Appropriateness of sham or placebo acupuncture for randomized controlled trials of acupuncture for nonspecific low back pain: a systematic review and meta-analysis

Yan Xiang1
Jinyuan He2
Rui Li1

1Department of Teaching, School of Acupuncture and Moxibustion and Tuina, Beijing University of Chinese Medicine, Beijing, China; 2Department of Cardiothoracic Surgery, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Objectives: To establish whether sham acupuncture (SA) or placebo acupuncture (PA) is more efficacious for reducing low back pain (LBP) than other routine treatments and to discuss whether SA or PA is appropriate for randomized controlled trials of acupuncture for LBP.

Methods: Six databases were searched on 31 May 2017. We included only randomized controlled trials of adults with LBP and lower back myofascial pain syndrome. The studies had at least two control arms: a sham-controlled acupuncture arm and a routine care or waiting list arm (people who did not receive acupuncture until the end of treatment). Trials were combined using meta-analysis methods when the data allowed statistical pooling. Pooled effect sizes were calculated by random effects models.

Results: This review identified 7 trials (1768 participants); all were included in the meta-analysis. We found statistically significant differences in pain reduction post-intervention between SA or PA and routine care or a waiting list, with a standardized mean difference of −0.36 (95% CI −0.54 to −0.18; I² statistic=16%; participants=624; studies=6) for the Visual Analog Scale and −0.35 (95% CI −0.49 to −0.20; I² statistic=0%; participants=736; studies=1) for the Chronic Pain Grade Scale; however, no significant difference was observed between SA or PA and routine care or no treatment for post-intervention function.

Conclusion: Compared with routine care or a waiting list, SA or PA was more efficacious for pain relief post-intervention. Concluding that SA or PA is appropriate for acupuncture research would be premature. Guidelines evaluating SA or PA control methods are needed to determine the specific effect of acupuncture over placebo.

Keywords: acupuncture, alternative medicine, backaches, pain management, placebos, controlled clinical trial, lumbago

Introduction

Nonspecific low back pain (LBP) is among the most common health complaints; it is the leading cause of years lived with disability worldwide and sixth in terms of overall disease burden (disability-adjusted life-years).1 Although LBP is usually a self-limiting and benign disease, a large variety of therapeutic interventions are available to treat it.2

Acupuncture, based on the concepts of traditional Chinese medicine (TCM), is one of the oldest and most popular complementary or alternative treatment methods. Although it is one of the most commonly used interventions to treat LBP, acupuncture
has mixed support. Despite a 2005 Cochrane review that reported the existence of some evidence supporting acupuncture for chronic LBP,\(^5\) a conclusion supported by another systematic review,\(^4\) the National Institute for Health and Care Excellence (NICE) in the UK removed acupuncture from their LBP guidelines in early 2016, citing a lack of evidence to indicate that it was more effective than sham treatment.\(^5\)

Sham acupuncture (SA), also called placebo acupuncture (PA), performed away from the acupuncture points established by TCM or without stimulation and manipulation to avoid eliciting “De Qi” sensations or using a non-penetrating technique, is used as control in scientific studies to determine the efficacy of acupuncture.\(^5,7\) However, in China, wrist-ankle acupuncture and abdomen acupuncture, both with shallow needling, are the primary treatments for LBP. It might be argued that the superficial or minimal acupuncture might stimulate C fibers in the skin to trigger some kind of analgesic effect.\(^8,9\) Of those sham-controlled clinical trials, several of them found no significant differences between real acupuncture and SA/PA in pain relief for LBP.\(^10–14\) As a result, controversy persists regarding whether acupuncture for LBP works primarily by the placebo effect. Meanwhile, some investigators argued that using SA/PA as a control is problematic because SA/PA might not meet the criteria of being physiologically inert.\(^15,16\)

Regardless, no systematic review and meta-analysis to date has addressed whether SA/PA has effects for LBP. We, therefore, performed a systematic review and meta-analysis of acupuncture for the treatment of LBP to determine whether SA/PA was more efficacious in reducing LBP than routine care and to discuss whether SA/PA is appropriate for use in randomized controlled trials (RCTs) of acupuncture for LBP.

