-quality RCTs provide the most robust evidence to date that acupuncture is a reasonable referral option for patients with chronic pain.

Accepted for Publication: May 28, 2012.
Published Online: September 10, 2012. doi:10.1001/archinternmed.2012.3654
Author Affiliations: Departments of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York (Dr Vickers and Ms Maschino); Center for Outcomes and Policy Research Dana Farber Cancer Institute, Boston, Massachusetts (Ms Cronin); Complementary and Integrated Medicine Research Unit, University of Southampton, Southampton, England (Dr Lewith); Department of Health Sciences, University of York, York, England (Dr MacPherson); Arthritis Research UK Primary Care Centre, Keele University, Newcastle-under-Lyme, Staffordshire, England (Dr Foster); Group Health Research Institute, Seattle, Washington (Dr Sherman); Institute for Social Medicine, Epidemiology, and Health Economics, Berlin, Germany (Dr Witt); and Institute of General Practice, Technische Universität München, Munich, Germany (Dr Linde).

Correspondence: Andrew J. Vickers, DPhil, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 E 63rd St, New York, NY 10065 (vickersa@mskcc.org).

Author Contributions: Dr Vickers had full access to all of the data in...
the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors gave comments on early drafts and approved the final version of the manuscript. Study concept and design: Vickers, Lewith, Foster, Witt, and Linde. Acquisition of data: Vickers, Maschino, MacPherson, Foster, and Witt. Analysis and interpretation of data: Vickers, Cronin, Maschino, Lewith, MacPherson, Foster, Sherman, Witt, and Linde. Drafting of the manuscript: Vickers and Maschino. Critical revision of the manuscript for important intellectual content: Vickers, Cronin, Maschino, Lewith, MacPherson, Foster, Sherman, Witt, and Linde. Statistical analysis: Vickers, Cronin, and Maschino. Obtained funding: Vickers and Linde. Administrative, technical, and material support: Lewith. Study supervision: Vickers.

Financial Disclosure: None reported.

Funding/Support: The Acupuncture Trialists’ Collaboration is funded by an R21 (AT004189) from the National Center for Complementary and Alternative Medicine (NCCAM) at the National Institutes of Health (NIH) to Dr Vickers and by a grant from the SamueI Institute. Dr MacPherson’s work has been supported in part by the UK National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10186). The views expressed in this publication are those of the author(s) and not necessarily those of the NCCAM NHS, the NIHR, or the Department of Health in England.

Role of the Sponsors: No sponsor had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

The Acupuncture Trialists’ Collaboration: Claire Allen, BS, Cochrane Collaboration Secretariat, Oxford, England; Mac Beckner, MIS, Information Technology and Data Management Center, SamueI Institute, Alexandria, Virginia; Brian Berman, MD, University of Maryland School of Medicine and Center for Integrative Medicine, College Park; Benno Brinkhaus, MD, Institute for Social Medicine, Epidemiology and Health Economics, Charité–University Medical Center, Berlin, Germany; Remy Coeytaux, MD, PhD, Department of Community and Family Medicine, Duke University, Durham, North Carolina; Angel M. Cronin, MS, Dana-Farber Cancer Institute, Boston, Massachusetts; Hans-Christian Diener, MD, PhD, Department of Neurology, University of Duisburg-Essen, Germany; Heinz G. Endres, MD, Ruhr–University Bochum, Bochum, Germany; Nadine Foster, DPhil, BSc(Hons), Arthritis Research UK Primary Care Centre, Keele University, Newcastle-under-Lyme, Staffordshire, England; Juan Antonio Guerra de Hoyos, MD, Andalusian Integral Plan for Pain Management, and Andalusian Health Service Project for Improving Primary Care Research, Sevilla, Spain; Michael Haake, MD, PhD, Department of Orthopedics and Traumatology, SLK-Hospitals, Heilbronn, Germany; Richard Hammerschlag, PhD, Oregon College of Oriental Medicine, Portland; Dominik Irmisch, MD, Interdisciplinary Pain Centre, University of Munich, Munich, Germany; Wayne B. Jonas, MD, SamueI Institute; Kai Kromfeld, PhD, Interdisciplinary Centre for Clinical Trials (IZKS Mainz), University Medical Centre Mainz, Mainz, Germany; Lixing Lao, PhD, University of Maryland and Center for Integrative Medicine, College Park; George Lewith, MD, FRCP, Complementary and Integrated Medicine Research Unit, Southampton Medical School, Southampton, England; Klaus Linde, MD, Institute of General Practice, Technische Universität München, Munich; Hugh MacPherson, PhD, Complementary Medicine Research Group, University of York, York, England; Eric Manheimer, MS, Center for Integrative Medicine, University of Maryland School of Medicine, College Park; Alexandra Maschino, BS, Memorial Sloan-Kettering Cancer Center, New York, New York; Dietmar Melchart, MD, PhD, Centre for Complementary Medicine Research (ZfK), Technische Universität München, Munich; Albrecht Molsberger, MD, PhD, German Acupuncture Research Group, Duesseldorf, Germany; Karen J. Sherman, PhD, MPH, Group Health Research Institute, Seattle, Washington; Hans Trampisch, PhD, Department of Medical Statistics and Epidemiology, Ruhr–University Bochum; Jorge Vas, MD, PhD, Pain Treatment Unit, Dos Hermanns Primary Care Health Center (Andalusia Public Health System), Dos Hermanns, Spain; Andrew J. Vickers (collaboration chair), DPhil, Memorial Sloan-Kettering Cancer Center; Norbert Victor, PhD (deceased), Institute of Medical Biometrics and Informatics, University of Heidelberg, Heidelberg, Germany; Peter White, PhD, School of Health Sciences, University of Southampton; Lyn Williamson, MD, MA (Oxon), MRCGP, FRCP, Great Western Hospital, Swindon, and Oxford University, Oxford, England; Stefan N. Willich, MD, MPH, MBA, Institute for Social Medicine, Epidemiology, and Health Economics, Charité University Medical Center, Berlin; Claudia M. Witt, MD, MBA, University Medical Center Charité and Institute for Social Medicine, Epidemiology and Health Economics, Berlin.


REFERENCES