**Methods**

**Criteria for considering studies for this review**

**Types of studies**

We included only RCTs, which provide the highest level of evidence to assess the effects of interventions. Moreover, the studies had at least two control arms: a sham-controlled acupuncture arm and a routine care or waiting list (people who did not receive acupuncture until the end of treatment) or no treatment arm. Articles published in English or Chinese without region restriction were included.

**Types of participants**

We included studies which recruited adults (>18 years) with LBP or myofascial pain syndrome in the lower back. RCTs that included subjects with LBP caused by specific pathological entities such as infection, metastatic disease, neoplasm, osteoarthritis, rheumatoid arthritis, inflammatory processes, radicular syndrome, or fractures were excluded. Patients with LBP associated with sciatica as the major symptom, pregnancy, post-partum status, and post-operative LBP were also excluded. Patients with acute, sub-acute (<12 weeks), or chronic LBP (>12 weeks) were included.

**Types of interventions**

We included studies in which acupuncture points were stimulated by needle insertion (with or without electroacupuncture), accompanied by a definite sensation of De Qi. The studies in which methods of stimulating acupuncture points by acupressure, transcutaneous electrical nerve stimulation, or infrared light for verum acupuncture were excluded. Bee venom acupuncture and ear acupuncture were excluded. Trials comparing two techniques of acupuncture were excluded. Trials assessing the efficacy of acupuncture as an adjunct treatment to other therapies compared to SA/PA as an adjunct treatment to other therapies were included.

**Types of outcome measures**

RCTs were included if they used at least one of the following two outcome measures considered salient in the field of LBP:

1. Pain intensity (e.g., Visual Analog Scale [VAS]);
2. Back-specific functional status (e.g., Roland–Morris Disability Questionnaire [RMDQ]).

The primary outcome was pain intensity. The secondary outcome measure was back-specific functional status.

**Search methods for identification of studies**

**Electronic searches**

We searched the following databases:

1. Cochrane Central Register of Controlled Trials (CENTRAL; the Cochrane Library, to Issue 5 of 12 May 2017) on 31 May 2017
3. EMBASE (Ovid SP, 1980 to May 2017) on 31 May 2017
4. China National Knowledge Infrastructure, the Wan Fang and the Wei Pu databases up to 31 May 2017

Keywords, free words, and MeSH terms including “acupuncture” OR “acupuncture therapy” OR “acupuncture points” AND “low back pain” OR “lower back pain” OR “backache” OR “lumbago” were used.
Appropriateness of sham or placebo acupuncture for RCTs

Searching other resources
Keywords of the references of all pertinent publications were traced, including systematic reviews and meta-analyses, to identify citations omitted by the electronic search.

Data collection and analysis
Selection of studies
For this review, one author generated the electronic search strategies in the CENTRAL, PubMed, EMBASE, and Chinese databases and downloaded the citations. Two authors independently applied the aforementioned inclusion criteria to identify trials. Any disagreements between them were resolved by consensus with a third review author acting as an arbiter. The authors of recent original studies were contacted to obtain more information when needed.

Data extraction and management
Two authors independently extracted data on the study characteristics, study population, duration of symptoms, treatment type, frequency and duration of the intervention, outcome measurements, and type of comparisons. A standard extraction form was used. Discrepancies were reassessed and consensus was reached by discussion. If necessary, a third author reviewed the data to reach consensus.

Assessment of risk of bias in included studies
Two authors independently assessed the risk of bias of each included trial using the criteria recommended in The Cochrane Handbook for Systematic Reviews of Interventions. Six domains of bias were assessed: selection bias, performance bias, detection (or measurement) bias, attrition bias, reporting bias, and other bias (registered or not, ethics approval obtained or not, participants gave the informed consent or not, etc.). For each study, we scored each criterion as “high risk”, “low risk”, or “unclear”. Any disagreements were resolved by consensus and a third author was consulted if disagreements persisted. We also used the risk of bias assessment of the included trials for grading the quality of the evidence.

Measures of treatment effect
The analyses compared SA/PA to no treatment or routine care. For continuous measures, the use of weighted mean differences was preferred to analyze the results when the outcome measures were identical. Standardized mean differences (SMD) were used when different instruments were used for the same outcome measurements.

Unit of analysis time frame
We extracted data from the outcomes measured immediately following the end of the sessions for up to 1 week after the end of the sessions.

Missing data management
When data were missing, attempts were made to obtain information by contacting the study authors.

Assessment of heterogeneity
Random effects models were used for all meta-analyses. This method is recommended by the CBN Group Editorial Board because the assumptions underlying the random effects model are better suited to the statistical combination of trials in this field.

Assessment of reporting biases
Publication bias was assessed using funnel plots generated using RevMan (version: 5.3.5 [Java 7 64 bit]) Analysis Software. We used the method of independent visual inspection by two review authors. A third review author was consulted in cases of disagreement.

Data synthesis
We entered all quantitative results into RevMan (version: 5.3.5 [Java 7 64 bit]) Analyses Software. A meta-analysis was considered when homogeneity existed in terms of population, comparison group, and outcome. The magnitude of the effect size was categorized as follows: 0.2 indicates a small effect, 0.5 a moderate effect, and 0.8 a large effect.

The GRADE approach was utilized to assess the quality of the evidence. This grading method for each outcome considered risk of bias, inconsistency of results, indirectness of evidence, imprecision, and publication bias.

Subgroup analysis and assessment of heterogeneity
Assessments not planned in this review were subgroup or meta-regression analyses and sensitivity analysis.

Results
Description of studies
The flow of studies is illustrated in Figure 1 study flow diagram. In this review, we identified seven trials for inclusion.
Included studies
In total, we included 7 trials (1768 participants). Four studies were conducted in Germany (1139 participants),10,13,24,25 one was conducted in the USA (323 participants),11 one was conducted in Spain (206 participants),14 and one was conducted in China (100 participants).26

Most patients included in these trials had chronic non-specific LBP. The control groups included conventional treatment, standard therapy, routine care, and waiting list.

The type of SA/PA varied among the included trials: four studies used a superficial insertion method at non-acupuncture points,10,13,24,25 one used a toothpick in a needle guide tube,11 and one penetrated at nonspecific acupuncture points that were then punctured following the usual procedure.14 The sham needles in the final study differed from regular needles in that they had blunt and retractable tips, although true acupuncture points were treated.26

Details of each included trial are presented in Table 1.

Risk of bias in included studies
Some degree of publication bias was suggested (Figure 2), which may be attributed in part to the small number of studies. A summary of the risk of bias for each article is shown in Figure 3.
Table 1: Study design and outcome measures

<table>
<thead>
<tr>
<th>Trials</th>
<th>Number of treatments/Frequency/Duration per treatment</th>
<th>Intervention group (N, Age, Sex)</th>
<th>Type of intervention</th>
<th>Control group (N, Age, Sex)</th>
<th>Type of comparison</th>
<th>Duration of pain</th>
<th>Duration of intervention</th>
<th>Outcome measures</th>
<th>Adverse effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkhaus et al&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12 sessions, 30 min per session&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N=73 (18M, 55F) Mean age=58</td>
<td>Minimal acupuncture</td>
<td>N=79 (25M, 54F) Mean age=59</td>
<td>Waiting list</td>
<td>&gt;6 months</td>
<td>8 weeks</td>
<td>VAS, FFbH-R score, PDI score</td>
<td>Reported</td>
</tr>
<tr>
<td>Cherkin et al&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10 sessions, 20 min per session&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N=162 (102M, 60F) Mean age=47</td>
<td>Simulated acupuncture</td>
<td>N=161 (97M, 64F) Mean age=46</td>
<td>Usual care</td>
<td>&gt;3 months</td>
<td>7 weeks</td>
<td>RMDQ, symptom bothersomeness</td>
<td>Reported</td>
</tr>
<tr>
<td>Haake et al&lt;sup&gt;c&lt;/sup&gt;</td>
<td>10 sessions, 30 min per session</td>
<td>N=387 (140M, 247F) Mean age=49</td>
<td>Sham acupuncture</td>
<td>N=388 (165M, 223F) Mean age=51</td>
<td>Standard therapy</td>
<td>&gt;6 months</td>
<td>6 weeks</td>
<td>FVAQ, HFAQ</td>
<td>Reported</td>
</tr>
<tr>
<td>Leibing et al&lt;sup&gt;d&lt;/sup&gt;</td>
<td>20 sessions, 30 min per session</td>
<td>N=45 (18M, 27F) Mean age=49</td>
<td>Sham acupuncture</td>
<td>N=46 (19M, 27F) Mean age=48</td>
<td>Control group</td>
<td>&gt;6 months</td>
<td>12 weeks</td>
<td>VAS, PDI</td>
<td>Reported</td>
</tr>
<tr>
<td>Molsberger et al&lt;sup&gt;e&lt;/sup&gt;</td>
<td>12 sessions, 3 per week, 30 min per session</td>
<td>N=61 (33M, 28F) Mean age=50</td>
<td>Sham acupuncture</td>
<td>N=60 (28M, 32F), Mean age=49</td>
<td>COT alone</td>
<td>≥6 weeks</td>
<td>4 weeks</td>
<td>VAS</td>
<td>No</td>
</tr>
<tr>
<td>Yuan et al&lt;sup&gt;f&lt;/sup&gt;</td>
<td>21 sessions, 20 min per session</td>
<td>N=50 (15M, 35F) Mean age=46</td>
<td>Sham acupuncture</td>
<td>N=50 (16M, 34F) Mean age=44</td>
<td>Waiting list</td>
<td>&lt;2 weeks</td>
<td>6 weeks</td>
<td>VAS, ODI</td>
<td>Reported</td>
</tr>
<tr>
<td>Vas et al&lt;sup&gt;g&lt;/sup&gt;</td>
<td>5 sessions, 20 min per session</td>
<td>N=68 (29M, 39F) Mean age=44; N=69 (35M, 34F) Mean age=44</td>
<td>Sham acupuncture; placebo acupuncture</td>
<td></td>
<td>Conventional treatment</td>
<td>&lt;2 weeks</td>
<td>2 weeks</td>
<td>VAS, RMDQ</td>
<td>No</td>
</tr>
</tbody>
</table>

**Notes:** <sup>a</sup>Usually two sessions in each of the first 4 weeks, followed by one session per week in the remaining 4 weeks. <sup>b</sup>Twice weekly for 3 weeks and then once weekly for 4 weeks.

**Abbreviations:** COT, conventional orthopedic therapy; CPGS, Chronic Pain Grade Scale; F, female; FFbH-R, Funktionsfragebogen Hannover-Rücken (back function questionnaire); HFAQ, Hanover Functional Ability Questionnaire; M, male; ODI, Oswestry Disability Index; PDI, Pain Disability Index; RMDQ, Roland-Morris Disability Questionnaire; VAS, Visual Analog Scale.
Figure 3 provides an overview of the methodological quality assessment of the included studies. Three of the trials had a low risk of bias. Two studies had a high risk of bias. The remaining two studies had an unclear risk of bias. All studies fully or partially fulfilled the categories of allocation concealment, selective outcome, incomplete outcome data, and other sources of bias. The two trials at high risk of bias were deemed as such because of the lack of blinding of participants and personnel.

Effects of interventions

The studies compared SA/PA to routine care or waiting list. Based on the current evidence, SA/PA was clearly more effective than routine care or waiting list for pain relief post-intervention, with an SMD of −0.36 (95% CI −0.54 to −0.18; F statistic=16%) in VAS and −0.35 (95% CI −0.49 to −0.20; F statistic=0%) in Chronic Pain Grade Scale (CPGS; Figure 4).

Five of the seven included trials measured functional disability outcomes, with two using the RMDQ, two using the Pain Disability Index (PDI), and one using the Oswestry Disability Index (ODI). No significant difference was observed between SA/PA and routine care or no treatment as measured immediately after the end of the sessions (RMDQ: SMD of −0.42; 95% CI −0.90 to 0.05; F statistic=66%; ODI: SMD of −0.30; 95% CI −0.69 to 0.10; Figure 5).

A summary of the findings for the main comparison is presented in Table 2.

Subgroup analysis

We added the subgroup analysis of VAS post-intervention, though we did not plan it in the protocol, because control groups in the included studies contained routine care and waiting list. Thus, we made the subgroup routine care versus waiting list. The results are consistent with pain differences in favor of SA/PA in both the routine care (mean difference [MD] −6.98 [95% CI −12.37, −1.59]) and waiting list subgroups (MD −13.70 [95% CI −20.32, −7.08]; Figure 6). The total MD and 95% CI were −9.66 (~−13 to 5.84) in favor of SA/PA. A similar finding was shown in chronic LBP subgroup for VAS (Figure 7). But there were no differences found in acute and chronic LBP subgroups for RMDQ (Figure 8).

Discussion

Summary of main results

Seven RCTs comprising a total of 1768 patients were included in this systematic review and meta-analysis. Moderately significant evidence was obtained for the
effectiveness of SA/PA for pain intensity as measured at the end of the treatments for LBP management. However, the review demonstrates scant support for the superior effectiveness of SA/PA compared with routine care or waiting list for back-specific functional status at the end of the treatments.

Overall completeness and applicability of evidence

We did not find a large effect size in the meta-analysis of continuous outcomes for pain intensity. All meta-analyses of continuous outcomes were performed using SMD values because the included trials used different measurement instruments for the outcomes of interest (pain and function). The disadvantage of using SMD values is that clinicians and patients are unlikely to relate to this way of presenting results. No serious adverse events were observed in these trials; the most common adverse event was increased pain after the massage sessions.

Quality of the evidence

In this review, we found two trials with unclear allocation of concealment bias and one trial with unclear selective reporting bias, suggesting that some studies may have exaggerated estimates of the intervention effect size compared with large trials. Two trials did not blind participants and personnel, suggesting that blinding patients and health-care providers was the most challenging methodological step in clinical trials of acupuncture, notwithstanding the recently improved methodological quality of the included RCTs.

Figure 3 Summary of risks of bias.

Notes: +, yes; -, no; ?, unclear.

Figure 4 Forest plot of comparison between SA/PA and routine care or waiting list for LBP. Outcome: pain intensity – VAS and CPGS.

Abbreviations: CPGS, Chronic Pain Grade Scale; IV, inverse variance; LBP, low back pain; PA, placebo acupuncture; SA, sham acupuncture; Std, standard; VAS, Visual Analog Scale.
nonspecific acupuncture points. Another type uses a non-penetration method, with sham needles having blunt tips. SA/P A includes several forms. In this review, one type is a superficial insertion into non-acupuncture points and nonspecific acupuncture points. Another type uses a non-penetration method, with sham needles having blunt and retractable tips.

For the points treated, some studies used a fixed protocol for all patients while others used a flexible set of points selected for each patient. Since both the methods are considered valid, this systematic review analyzed them together.

Limitations of this meta-analysis include the heterogeneity of the participants, because we included people with both acute and chronic nonspecific LBP, though we had subgroups. If there had been more literature, the evidence of the results would be more convincing with analyzing the types of LBP (acute, subacute, and chronic) separately, as they have different clinical approaches and disease trajectories. Additionally, the small number of studies for each SA/P A technique is another limitation. Furthermore, while no regions were excluded, the number of non-English-language journals indexed in electronic databases such as MEDLINE and EMBASE is limited. Should additional relevant trials be identified, this review will be updated.

Potential biases in the review process

Acupuncture is not a standardized therapy, and many variables may affect its potential effect on LBP, such as the needling technique, duration, frequency and number of sessions, points treated, manipulation, stimulus intensity, acupuncturists’ experience, patient heterogeneity, and confounding variables, such as co-interventions or the emotional effect of counseling by the acupuncturists.

SA/P A includes several forms. In this review, one type is a superficial insertion into non-acupuncture points and nonspecific acupuncture points. Another type uses a non-penetration method, with sham needles having blunt and retractable tips. For the points treated, some studies used a fixed protocol for all patients while others used a flexible set of points selected for each patient. Since both the methods are considered valid, this systematic review analyzed them together.

Limitations of this meta-analysis include the heterogeneity of the participants, because we included people with both acute and chronic nonspecific LBP, though we had subgroups. If there had been more literature, the evidence of the results would be more convincing with analyzing the types of LBP (acute, subacute, and chronic) separately, as they have different clinical approaches and disease trajectories. Additionally, the small number of studies for each SA/P A technique is another limitation. Furthermore, while no regions were excluded, the number of non-English-language journals indexed in electronic databases such as MEDLINE and EMBASE is limited. Should additional relevant trials be identified, this review will be updated.

Conclusion

Implications for practice

Notably, SA/P A was originally developed for use in acupuncture trials to determine the specific effects of needling. In clinical trials, procedures similar to real acupuncture may bias the results, and the sham or placebo control should be indistinguishable from the active treatment and yet be physiologically inert. Without meeting both criteria, it may not be appropriate to regard some procedures as sham or placebo controls within systematic reviews. While this review found evidence that SA/P A is more efficacious than routine care or waiting list in terms of pain relief at the end of the treatments, no evidence existed for improved function.

Additionally, variations of acupuncture must be considered. Based on TCM theory, all acupuncture procedures (e.g., points used, method of stimulation, and number of treatment sessions) must be performed according to individual differences; thus, acupuncture modalities vary among studies and are difficult to master and unify. Therefore, studies that examine the specific sham or placebo techniques used, including needle placement, needle insertion, acupuncture points or non-acupuncture points or nonspecific acupuncture points, and the acupuncturists’ experience are required, so as to differ distinguish from the real acupuncture. Furthermore, future research should focus on the standardization of outcome measures and the duration and frequency of sessions.

Besides, because acupuncture functions as a somatosensory-guided mind-body therapy, SA/P A could similarly enhance bodily sensations around the treatment site and
# Table 2 Summary of findings for the main comparison

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Corresponding risk</th>
<th>Relative effect (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>The mean pain intensity immediately after the treatment – pain (VAS) ranged across control groups from 39 to 58.6 mm</td>
<td>The mean pain intensity immediately after the treatment – pain (VAS) in the intervention groups was 0.36 SD lower (0.54 to 0.18 lower)</td>
<td>624 (6 studies)</td>
<td>⊕⊕⊝ moderate⁹</td>
<td>SMD = −0.36</td>
<td>(−0.54 to −0.18)</td>
</tr>
<tr>
<td>Pain intensity immediately after the treatment – pain (VAS)</td>
<td>Pain intensity immediately after the treatment – pain (CPGS) across control groups is 57.1</td>
<td>The mean pain intensity immediately after the treatment – pain (CPGS) in the intervention groups was 0.35 SD lower (0.49 to 0.2 lower)</td>
<td>736 (1 study)</td>
<td>⊕⊕⊕⊕ high</td>
<td>(−0.49 to −0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain intensity immediately after the treatment – Pain (CPGS)</td>
<td>Disability immediately after the treatment – disability (RMDQ) ranged across control groups from 8.9 to 39</td>
<td>The mean disability immediately after the treatment – disability (RMDQ) in the intervention groups was 0.11 SD higher (0.78 lower to 1 higher)</td>
<td>508 (3 studies)</td>
<td>⊕⊕⊕ moderate</td>
<td>SMD = 0.11</td>
<td>(−0.78 to 1)</td>
<td></td>
</tr>
<tr>
<td>Disability immediately after the treatment – Disability (RMDQ)</td>
<td>Disability immediately after the treatment – disability (PDI) ranged across control groups from 21.5 to 22.3</td>
<td>The mean disability immediately after the treatment – disability (PDI) in the intervention groups was 0.42 SD lower (0.9 lower to 0.05 higher)</td>
<td>223 (2 studies)</td>
<td>⊕⊕⊕⊕ low⁹</td>
<td>SMD = −0.42</td>
<td>(−0.9 to 0.05)</td>
<td></td>
</tr>
<tr>
<td>Disability immediately after the treatment – Disability (PDI)</td>
<td>Disability immediately after the treatment – disability (ODI) across control groups is 49.7</td>
<td>The mean disability immediately after the treatment – disability (ODI) in the intervention groups was 0.3 SD lower (0.69 lower to 0.1 higher)</td>
<td>100 (1 study)</td>
<td>⊕⊕⊕⊕ low⁹</td>
<td>SMD = −0.3</td>
<td>(−0.69 to 0.1)</td>
<td></td>
</tr>
<tr>
<td>Disability immediately after the treatment – Disability (ODI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
* The basis for the assumed risk (e.g., the median control group risk across studies). The corresponding risk (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). GRADE Working Group grades of evidence. High quality (⊕⊕⊕⊕): further research is very unlikely to change our confidence in the estimate of effect. Moderate quality (⊕⊕⊕): further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. Low quality (⊕⊕): further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Very low quality (⊕): we are very uncertain about the estimate. *Two articles lack blinding of participants and personnel.* 95% CI includes no effect.

**Abbreviations:** CPGS, Chronic Pain Grade Scale; ODI, Oswestry Disability Index; PDI, Pain Disability Index; RMDQ, Roland-Morris Disability Questionnaire; SMD, standard mean difference; VAS, Visual Analog Scale.
### 1.3.1 Pain intensity: acupuncture vs routine care

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Routine care Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Fixed, 95% CI</th>
<th>Std mean difference IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leibing et al13</td>
<td>32 (22)</td>
<td>44 (17)</td>
<td>39</td>
<td>38.8%</td>
<td>-12.00 [-20.66, -3.34]</td>
</tr>
<tr>
<td>Molsberger et al27</td>
<td>36 (19)</td>
<td>39 (21)</td>
<td>58</td>
<td>54.7%</td>
<td>-3.00 [-10.29, 4.29]</td>
</tr>
<tr>
<td>Vas et al14</td>
<td>33 (50.8)</td>
<td>49 (72.1)</td>
<td>30</td>
<td>3.5%</td>
<td>-16.00 [-44.73, 12.73]</td>
</tr>
<tr>
<td>Vas et al14</td>
<td>46 (70.3)</td>
<td>49 (72.1)</td>
<td>30</td>
<td>3.0%</td>
<td>-4.00 [-35.10, 27.10]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>223</td>
<td>157</td>
<td>100.0%</td>
<td>-6.98 [-12.37, -1.59]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $I^2 = 2.85$, $df = 3$ ($p = 0.42$); $I^2 = 0%$
Test for overall effect: $Z = 2.54$ ($p = 0.01$)
Test for subgroup differences: $X^2 = 2.38$, $df = 1$ ($p = 0.58$); $I^2 = 0%$

### 1.3.2 Pain intensity: acupuncture vs waiting list

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Waiting list Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Fixed, 95% CI</th>
<th>Std mean difference IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkhaus et al10</td>
<td>43.7 (29.8)</td>
<td>58.6 (25.1)</td>
<td>74</td>
<td>53.8%</td>
<td>-14.90 [-23.92, -5.88]</td>
</tr>
<tr>
<td>Yuan et al28</td>
<td>46 (27)</td>
<td>58.3 (22.5)</td>
<td>50</td>
<td>46.2%</td>
<td>-12.30 [-22.04, -2.56]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>125</td>
<td>124</td>
<td>100.0%</td>
<td>-13.70 [-20.32, -7.08]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $X^2 = 0.15$, $df = 1$ ($p = 0.70$); $I^2 = 0%$
Test for overall effect: $Z = 4.06$ ($p < 0.0001$)

### Figure 6 Subgroup of VAS post-intervention routine care versus waiting list.

#### Abbreviations: IV, inverse variance; Std, standard; VAS, Visual Analog Scale.

### 1.4.1 Acute LBP

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Routine care Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Fixed, 95% CI</th>
<th>Std mean difference IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vas et al14</td>
<td>33 (50.8)</td>
<td>49 (72.1)</td>
<td>30</td>
<td>2.1%</td>
<td>-16.00 [-44.73, 12.73]</td>
</tr>
<tr>
<td>Vas et al14</td>
<td>45 (70.3)</td>
<td>50 (86.2)</td>
<td>30</td>
<td>1.8%</td>
<td>-4.00 [-35.10, 27.10]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>125</td>
<td>60</td>
<td>3.9%</td>
<td>-10.47 [-31.58, 10.63]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $X^2 = 0.31$, $df = 1$ ($p = 0.58$); $I^2 = 0%$
Test for overall effect: $Z = 0.97$ ($p = 0.33$)

### 1.4.2 Chronic LBP

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Routine care Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Fixed, 95% CI</th>
<th>Std mean difference IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkhaus et al10</td>
<td>43.7 (29.8)</td>
<td>58.2 (25.1)</td>
<td>74</td>
<td>21.5%</td>
<td>-14.90 [-23.92, -5.88]</td>
</tr>
<tr>
<td>Leibing et al13</td>
<td>32 (22)</td>
<td>44 (17)</td>
<td>39</td>
<td>23.3%</td>
<td>-12.00 [-20.66, -3.34]</td>
</tr>
<tr>
<td>Molsberger et al27</td>
<td>36 (19)</td>
<td>39 (21)</td>
<td>58</td>
<td>32.9%</td>
<td>-3.00 [-10.29, 4.29]</td>
</tr>
<tr>
<td>Yuan et al28</td>
<td>46 (27)</td>
<td>50 (83.2)</td>
<td>50</td>
<td>18.4%</td>
<td>-12.30 [-22.04, -2.56]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>218</td>
<td>221</td>
<td>96.1%</td>
<td>-9.62 [-13.89, -5.36]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $X^2 = 5.07$, $df = 3$ ($p = 0.17$); $I^2 = 41%$
Test for overall effect: $Z = 4.42$ ($p = 0.00001$)
Test for subgroup differences: $X^2 = 5.38$, $df = 5$ ($p = 0.37$); $I^2 = 7%$
Test for overall effect: $Z = 5.53$ ($p = 0.00001$)
Test for subgroup differences: $X^2 = 0.01$, $df = 1$ ($p = 0.94$); $I^2 = 0%$

### Figure 7 Subgroup of VAS – acute LBP versus chronic LBP.

#### Abbreviations: IV, inverse variance; LBP, low back pain; VAS, Visual Analog Scale.

### 1.5.1 Acute LBP

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Routine care Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Random, 95% CI</th>
<th>Std mean difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vas et al14</td>
<td>57.2 (62.9)</td>
<td>39 (62.9)</td>
<td>30</td>
<td>29.1%</td>
<td>18.20 [-9.16, 45.56]</td>
</tr>
<tr>
<td>Vas et al14</td>
<td>77.6 (45.3)</td>
<td>39 (62.9)</td>
<td>30</td>
<td>30.4%</td>
<td>13.42 [-36.78]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>125</td>
<td>60</td>
<td>59.5%</td>
<td>29.13 [19.9, 49.07]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $X^2 = 28.13$, $df = 16$ ($p = 0.08$); $I^2 = 14%$
Test for overall effect: $Z = 2.86$ ($p = 0.004$)

### 1.5.2 Chronic LBP

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Sham acupuncture Mean (SD)</th>
<th>Routine care Mean (SD)</th>
<th>Weight</th>
<th>Mean difference IV, Random, 95% CI</th>
<th>Std mean difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cherkin et al11</td>
<td>5.4 (4.9)</td>
<td>8.9</td>
<td>6</td>
<td>40.5%</td>
<td>-4.70 [-5.50, -3.90]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>162</td>
<td>161</td>
<td>40.5%</td>
<td>-4.70 [-5.50, -3.90]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: not applicable
Test for overall effect: $Z = 5.74$ ($p = 0.00001$)

### Figure 8 Subgroup of RMDQ – acute LBP versus chronic LBP.

#### Abbreviations: IV, inverse variance; LBP, low back pain; RMDQ, Roland-Morris Disability Questionnaire.
induce potential effects related to endogenous pain modulation in the brain.27–29

Thus, we should avoid prematurely concluding that SA/PA is appropriate for acupuncture research. Moreover, guidelines should be developed to assess acupuncture sham–placebo control methods to address the specific effect of acupuncture over placebo.

Implications for research

To provide homogeneous information for future reviews, trial authors are encouraged to use the Consolidated Standards of Reporting Trials (CONSORT) statement as a model for reporting their trials (www.consort-statement.org) and the Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA) criteria30 to report the interventions. From the available trials we included in this meta-analysis, superficial stimulation appears a more promising treatment for LBP than routine care in relieving pain post-intervention. Further studies to assess the superiority of SA/PA are required, as are those with larger sample sizes and those that assess the role of session length by including two (or more) levels of this variable; moreover, the experience of the therapist should also be considered by including practitioners with different levels of experience and training.

Acknowledgment

Prospero registration number: CRD42017070122.

Disclosure

The authors report no conflicts of interest in this work.

References
